

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده: شیمی

مطالعه ارتباط کمی ساختار - فعالیت مشتقات بنزیمیدازول به عنوان دسته‌ی

جدیدی از داروهای ضدانگل

دانشجو: طاهره موذن

استاد راهنما:

دکتر منصور عرب چم جنگلی

استاد مشاور:

دکتر ناصر گودرزی

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ۱۳۹۰



تقدیم به

همسر مهربان

و

پسر عزیزم

تشکر و قدردانی

..... و ما اگر در مقابل نیکان روزگار با تمام وجود سپاسگزار نباشیم، برگمی هستیم که از شاخسار هستی جا مانده ایم.....

اکنون که به یاری ایزد یکتا دوره کارشناسی ارشد خود را به پایان رسانده ام، وظیفه‌ی خود می‌دانم که از زحمات عزیزانی که مرا در انجام این کار یاری نموده‌اند، تشکر نمایم. در ابتدا از استاد راهنمای گرامیم جناب آقای دکتر عرب به خاطر راهنمایی‌های بی‌دریغ و حمایت‌های همه‌جانبه‌ی ایشان سپاسگزاری نموده، برای ایشان سلامتی و موفقیت روزافزون آرزو می‌نمایم. از استاد مشاور ارجمند جناب آقای دکتر گودرزی که از نقطه نظرات سودمند و مساعدتهای ایشان بهره‌مند شده‌ام کمال تشکر را دارم.

از اساتید داور محترم جناب آقای دکتر باقریان و سرکار خانم دکتر مصدرالامور به خاطر مطالعه‌ی پایان‌نامه و شرکت در جلسه دفاع قدردانی می‌نمایم. از دوستان عزیزم بخصوص خانم اشرفی و برومند که مرا در انجام این پروژه یاری نمودند و با مهربانی پاسخگوی پرسش‌های من بودند کمال تشکر را دارم.

زحمات بی‌دریغ و حمایت‌های پدر و مادر عزیزم در طول دوران تحصیل، همواره موجب دلگرمی و آسودگی خاطر برایم بوده است و بیشک در سایه‌ی حمایت‌های آنها امروز مرحله‌ی علمی دیگری را با موفقیت به پایان رسانده‌ام. زحمات آنها را ارج می‌نهم و امیدوارم فرزند شایسته و قدرشناسی برای آنها باشم.

همسرم رضا و پسرم محمدرسول به خاطر شکیبایی در این دوران سخت، شایسته‌ی خالصترین تشکر و گرم‌ترین قدردانی

هستند از این رو این پایان نامه را به ایشان تقدیم می-  
دارم .

## طاهره مؤذن

بهمن ۱۳۹۹

چکیده:

مشتقات بنزیمیدازول، یک طبقه‌ی جدید از داروهای ضدانگل هستند که برای درمان عفونت‌های ناشی از انگل‌های متفاوتی مانند گیاردیا اینتستینالیس و تریکوموناس واژینالیس استفاده می‌شوند. در بخش اول پروژه، رابطه کمی ساختار- فعالیت (QSAR) ۴۲ ترکیب از مشتقات بنزیمیدازول به عنوان داروهای ضد انگل گیاردیا اینتستینالیس، با استفاده از شبکه عصبی مصنوعی با الگوریتم تنظیم بایزین مورد مطالعه قرار گرفت و با روش رگرسیون خطی چندگانه مقایسه گردید. برای انتخاب مهمترین توصیفگرها، رگرسیون خطی مرحله‌ای مورد استفاده قرار گرفت. سپس توصیفگرهای مهم به عنوان ورودی برای ایجاد مدل های QSAR با استفاده از رگرسیون خطی چندگانه (MLR) و شبکه عصبی (ANN)، استفاده شدند. سری داده‌ها به طور تصادفی به سه سری آموزش، ارزیابی و تست تقسیم گردید و مدل‌های رگرسیون خطی چندگانه و شبکه عصبی مصنوعی ساخته شدند. اعتبار این مدل‌ها علاوه بر به کارگیری سری تست، توسط تکنیک رد مرحله‌ای تک تک و  $Y$ - تصادفی بررسی شد. هر دو روش خطی و غیر خطی توانایی پیش‌بینی خوبی دارند، اما مدل ANN نتایج صحیح‌تری دارد. خطای مجذور میانگین (MSE) برای سری تست توسط مدل های MLR و ANN به ترتیب ۰/۰۴۷ و ۰/۰۳۹ بودند.

در بخش دوم پروژه، رابطه کمی ساختار - فعالیت ۴۹ ترکیب از مشتقات بنزیمیدازول به عنوان داروهای ضد انگل تریکوموناس واژینالیس، مورد پژوهش قرار گرفت. در این تحقیق، سری داده‌ها به طور تصادفی به سه سری آموزش، ارزیابی و تست تقسیم شد که هر یک به ترتیب شامل ۲۹، ۱۰ و ۱۰ ترکیب می باشند. مدل رگرسیون خطی چندگانه، با استفاده از ۵ توصیف کننده با قدرت بالایی

مقادیر  $pIC_{50}$  مربوط به هر یک از ترکیبات را پیش بینی نمود. خطای مجذور میانگین (mse) برای

سری تست توسط مدل های MLR و ANN به ترتیب ۰/۰۸۷ و ۰/۰۸۹ بودند.

**کلیدواژه‌ها:** ارتباط کمی ساختار- فعالیت، بنزیمیدازول، شبکه عصبی مصنوعی، رگرسیون خطی

چندگانه

نتایج حاصل از این پایان نامه در پوستری تحت عنوان:

Use of Bayesian regularized Artificial Neural Network in a QSAR study of benzimidazole derivatives as new of anti protozoal compounds

در هجدهمین سمینار شیمی تجزیه ایران در دانشگاه زاهدان

و در پوستری تحت عنوان

Linear and nonlinear QSAR study benzimidazole derivatives as novel Anti parasitic chemicals

در سومین سمینار کمومتریکس ایران در تبریز پذیرفته شد.

## فهرست مطالب

### فصل اول

- ۱-۱- بیماری های انگلی..... ۱
- ۱-۱-۱- انگل و تاریخچه ی رشد و تکامل آن..... ۲
- ۱-۱-۱-۱- شرایط محیطی..... ۳
- ۱-۲- انتقال عامل بیماریزای عفونی..... ۴
- ۱-۳- علائم..... ۴
- ۱-۴- درمان..... ۵
- ۱-۵- گیار دیاسیس..... ۶
- ۱-۵-۱- تاریخچه ی گیار دیا..... ۶
- ۱-۵-۱-۲- سیکل زندگی گیار دیا..... ۷
- ۱-۳-۵- علائم و تشخیص..... ۹
- ۱-۴-۵- درمان..... ۹
- ۱-۶-۱-۱- تریکومونیا سیس..... ۱۰
- ۱-۶-۱-۱- تمایز تریکوموناس واژینالیس با دیگر تک یا خنگان..... ۱۱
- ۱-۶-۱-۲- بیماریزایی تریکوموناس واژینالیس..... ۱۱
- ۱-۳-۶- علائم و عوارض..... ۱۲
- ۱-۴-۶- درمان..... ۱۲
- ۲-۱- پیشینه ی کارهای انجام شده..... ۱۳

### فصل دوم

- ۱-۲- ارتباط کمی ساختار- فعالیت..... ۱۶
- ۲-۲- مراحل مطالعه ی کمی ساختار- فعالیت (QSAR)..... ۱۶
- ۱-۲-۲- فراهم کردن سری داده ها..... ۱۷
- ۲-۲-۲- رسم و بهینه سازی ساختار هندسی..... ۱۸
- ۳-۲-۲- استخراج توصیفگرها..... ۱۹
- ۴-۲-۲- انتخاب بهترین توصیف کننده ها..... ۲۱
- ۱-۴-۲-۲- روش ورود اجباری..... ۲۲
- ۲-۴-۲-۲- روش جلوبرنده..... ۲۲
- ۳-۴-۲-۲- روش عقب برنده..... ۲۳
- ۴-۴-۲-۲- روش مرحله ای..... ۲۳



۲۳	..... ساختن مدل مناسب
۲۴	..... مدل سازی به کمک روش MLR
۲۵	..... شبکه عصبی مصنوعی
۲۷	..... سیستم عصبی طبیعی
۲۸	..... نرون مصنوعی
۳۰	..... توابع انتقال نرون
۳۱	..... انواع شبکه‌های عصبی از نظر برگشت پذیری
۳۲	..... معماری شبکه‌های پیشخور
۳۳	..... مراحل آموزش در شبکه‌های پیشخور
۳۳	..... ایجاد شبکه
۳۳	..... مقدار دهی آغازین به وزن‌ها
۳۳	..... آموزش شبکه
۳۴	..... روش‌های آموزش
۳۵	..... آموزش شبکه‌های پیشخور با تکنیک پس انتشار
۳۹	..... بهبود تعمیم
۴۲	..... ارزیابی مدل
۴۴	..... تکنیک‌های سنجش اعتبار
۴۴	..... ارزیابی متقاطع
۴۴	..... اعتبارسنجی بیرونی
۴۵	..... تقسیم مجموعه به آموزش و ارزیابی
۴۵	..... آزمون Y- تصادفی
۴۵	..... نرم‌افزارهای مورد استفاده در تحقیق
۴۶	..... نرم‌افزار Hyperchem
۴۶	..... نرم‌افزار Dragon
۴۷	..... نرم‌افزار SPSS
۴۷	..... نرم‌افزار MATLAB

### فصل سوم

۴۹	..... مدل سازی فعالیت دارویی مشتقات بنزیمیدازول به عنوان داروهای ضد انگل
۴۹	..... ۱-۱-۳ سری داده‌ها
۵۲	..... ۲-۱-۳ رسم و بهینه سازی ساختار هندسی مولکول‌ها
۵۲	..... ۳-۱-۳ محاسبه توصیفگرهای مولکولی
۵۲	..... ۴-۱-۳ انتخاب بهترین توصیف کننده‌ها
۵۳	..... ۵-۱-۳ به دست آوردن مدل نهایی به روش رگرسیون خطی چندگانه (MLR)
۵۷	..... ۶-۱-۳ مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN)
۵۸	..... ۷-۱-۳ بهینه سازی پارامترهای شبکه عصبی
۸۹	..... ۱-۷-۱-۳ انتخاب تعداد لایه‌های شبکه
۵۸	..... ۲-۷-۱-۳ انتخاب نوع الگوریتم آموزشی، نوع تابع انتقال، تعداد متغیرهای ورودی و تعداد گره‌های لایه ی پنهان

۶۴	۳-۷-۱-۳- بهینه سازی پارامتر $\mu$ .....
۵۶	۳-۷-۱-۴- بهینه سازی تعداد دورهای آموزشی.....
۶۶	۳-۷-۱-۵- مشخصات و هندسه ی شبکه عصبی بهینه.....
۶۷	۳-۲-۲- ارزیابی مدل های خطی و غیرخطی ارائه شده و مقایسه ی مدل ها.....
۶۸	۳-۲-۱- ارزیابی مدل ها با استفاده از سری ارزیابی.....
۶۹	۳-۲-۲- ارزیابی مدل ها با استفاده از سری تست.....
۷۰	۳-۲-۳- ارزیابی مدل ها به روش رد مرحله ای تک تک.....
۷۴	۳-۲-۴- آزمون Y- تصادفی.....
۷۵	۳-۲-۵- ارزیابی مدل ها ی ارائه شده با استفاده از پارامترهای آماری.....
۷۵	۳-۲-۶- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی.....
۷۷	۳-۲-۷- بررسی ارتباط توصیف کننده های ساختاری وارد شده در مدل و فعالیت ضد انگلی.....
۷۷	۳-۲-۷-۱- توصیفگرهای RDF.....
۷۸	۳-۲-۷-۲- توصیفگرهای 2D-autocorrelation.....
۷۹	۳-۲-۷-۳- توصیف کننده های BCUT.....
۸۰	۳-۲-۷-۴- توصیف کننده های GETAWAY.....
۸۱	۳-۲- نتیجه گیری نهایی.....

#### فصل چهارم

۸۲	۴-۱- مدل سازی فعالیت دارویی مشتقات بنزیمیدازول به عنوان داروهای ضد انگل.....
۸۲	۴-۱-۱- نرم افزارهای مورد استفاده.....
۸۲	۴-۱-۲- سری داده ها.....
۸۶	۴-۱-۳- رسم و بهینه سازی ساختار مولکول ها.....
۸۶	۴-۱-۴- محاسبه و انتخاب توصیفگرهای مهم با روش رگرسیون خطی چندگانه.....
۹۱	۴-۱-۵- مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN).....
۹۱	۴-۱-۶- بهینه سازی پارامترهای شبکه عصبی.....
۹۱	۴-۱-۶-۱- انتخاب تعداد لایه های شبکه.....
۹۲	۴-۱-۶-۲- انتخاب تعداد ورودی های شبکه، نوع تابع انتقال و آموزش و تعداد نرون های لایه ی پنهان.....
۹۷	۴-۱-۶-۳- انتخاب تعداد دورهای آموزش.....
۹۹	۴-۱-۶-۴- بهینه سازی پارامتر $\mu$ .....
۹۹	۴-۱-۶-۵- مشخصات و هندسه ی شبکه عصبی بهینه.....
۱۰۰	۴-۲- ارزیابی مدل های بدست آمده و مقایسه مدل ها.....
۱۰۰	۴-۲-۱- ارزیابی مدل ها با استفاده از سری ارزیابی.....
۱۰۱	۴-۲-۲- ارزیابی مدل ها با استفاده از سری تست.....
۱۰۱	۴-۲-۳- ارزیابی مدل ها به روش رد مرحله ای تک تک.....
۱۰۶	۴-۲-۴- آزمون Y- تصادفی.....
۱۰۷	۴-۲-۵- ارزیابی مدل ها ی ارائه شده با استفاده از پارامترهای آماری.....
۱۰۷	۴-۲-۶- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی.....
۱۰۸	۴-۲-۷- بررسی ارتباط توصیفگرهای انتخاب شده با فعالیت بازدارندگی.....
۱۰۹	۴-۲-۷-۱- توصیفگرهای Topological.....

۱۱۰	.....3D MORSE توصیفگرهای
۱۱۰	.....Charge توصیفگرهای
۱۱۱	..... functional groups توصیفگرهای
۱۱۲	..... نتیجه‌گیری نهایی
۱۱۲	..... آینده نگری
۱۱۳	..... مراجع و منابع

## فهرست اشکال

۳	..... شکل (۱-۱) - شماتیکی از انگل
۸	..... شکل (۲-۱) - سیکل زندگی گیاردیا اینتستینالیس
۱۷	..... شکل (۱-۲) مراحل QSAR
۲۸	..... شکل (۲-۲) - شماتیک سلول عصبی طبیعی
۲۹	..... شکل (۳-۲) - مدل نرون مصنوعی
۳۰	..... شکل (۴-۲) - نرون با یک بردار ورودی
۳۱	..... شکل (۵-۲) - برخی از توابع انتقال مورد استفاده در شبکه های عصبی
۳۲	..... شکل (۶-۲) - انواع شبکه‌های عصبی: الف: شبکه عصبی پیشخور ب: شبکه عصبی برگشتی
۳۳	..... شکل (۷-۲) - شبکه پیشخور tansig/pureline
۳۵	..... شکل (۸-۲) - شیوه آموزش نظارت شده
۳۷	..... شکل (۹-۲) - نحوه‌ی عملکرد تکنیک پس‌انتشار در یک شبکه عصبی پیشخور
۴۱	..... شکل (۱۰-۲) - تغییرات خطای سری آموزش و سری ارزیابی
۵۰	..... شکل (۱-۳) - ساختار عمومی مشتقات بنزیمیدازول
۶۱	..... شکل (۲-۳) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش لونیبرگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی
۶۲	..... شکل (۳-۳) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش لونیبرگ-ماکوارت و تابع انتقال تانژانت سیگموئید
۶۳	..... شکل (۴-۳) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

- شکل (۳-۵) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی..... ۶۴
- شکل (۳-۶) - نمودار میانگین مربع خطا بر حسب پارامتر  $\mu$ ..... ۶۵
- شکل (۳-۷) - نمودار نمودار میانگین مربع خطا بر حسب تعداد دوره‌های آموزش متفاوت..... ۶۶
- شکل (۳-۸) - هندسه شبکه عصبی بهینه..... ۶۸
- شکل (۳-۹) - نمودار مقادیر پیش بینی شده  $pIC_{50}$  توسط الف) روش MLR ب) شبکه عصبی مصنوعی بر حسب مقادیر تجربی برای سری ارزیابی..... ۷۰
- شکل (۳-۱۰) - نمودار مقادیر پیش بینی شده  $pIC_{50}$  توسط الف) روش MLR ب) شبکه عصبی مصنوعی بر حسب مقادیر تجربی برای سری تست..... ۷۱
- شکل (۳-۱۱) - نمودار مقادیر پیش بینی شده  $pIC_{50}$  توسط الف) روش MLR ب) شبکه عصبی مصنوعی بر حسب مقادیر تجربی به روش رد مرحله ای تک تک..... ۷۴
- شکل (۳-۱۲) - نمودار باقیمانده ها برای کل داده ها الف) با مدل MLR ب) با مدل ANN..... ۷۴
- شکل (۳-۱۳) - مشارکت توصیفگرها در شبکه عصبی بهینه..... ۷۷
- شکل (۴-۱) - اسکلت مولکولی ترکیبات مورد بررسی..... ۸۴
- شکل (۴-۲) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش لونیبرگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی..... ۹۴
- شکل (۴-۳) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش لونیبرگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی..... ۹۵
- شکل (۴-۴) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی..... ۹۶
- شکل (۴-۵) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی..... ۹۷
- شکل (۴-۶) - نمودار میانگین مربع خطاها بر حسب تعداد دوره‌های آموزش متفاوت..... ۹۸
- شکل (۴-۷) - نمودار میانگین مربع خطاها بر حسب پارامتر..... ۹۹
- شکل (۴-۸) - هندسه شبکه عصبی بهینه..... ۱۰۰
- شکل (۴-۹) - نمودار مقادیر پیش بینی  $pIC_{50}$  بر حسب مقادیر تجربی برای سری ارزیابی الف) با مدل MLR ب) با مدل ANN..... ۱۰۲

- شکل (۴-۱۰) - نمودار مقادیر پیش‌بینی شده  $pIC_{50}$  بر حسب مقادیر تجربی برای سری تست الف) با مدل MLR (ب) با مدل ANN ..... ۱۰۳
- شکل (۴-۱۱) - نمودار مقادیر پیش‌بینی  $pIC_{50}$  بر حسب مقادیر تجربی برای کل داده‌ها الف) با مدل MLR (ب) با مدل ANN ..... ۱۰۶
- شکل (۴-۱۲) - نمودار باقیمانده‌ها بر حسب مقادیر واقعی برای کل داده‌ها الف) با مدل MLR (ب) با مدل ANN ..... ۱۰۶
- شکل (۴-۱۳) - مشارکت توصیفگرها در شبکه عصبی بهینه ..... ۱۰۹

### فهرست جداول

- جدول (۲-۱) - توصیفگرهای محاسبه شده توسط نرم افزار Dragon ..... ۴۷
- جدول (۳-۱) - ساختار و مقادیر  $pIC_{50}$  ترکیبات مورد بررسی ..... ۵۱
- جدول (۳-۲) - کل توصیفگرهای انتخاب شده بوسیله ی رگرسیون خطی چندگانه مرحله به مرحله ..... ۵۴
- جدول (۳-۳) - مقادیر  $mse$  و  $R$  محاسبه شده برای سری ارزیابی برای مدل های ۱۴-۲ ..... ۵۵
- جدول (۳-۴) - توصیف کننده های وارد شده در مدل MLR منتخب ..... ۵۶
- جدول (۳-۵) - ماتریس همبستگی توصیفگرهای منتخب ..... ۵۶
- جدول (۳-۶) - مقادیر عددی توصیفگرهای وارد شده در مدل MLR نهایی ..... ۵۷
- جدول (۳-۷) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان ..... ۶۱
- جدول (۳-۸) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان ..... ۶۲
- جدول (۳-۹) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان ..... ۶۳
- جدول (۳-۱۰) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان ..... ۶۴
- جدول (۳-۱۱) - مقادیر خطای شبکه با مقادیر  $\mu$  ..... ۶۵
- جدول (۳-۱۲) - مقادیر میانگین مربع خطا در تعداد دوره های آموزشی متفاوت ..... ۶۶
- جدول (۳-۱۳) - مشخصات شبکه بهینه ..... ۶۷
- جدول (۳-۱۴) - توصیفگرهای مورد استفاده در شبکه عصبی ..... ۶۷

- جدول (۳-۱۵) - نتایج حاصل از مدل رگرسیون خطی چندگانه و شبکه عصبی برای سری ارزیابی ..... ۶۹
- جدول (۳-۱۶) - نتایج حاصل از مدل رگرسیون خطی چندگانه و شبکه عصبی برای سری تست ..... ۷۰
- جدول (۳-۱۷) - نتایج حاصل از ارزیابی مدل رگرسیون خطی چندگانه و شبکه عصبی با استفاده از روش رد مرحله ای تک تک ..... ۷۲
- جدول (۳-۱۸) - مقادیر ضریب همبستگی برای سری تست و سری ارزیابی بعد از چندین تست Y- تصادفی ..... ۷۵
- جدول (۳-۱۹) - پارامترهای آماری محاسبه شده برای مدل رگرسیون خطی چندگانه و شبکه عصبی ..... ۷۶
- جدول (۳-۲۰) - مثال هایی از اثر توصیفگر MATS7e بر فعالیت دارویی ..... ۸۰
- جدول (۴-۱) - سری داده های بکار رفته در مدل سازی ..... ۸۴
- جدول (۴-۲) - مقایسه آماره های چند مدل نهایی به دست آمده از روش MLR ..... ۸۸
- جدول (۴-۳) - توصیف کننده های وارد شده در مدل MLR به همراه کلاس مربوطه ..... ۸۹
- جدول (۴-۴) - ماتریس همبستگی توصیف کننده های به کار رفته در مدل نهایی ..... ۸۹
- جدول (۴-۵) - مقادیر عددی توصیفگرهای منتخب در مدل MLR نهایی ..... ۹۰
- جدول (۴-۶) - مقادیر میانگین مربعات خطاهای سری ارزیابی برای شبکه با الگوریتم آموزشی لونیبرگ - مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان ..... ۹۴
- جدول (۴-۷) - مقادیر میانگین مربعات خطاهای سری ارزیابی برای شبکه با الگوریتم آموزشی لونیبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان ..... ۹۵
- جدول (۴-۸) - مقادیر میانگین مربعات خطاهای سری ارزیابی برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان ..... ۹۶
- جدول (۴-۹) - مقادیر میانگین مربعات خطاهای سری ارزیابی 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 (۴-۱۰) - مقادیر میانگین مربع خطای شبکه با تغییر تعداد دوره های آموزشی ..... ۹۸
- جدول (۴-۱۱) - مقادیر میانگین مربع خطا در مقادیر متفاوت  $\mu$  ..... ۹۹
- جدول (۴-۱۲) - پارامترهای بهینه شده شبکه عصبی ..... ۱۰۰
- جدول (۴-۱۳) - نتایج حاصل از ارزیابی مدل های برتر MLR و ANN با استفاده از سری ارزیابی ..... ۱۰۱
- جدول (۴-۱۴) - نتایج حاصل از ارزیابی مدل های برتر MLR و ANN با استفاده از سری تست ..... ۱۰۲
- جدول (۴-۱۵) - نتایج حاصل از ارزیابی مدل MLR و ANN با روش رد مرحله ای تک تک ..... ۱۰۴

- جدول (۴-۱۶) - مقادیر R برای سری ارزیابی و سری تست بعد از چندین تست Y- تصادفی ..... ۱۰۷
- جدول (۴-۱۷) - پارامترهای آماری محاسبه شده ..... ۱۰۸
- جدول (۴-۱۸) - چند مثال از اثر توصیفگر X0sol ..... ۱۱۰
- جدول (۴-۱۹) - چند مثال از اثر توصیفگر nCONR<sub>2</sub>ph و nCOORph بر فعالیت دارویی ..... ۱۱۲

# مقدمه



## ۱-۱- بیماریهای انگلی

امروزه در بسیاری از نقاط جهان، بیماریهای مسری مشکلات زیادی را برای جوامع انسانی بوجود آورده است؛ این در حالی است که آسیبها و صدمات و خسارات، کشورهای در حال توسعه را در مقایسه با کشورهای صنعتی بیشتر تحت تأثیر قرار می دهد. در این میان بیماریهای انگلی با انتشاری گسترده به عللی از قبیل فقر، سوء تغذیه، بی سوادی، ازدیاد بی رویه جمعیت، فقدان تسهیلات بهداشتی و دهها عامل دیگر بخش عمده ای از مشکلات را به خود اختصاص می دهد، به گونه ای که در برخی مناطق جهان سهم بیماریهای انگلی در ایجاد خسارات اجتماعی، اقتصادی با برخی بیماریها نظیر سل، بیماریهای مقاربتی و عفونتهای حاد تنفسی برابری می کند [۱].

بیماری انگلی بیماری است که بوسیله ی انگل ایجاد و سرایت داده می شود. این بیماریها می توانند همه ی ارگانسیم های جاندار شامل گیاهان و پستانداران را تحت تأثیر قرار دهند. مطالعه ی بیماریهای انگلی، انگل شناسی<sup>۱</sup> نامیده می شود. بعضی انگلها مثل توکسوپلازما گاندی<sup>۲</sup> و مالاریا مستقیماً باعث بیماری می شوند اما ارگانسیم های دیگر بوسیله ی زهری که تولید می کنند باعث بیماری می شوند.

سرایت و انتقال بیماریهای انگلی به عوامل سه گانه ی زیر بستگی دارد:

۱- وجود میزبان حساس

۲- وجود منبع آلودگی

۳- طرق انتقال آلودگی

در اینجا سه نوع اصلی ارگانسیم هایی که باعث بیماریهای انگلی می شوند ذکر می شود که عبارتند

از:

❖ تک یاخته ← باعث عفونت تک یاخته ای می شود.

1-Parasitology

2-Toxoplasma gondii

❖ کرم روده ← باعث بیماری کرمی می‌شود.

❖ انگل خارجی

تک یاخته و کرم معمولاً انگل داخلی هستند یعنی داخل بدن میزبان زندگی می‌کنند مانند کرم روده‌ای، در حالیکه انگل خارجی معمولاً روی سطح میزبان زندگی می‌کند مانند شپشه‌ها [۲].

از آنجا که بیماریهای انگلی دارای سیر مزمن بوده و کمتر با آثار و علایم بالینی همراه هستند لذا افراد آلوده ممکن است بعد از بهبود ظاهری و بدون داشتن هر گونه علایم کلینیکی به صورت ناقلین بیماری در آیند و خود نقش آلوده را برای افراد سالم به عهده بگیرند. به عبارت دیگر نظر به اینکه در افراد ناقل، تعادل بیولوژیکی خاص بین انگل و میزبان بوجود می‌آید، لذا اینگونه ناقلین در حالیکه از صدمات انگل محفوظ هستند در عین حال منبع و ناقل عامل بیماری می‌باشند. بدین جهت آگاهی از نقش بیماریهای انگلی و راههای پیشگیری از آنها از اهمیت خاصی برخوردار است.

### ۱-۱-۱- انگل و تاریخچه ی رشد و تکامل آن

انگل یا پارازیت<sup>۱</sup> (شکل ۱-۱) به موجودی اطلاق می‌شود که در داخل یا خارج پیکر موجود زنده‌ی دیگری به طور موقت مستقر شده و برای رشد و تکثیر خود از محیط مناسب و مواد غذایی موجود میزبان استفاده می‌کند. لزومی ندارد انگل به میزبان خود آسیب برساند، گاهی هم بصورت مسالمت‌آمیز زندگی می‌کنند. به هر حال خاصیت انگلی این است که بیماری‌زایی داشته باشد و خاصیت میزبان هم دفاع از خود است. از جمله خواص انگل، عفونی کردن، تهاجم، آسیب‌زایی و سم‌زایی است. اگر انگل بتواند به میزبان خود آسیب کافی برساند تغییراتی در بدن میزبان بوجود می‌آید که تظاهرات آن به صورت بیماری خودنمایی می‌کند.

---

<sup>1</sup>-Parasit

در مورد خواستگاه انگلها از اجداد آزادزی، همه ی نظریات توافق دارند که انگلها به طور اتفاقی وارد بدن میزبان شده و پس از انجام موتاسیونهای مختلف، برای محیط جدید سازش یافته و این سازشها یکباره حاصل شده و به تطابق اولیه و در نهایت به تغییر مسیر زندگی آنها منجر شده است [۳].



شکل (۱-۱)- تصویری از یک انگل یا پارازیت [۳]

#### ۱-۱-۱-۱- شرایط محیطی

مهمترین محیط های مناسب انگل در بدن مهره داران، دستگاه گوارشی، خون، ریه ها، حفره عدسی و بافتها می باشد. احتمالاً مواد غذایی متنوع از جمله قندها، پروتئینها، چربیها و آنزیمها با pH مناسب و فقدان اکسیژن و وضع ساختمان اندامها در ایجاد شرایط مساعد نقش دارند. گاهی انگلها محیط بدن میزبان را به نفع خود تغییر می دهند، مثلاً آلودگی به سستدها<sup>۱</sup> باعث کاهش pH اسیدی روده می شود. از محیط های مناسب دیگر برای انگلها، خون و روده است [۳].

<sup>1</sup>-Cestod

## ۱-۱-۲- انتقال عامل بیماریزای عفونی

آشنایی با روشهای انتقال بیماریهای عفونی برای مقابله با آنها حائز اهمیت است. در صورت آشنایی با راههای انتقال بیماری می توان اکثر بیماریهای عفونی را ریشه کن کرد که این عمل با ایجاد وقفه در گردش عامل بیماریزا در طبیعت بوجود می آید. میکرووبهای بیماریزا به چهار روش عمده انتقال می یابند:

۱- انتقال از طریق تماس مستقیم

۲- انتقال بوسیله ی آب و مواد غذایی

۳- انتقال با تلقیح مستقیم (نیش حشرات)

۴- انتقال از طریق هوا

عموماً انگلها از طریق پوست یا دهان وارد بدن شخص می شوند. به طور تخمینی، ۲۴۰ بیماری از حیوانات به انسانها از طریق انگل می تواند منتشر شود.

## ۱-۱-۳- علائم

علائم بیماریهای انگلی ممکن است همیشه مشاهده نشود. در واقع ، بعضی علائم شامل کمخونی و کمبود هورمون می شود. بعضی علائم ایجاد شده بوسیله ی هجوم کرمهای مختلف شامل خارش مقعد یا سطح تناسلی، درد شکمی، کاهش وزن ، افزایش اشتها، انسداد روده، اسهال و استفراغ که منجر به آبدایی می شود، مشکلات خواب، حضور کرمها در اسهال، بیقراری، آلرژی، خستگی و عصبیت می باشد [۲].

## ۱-۱-۴- درمان

داروهای زیادی برای درمان عفونتهای انگلی بکار برده می‌شود که خودشان عامل بالقوه‌ای در تغییر ساختار میزبان هستند. این داروها ابتدا در آزمایشگاه بر روی موشها تست می‌شوند. فقط موشهای فاقد عفونتهای انگلی، کاملاً برای کارهای تحقیقاتی مناسب هستند. از میان آنها، تیبندازول<sup>۱</sup> غیرمستوفکننده‌ی ایمنی و غیرآتشگیر در موشهاست. لوامیزول<sup>۲</sup> بعنوان یک کرم‌کش قوی هم در انسان و هم در طب دامپزشکی بکار می‌رود. فنبندازول<sup>۳</sup> نیز کاربردهای زیادی دارد. آلبندازول<sup>۴</sup> و بنزیمیدازول<sup>۵</sup>های دیگر با خواص کرم‌کشی و همچنین ایورمکتین<sup>۶</sup> که یک لاکتون دی‌ساکارید ماکروحلقه و عضوی از خانواده‌ی کرم‌کش آورمکتین<sup>۷</sup> است نیز از داروهای مفید می‌باشند [۴].

برای کاهش شدت عفونتهای کرمی، مدیریت تغذیه به تنهایی یا به همراه داروهای کرم‌کش، می‌تواند مفید باشد، اگر چه طرحهای مدیریت تغذیه اغلب عمل‌نشدنی است. برای درمان این عفونتها، واکسنهایی نیز تولید شدند، اما در بسیاری از کشورها فقدان زیربنای مناسب، گرانی واکسنها، مشکوک بودن ارزش آنها در افراد دچار سوء تغذیه از کاربرد وسیعشان جلوگیری می‌کند [۵].

در این پروژه، تحقیقاتی روی داروهای دارای اثر بازدارندگی روی دو انگل تریکوموناس واژینالیس<sup>۸</sup> و گیاردیا اینتستینالیس<sup>۹</sup> نسبت به میزبان به عمل آمده است که ابتدا شرحی از این دو انگل، بیماری، علل و راههای درمان آنها و پس از آن نتایج تحقیقات بیان می‌شود.

---

1-Thiabendazole

2-Levamisole

3-Fenbendazole

4-Albendazole

5-Benzimidazole

6-Ivermectin

7-Avermectin

8-Trichomonas vaginalis

9-Giardia Intestinalis

## ۱-۱-۵- گیاردیاسیس<sup>۱</sup>

بیشتر از ۲ میلیون نفر از مردم جهان با انگل روده ای عفونی شده‌اند. در میان این افراد، گیاردیا<sup>۲</sup>، عامل مسبب بیماری گیاردیاسیس، اساساً جمعیت کودکان را تحت تأثیر قرار می‌دهد و در آب و هوای گرم بیشتر رؤیت می‌شود. این بیماری باعث کاهش رشد و کاهش پیشرفت ذهنی کودکان می‌شود. انگل گیاردیا سرعت شیوع بسیار بالایی در میان بچه‌های کوچکتر در کشورهای جهان سوم دارد و تخمهای آن بسیار عفونی هستند و فقط ۱۰ تخم آن می‌تواند باعث عفونت در یک شخص شود. این بیماری از راه آب و غذای آلوده و تماس جنسی منتقل می‌شود. انگل گیاردیا علاوه بر انسان، مرغ، گاو، گوسفند، آهو، سگ و گربه را نیز عفونی می‌کند [۶].

## ۱-۱-۵- تاریخچه ی گیاردیا

گیاردیا یک تک‌یاخته‌ی گلایی شکل تاژکدار است که در دو حالت تخم و تروفوزوئیت<sup>۳</sup> یافت می‌شود و  $10-20\mu\text{m}$  طول دارد. بدن متوسط و صفحه‌های مکند از خصوصیات دیگر گیاردیا می‌باشند. گیاردیا اینتستینالیس<sup>۴</sup> اولین بار توسط یکمرد کاسب هلندی، آنتونی وان لی ون هوک<sup>۵</sup> (که به عنوان پدر علم میکروب شناسی شناخته شده است)، در سال ۱۶۸۱ هنگامیکه مدفوع اسهالی خودش را زیر میکروسکوپش آزمایش می‌کرد مشاهده شد. وان سپس بین این جانور مشاهده شده و علم پاتولوژی ارتباط برقرار کرد.

۲۰۰ سال بعد از آنتونی وان، در سال ۱۸۵۹ یک پزشک اهل چکوسلواکی به نام ویلم لامبل<sup>۶</sup> و

همچنین در سال ۱۸۹۵ آلفرد گیارد<sup>۷</sup>، این جانور را مشاهده و توصیف کردند. این گونه‌ها بعدها به نام

<sup>1</sup>-Giardiasis

<sup>2</sup>-Giardia

<sup>3</sup>-Terophozoi

<sup>4</sup>-Giardia Intestinalis

<sup>5</sup>-Antony van Leeuwenhoek

<sup>6</sup>-Vilem Lambl

<sup>7</sup>-Alfred Giard

لامبل، گiardia لامبلیا<sup>۱</sup>، نامیده شدند. بنابراین گiardia اینتستینالیس اولین انگل تک‌یاخته بود که توصیف شد و نقش آن بعنوان یک ارگانیزم بیماریزا تا سال ۱۹۷۰ تصدیق نشد. قبل از آن زمان، تصور می‌شد که این ارگانیزم، یک ارگانیزم بی‌ضرر هم‌غذای روده است [۷].

### ۱-۱-۵-۲- سیکل زندگی گiardia

چرخه‌ی زندگی انگل گiardia دارای مراحل زیر است :

مرحله‌ی ۱ و ۲ - تخمهای گiardia بوسیله مصرف غذا یا آب آلوده، تماس جنسی و... بلعیده می‌شوند و می‌توانند چندین ماه داخل بدن شخص زنده بمانند و همچنین نسبت به کلرینه شدن، تابش UV و انجماد مقاوم هستند.

مرحله‌ی ۳ - هنگامی که تخمها بلعیده شدند، کاهش pH اسید معده تولید اگزیستیشن<sup>۲</sup> می‌کند که مرحله خارج شدن از تخم نامیده می‌شود و تروفوزوئیتها را منتشر می‌کند و با هر تخم دو تروفوزوئیت منتشر می‌شود.

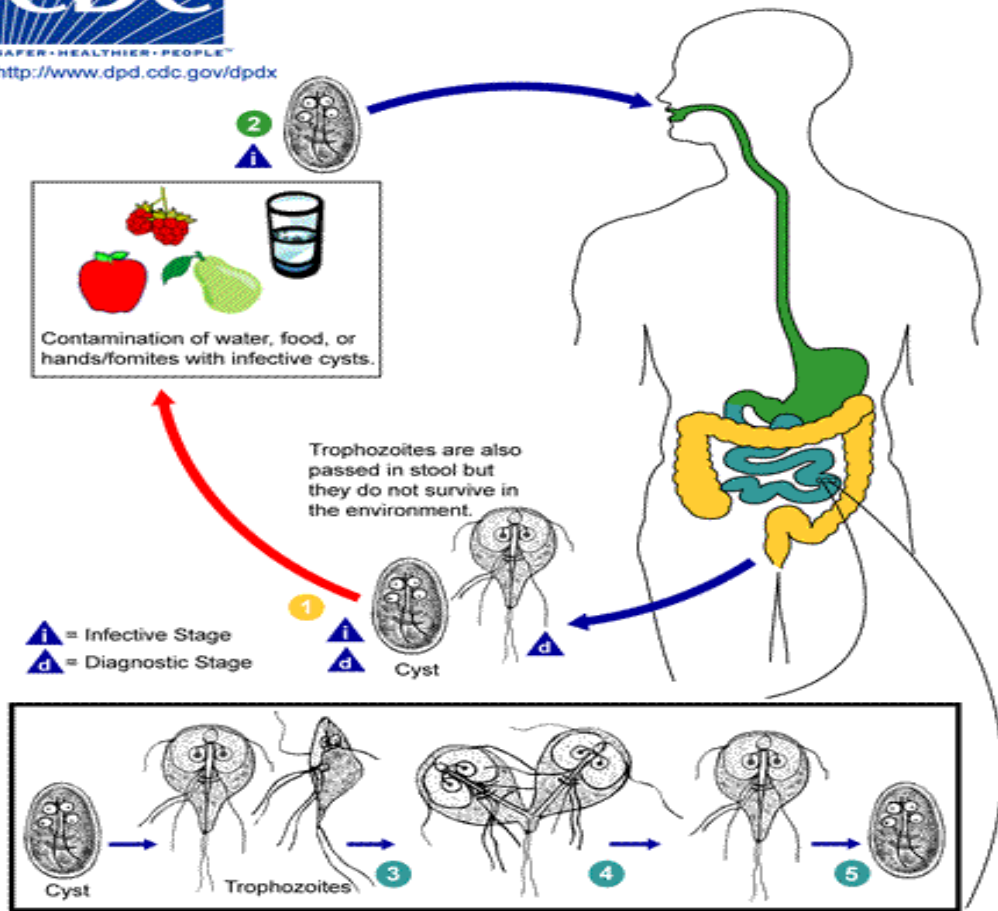
مرحله‌ی ۴ - داخل روده‌ی کوچک، تروفوزوئیتها هر ۵ ساعت بطور غیر جنسی تکثیر می‌شوند و هر یک یا آزادانه شناور شده یا با مخاط حفره همراه می‌شوند، این انگل به موکوس که می‌چسبد ماده‌ی مغذی که از دیواره‌ی روده می‌آید را جذب کرده و ایجاد اسهال را به دنبال می‌آورد.

مرحله‌ی ۵ - بعضی تروفوزوئیتها در روده‌ی کوچک در کیسه‌ای قرار می‌گیرند. این مرحله که ان-سیستیشن<sup>۳</sup> یا داخل تخم شدن نامیده می‌شود، اغلب اوقات در نتیجه‌ی مواجه شدن با نمکهای صفرای اسیدهای چرب در یک محیط بسیار قلیایی رخ می‌دهد. سپس هم تخمها و هم تروفوزوئیتها از طریق مدفوع خارج شده و فوراً یا کمی پس از آن عفونی می‌شوند و می‌توانند به شخص دیگری منتقل شوند (شکل ۱-۲) [۸].

<sup>1</sup>-Giardia Lamblia

<sup>2</sup>-Excystation

<sup>3</sup>-Encystation



شکل (۱-۲) - سیکل زندگی گیاردیا اینتستینالیس [۸]

### ۱-۱-۵-۳- علائم و تشخیص

حدود  $\frac{2}{3}$  افراد عفونی شده با گیاردیاسیس بدون علامت باقی می‌مانند. هنگامیکه نشانه‌ها رخ می‌دهند، بیماری با یک دوره‌ی نهفتگی ۳ تا ۲۵ روز دنبال می‌شود. تعدادی از نشانه‌های معمول گیاردیاسیس عبارتند از: نفخ کردن، تنفس مشکل، دفع آب، اسهال، خستگی، تب، کاهش اشتها، حالت تهوع، کاهش وزن.

در گذشته، تشخیص گیاردیا بوسیله‌ی آشکارسازی مستقیم تروفوزوئیتها یا تخم‌ها در یک نمونه مدفوع بوسیله‌ی آزمایش انگل انجام می‌شد. در سالهای اخیر، تستهای تشخیص سریع که روشهای



آشکارسازی آنتی ژن را بکار می‌برند بسیار بکار برده شده‌اند. این تستها در مقایسه با آزمایش انگل هم بسیار ویژه و هم حساس هستند [۹].

## ۱-۱-۵-۴-درمان

اولین داروهای انتخابی در درمان گیاردیاسیس، مترونیدازول و تینیدازول<sup>۱</sup> هستند، اما مترونیدازول که یک نیتروایمیدازول است، دارای چندین اثر سمی جانبی ناخوشایند مثل سوزش بلغم شکمی و تخریب اسپرما می‌باشد و نیز مقاومت دارو به انگل، باعث محدود شدن کاربرد آن در درمان گیاردیاسیس شده است. به خاطر وجود این مشکلات، نیاز مبرمی در تهیه‌ی داروهای جدید مؤثر ایجاد شد و داروهای جدیدی برای درمان گیاردیاسیس تولید شدند [۱۰]:

- نیتازوکسانید<sup>۲</sup>: که از رشد تروفوزوئیتها از طریق قطع متابولیسم انرژی‌شان جلوگیری می‌کند و به فرم مایع برای کودکان در دسترس است.
- پارمومایسین<sup>۳</sup>: برای عفونتهای متنوع در زنان باردار بکار می‌رود اما اثرات جانبی بدتری دارد و کاربرد طولانی مدت آن توصیه نمی‌شود.
- فورازولیدون<sup>۴</sup>: به فرم مایع در دسترس است، نسبت به دیگر داروهای درمانی کمتر مؤثر است.
- همچنين آلبندازول<sup>۵</sup> و مبندازول<sup>۶</sup> داروهای مؤثری در درمان هستند.

---

<sup>1</sup>-Tinidazole

<sup>2</sup>-Nitazoxanid

<sup>3</sup>-Parmomycin

<sup>4</sup>-Furazolidone

<sup>5</sup>-Albendazole

<sup>6</sup>-Mebendazol

## ۱-۱-۶- تریکومونیا سیس<sup>۱</sup>

تریکومونیا سیس عفونت انتقالی جنسی (STI<sup>۲</sup>) است که توسط انگل تریکوموناس واژینالیس<sup>۳</sup> (T.V) ایجاد می‌شود و پس از عفونتهای ویروسی، شایع ترین بیماری منتقله از راه جنسی می‌باشد. در سال ۱۹۹۸، پنج میلیون مورد از این بیماری مشاهده شد. این بیماری فراگیر است و حالتهای آن اغلب بدون نشانه است. بیماری تریکومونیا سیس در هر دو جنس زن و مرد دیده می‌شود اما علائم آن در زنان شایعتر است و در زنان جوان فعال از نظر جنسی قابل درمان می‌باشد. شایعترین محل عفونت در زنان، واژن و در مردان، مجرای ادراری می‌باشد [۱۱].

انگل تریکوموناس، تک یاخته‌ای است که فقط به یک فرم دیده می‌شود و آن هم فرم تروفوزوئیتی است. این انگل، گلابی شکل و قسمت قدامی آن پهن است، دارای دهان سلولی و جسم اکسوتیل<sup>۴</sup> است، به طوریکه قسمتی از آن از انتهای انگل خارج می‌شود. همچنین دارای هسته با کاریوزوم<sup>۵</sup> بزرگ می‌باشد. در قسمت قدام دارای ۴ تاژک آزاد و تاژک پنجم چسبیده به دم تک‌یاخته است بطوریکه  $\frac{2}{3}$  طول بدن تک‌یاخته را طی کرده و ایجاد پرده‌ی موج می‌کند [۱۱].

انسان تنها مخزن شناخته شده‌ی این تک‌یاخته است. در کشورهای توسعه یافته بیش از ۵۰ درصد بیماران مراجعه کننده به کلینیک بیماریهای مقاربتی به تریکومونیا سیس مبتلا هستند [۱۱].

## ۱-۱-۶-۱- تمایز تریکوموناس واژینالیس با دیگر تک‌یاختگان

همانطور که اثبات شده است وظایف اصلی میتوکندری، تولید انرژی (ATP) و آزاد کردن پروتئین‌هایی مثل سیتوکروم C<sup>۶</sup> از فضای بین غشایی میتوکندری به سیتوسل<sup>۷</sup> است.

1-Trichomoniasis

2-Sexually transmitted infection

3-Trichomonas Vaginalis

4-Axotil object

5-Karyosome

6-Cytochrome

7-Cytosol

یک تغییر نژاد باعث تمایز یوکاریوتها شد که این نژادهای ایجاد شده، قارچها، آمیبها و تاژکدارها را شامل می شود که فاقد ارگانهای یوکاریوتی مثل میتوکندری و پراکسی زوم هستند. این ارگانیسمها که تکیاختگان آغازی فاقد میتوکندری نامیده می شوند به دو نوع متمایز تقسیم می شوند:

نوع ۱- که گیاردیا اینتستینالیس به آن تعلق دارد و فاقد هر دو هیدروژنوزوم و میتوکندری است.

نوع ۲- که تریکوموناس واژینالیس به آن تعلق دارد و با حضور هیدروژنوزوم اختصاصی می شود.

در T.V دانه های ریزی به نام هیدروژنوزوم ها وجود دارد که تولید  $H_2$  می کنند که محصول سوخت و ساز و غیرمعمول در میان یوکاریوتهاست. حضور آنزیمهایی مثل فرودوکسین، اکسیدوردوکتاز و هیدروژناز که در میتوکندری یافت نمی شوند و فقدان وظایف میتوکندری مثل اکسیداسیون و فسفریلاسیون خصیصه اصلی این دانه های ریز منحصر به فرد است [۱۲].

### ۱-۶-۲- بیماری زایی تریکوموناس واژینالیس

بیماری زایی این انگل به این شکل است که انگل در محیط اسیدی با  $pH = 5/5$  شروع به تکثیر و رشد می کند. لذا این انگل در محیط های با  $pH = 4-4/5$  تکثیر پیدا نمی کند. اما چنانچه به دلایلی تغییری در  $pH$  محیط ایجاد شود و به طرف  $5/5$  برود محیط برای تکثیر انگل مناسب می شود. از دلایل دیگر رشد انگل بویژه در خانمهای بالغ این است که در دیواره سلولهای مخاط واژن گلیکوژن بوجود می آید که یکی از مواد غذایی مهم مورد نیاز انگل است. این انگل در دختران نابالغ و خانمهای یائسه بوجود نمی آید، زیرا در این افراد  $pH$  در حد خنثی یعنی ۷ است و همچنین سلولهای مخاط واژن فاقد گلیکوژن هستند که ماده غذایی مناسب برای انگل است.

## ۱-۱-۶-۳- علائم و عوارض

اکثر مردان مبتلا به این انگل بدون علامت هستند، اگر چه بعضی مردان گهگاهی ترشح رقیق و سفید از آلت، درد یا احساس سوزش هنگام دفع ادرار و التهاب پوست ختنه‌گاه را تجربه می‌کنند. علائم عفونت در زنان شامل ترشح کف آلود سبز یا زرد رنگ بدبو، سوزش و التهاب و درد در اطراف مهبل هنگام مقاربت و ادرار و در موارد نادر، درد در ناحیه ی زیر شکم می‌باشد. علائم معمولاً در زنان بین ۲۸-۵ روز بعد از مواجهه با انگل رخ می‌دهد. ابتلا به این عفونت در زنان باردار منجر به زایمان زودرس و تولد نوزادان با وزن کم و سقط جنین می‌شود. امروزه اعتقاد بر این است که T.V با از بین بردن پروتئاز ترشعی گلبولهای سفید، که آنرا مسئول حفاظت سلولهای غشایی موکوس واژن در برابر ویروس HIV<sup>۱</sup> می‌دانند، باعث افزایش خطر ابتلا به بیماری ایدز<sup>۲</sup> می‌گردد.

## ۱-۱-۶-۴- درمان

تحقیقات نشان داده است که گیاهان دارویی به دلیل سازگاری بیشتر با بدن و قیمت کمتر و توصیه طب سنتی در درمان تریکومونیاسیس کاربرد زیادی دارند. از جمله این گیاهان دارویی می‌توان از اسطوخودوس، سیر و آنغوزه نام برد [۱۳]. تریکومونیاسیس با تک دوز قرص مترونیدازول قابل درمان است. ولی گزارشات عدیده‌ای از بسیاری از کشورها مبنی بر شیوع مقاومت به داروهای مترونیدازول و عوارض کارسینوژنیک و تراژوژنیک آن (خصوصاً در سه ماه اول بارداری) محققان را بر آن داشت تا به جستجوی داروهای جایگزین با عوارض جانبی کمتر بپردازند، که تینیدازول و مبندازول از این داروهای جدید می‌باشند. مبندازول باعث بی‌حرکتی و مرگ کرم‌ها توسط مهار انتخابی و غیر قابل برگشت جذب گلوکز در آنها می‌شود. مبندازول جذب روده‌ای خوبی ندارد و پس از مصرف، مقدار زیادی از آن بدون تغییر در مدفوع یافت می‌شود. دارو اثر درمانی خود را به آهستگی

1- Human Immunodeficiency Virus

2-AIDS

(غالباً ظرف سه روز) نشان می‌دهد. علائم تریکوموناس در مردان حتی بدون درمان در عرض چند هفته رفع می‌شود، اگر چه انتقال عفونت به همسر در اثر شخص کماکان صورت می‌گیرد. به همین علت زن و شوهر بایستی با هم درمان شوند تا عفونت ریشه کن شود. افرادی که تحت درمان قرار می‌گیرند تا زمان بهبودی کامل و رفع علائم در خود و همسرشان بایستی از مقاربت خودداری کنند.

## ۱-۲- پیشینه‌ی کارهای انجام شده

یووانی ماررو<sup>۱</sup> و همکارانش در سال ۲۰۰۵، از آنالیز تبعیض خطی<sup>۲</sup> (LDA) برای طبقه‌بندی مولکولها با در نظر گرفتن فعالیت ضد انگلی‌شان استفاده کردند. در این مطالعه‌ی کمی ساختار - فعالیت<sup>۳</sup>، مجموعه داده‌های ۱۰۷ عضوی با تنوع مولکولی زیاد بکار برده شده است و داده‌ها به دو سری تست و آموزش تقسیم شده‌اند که سری تست دارای ۱۷ و سری آموزش دارای ۹۰ عضو است. در این مطالعه، کیفیت مدل‌ها با بررسی پارامتر لاندای ویلکز،  $D_2$  و  $F$  بررسی شده است [۱۴].

یووانی ماررو و همکارانش در سال ۲۰۰۶ در ادامه‌ی کار قبلی خود برای کشف مواد شیمیایی ضد تریکومونیاسی غیرسمی و قویتر، یک توصیفگر مولکولی جدید و آنالیز تبعیض خطی (LDA) را گزارش نموده‌اند. در این آنالیز، ضریب تعیین ۰/۸۹ و ۰/۷۹ به ترتیب برای سری آموزش و تست گزارش شده است. با استفاده از تست ارزیابی تقاطعی، قدرت پیشگویی بالاتری برای مدل‌های بدست آمده حاصل شده است و علاوه بر آن، با آنالیز رگرسیون استاندارد<sup>۴</sup>، کیفیت آماری این مدل‌ها را تقویت کرده‌اند. آنتونیو و همکارانش در این تحقیق، ۵ ترکیب شیمیایی جدید با خصلت درمانی قوی ضد تریکومونیاسی را نشان داده‌اند و قدرت پیشگویی معادلات QSPR/QSAR را با تکنیکهای ارزیابی

---

<sup>1</sup>-Yovani Marrero-Ponc

<sup>2</sup>-Linear Discrimination Analysis

<sup>3</sup>-Quantitative structure-Activity relationship(QSAR)

<sup>4</sup>-Canonical regression analysis

داخلی (حذف گروهی<sup>۱</sup>(LGO)) و خارجی (کاربرد یک سری تست و یک سری پیشگویی خارجی) بررسی کرده‌اند [۱۵].

هاکان جوکر<sup>۲</sup> و همکارانش در سال ۲۰۰۸ مطالعات QSAR را بوسیله‌ی آنالیز رگرسیون چندگانه برای ۳۲ ترکیب فعال آنیلینوبنزیمیدازول انجام داده‌اند. در این آنالیز ضمن دستیابی به معادله‌ی خط، به نتایج  $r^2=0/90$  و  $s=0/281$  و  $F=65/52$  برای این داروهای ضد استافیلوکوکوس اورئاز دست یافته‌اند. این آنالیز به منظور حصول رابطه‌ی بین فعالیت آنتی استافیلوکوکوسی و مقادیرهای  $\log k_w$  برون‌یابی شده انجام شده است که  $k_w$  همان فاکتور ظرفیت می‌باشد. هاکان در کار خود از چربی‌دوستی عوامل شیمیایی و برای توصیف آن، از لگاریتم ضریب تقسیم آب- اکتانول ( $\log p$ ) استفاده کرده است [۱۶].

هونگ گونگ<sup>۳</sup> و همکارانش در سال ۲۰۱۰ بمنظور مطالعه‌ی کمی رابطه‌ی ساختار- فعالیت مجموعه‌ی ناهمگنی از عوامل آنتی باکتریال در مقابل کاندیدا آلبیکنز<sup>۴</sup> روش‌درخت‌های رگرسیون و طبقه بندی<sup>۵</sup> و رگرسیون خطی چندگانه را در مقایسه با هم بکار برده است. نتایج آزمایشگاهی، همبستگی بالایی بین ساختار مجموعه و شدت آنتی باکتریایی بر ضد کاندیدا بوسیله‌ی روش CART نشان داده‌اند. جیان در این آنالیز مجموعه‌ی داده‌ها را به دو سری آموزش و تست تقسیم کرده است. روش MLR ضرایب تعیین  $0/88$  و  $0/89$  به ترتیب برای سری آموزش و تست نتیجه داده است در حالیکه روش CART ضرایب تعیین  $0/98$  و  $0/92$  به ترتیب برای سری آموزش و تست نتیجه داده است. جیان بعد از مقایسه‌ی این دو روش، به برتری روش CART نسبت به MLR دست یافته است [۱۷].

---

<sup>1</sup>-Leave- group- out

<sup>2</sup>-Hakan Goker

<sup>3</sup>-Hong gong

<sup>4</sup>-Candida Albicans

<sup>5</sup>-Classification and Regression trees(CART)

در تحقیق حاضر، ارتباط کمی ساختار-فعالیت (QSAR) برای پیش بینی فعالیت دارویی یکسری  
از مشتقات بنزیمیدازول به عنوان ترکیبات ضد انگلی بررسی شده است. بر اساس دانش ما و مطالعات  
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  
چندگانه 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.

## ارتباط کمی ساختار- فعالیت



## ۲-۱-ارتباط کمی ساختار- فعالیت

از نظر شیمیدانان، فعالیتها و خواص یک ترکیب ناشی از ویژگیهای ساختاری آن است. هر گاه مطالعات بصورت ارتباط بین ساختار مولکولی و خواص مشاهده شدهی مولکول انجام گیرد به آن ارتباط کمی ساختار- خاصیت (QSPR)<sup>۱</sup> گویند مانند یافتن رابطه‌ای بین خواصی نظیر نقطه‌ی جوش و ساختار مولکولی. اما وقتی خواصی از نوع بیولوژیکی (مانند فعالیت دارو) توصیف می‌شود، اشاره به ارتباط کمی ساختار- فعالیت (QSAR) خواهد داشت [۱۸].

کورن هانش<sup>۲</sup> برای اولین بار در سال ۱۹۶۴ مفهوم رابطه‌ی کمی ساختار- فعالیت (QSAR) را مطرح کرد. او نشان داد که توصیفگرهای کمی مولکولی در ترکیبات دارویی با ویژگیهای بیولوژیک این ترکیبات ارتباط تنگاتنگی دارند [۱۹]. در حال حاضر تغییر و اصلاح ساختار یک ترکیب، عمده‌ترین روش در دستیابی به داروهای جدید می‌باشد. تغییرات ساختاری منجر به پیدایش ویژگیهای جدید می‌شود که در نهایت موجب تغییر پخش دارو، تغییر در کیفیت اثر متقابل با جایگاههای فعال آنزیم و گیرنده‌ها و در نهایت تغییر سرعت واکنشها در جایگاههای مورد نظر می‌گردد [۲۰].

در سال ۱۹۶۴ هانش و همکارانش بین فعالیت رشد گیاهی مشتقات فنوکسی استیک اسید و چگالی الکترونی در موقعیت اورتو، ارتباطی مثبت یافتند و برای اولین بار روش رگرسیون خطی چندگانه را برای بدست آوردن ارتباط کمی ساختار- فعالیت بکار بردند [۲۰].

## ۲-۲- مراحل مطالعه ارتباط کمی ساختار – فعالیت (QSAR)

همانطور که در شکل (۲-۱) نشان داده شده است، در مطالعات QSAR برای بیان ارتباط ساختار مولکول با فعالیت بیولوژیکی آن، باید مراحل زیر به ترتیب طی شوند [۲۱ و ۲۲]:

۱- فراهم کردن سری داده‌ها

<sup>1</sup>-Quantitative Structure-Property Relationship

<sup>2</sup>-Corren Hansch

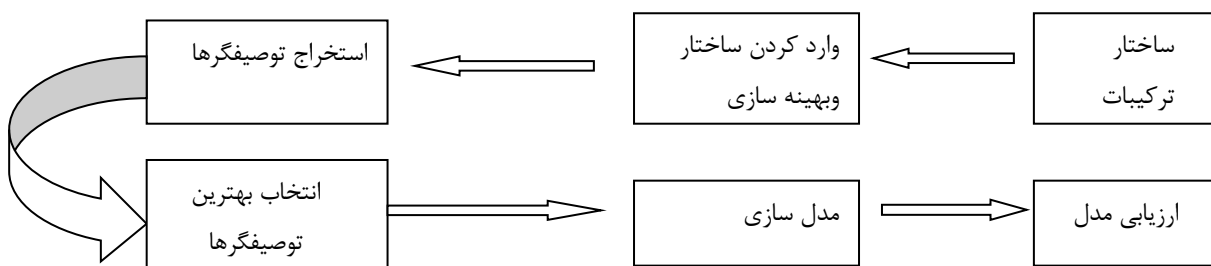
۲- رسم و بهینه‌سازی ساختار هندسی

۳- استخراج توصیف‌کننده‌ها یا متغیرهای مستقل

۴- انتخاب بهترین توصیف‌کننده‌ها برای ورود به مدل

۵- ساختن مدل مناسب برای پیش‌بینی متغیر وابسته

۶- ارزیابی قدرت پیش‌گویی مدل



شکل (۱-۲) - مراحل QSAR

## ۲-۲-۱- فراهم کردن سری داده‌ها

در مطالعات QSAR اولین مرحله، جمع‌آوری فهرستی از ترکیبات است که کمیتهای تجربی خاصی برای آنها توسط محققین گزارش شده است. عموماً جمع‌آوری مولکولهای سری داده‌ها از طریق جستجو در مقالات و مجلات علمی صورت می‌گیرد. از میان مجموعه مولکولهای جمع‌آوری شده، تعدادی که داده‌های گزارش شده در مورد آنها در شرایط عملی یکسانی بدست آمده باشند به عنوان سری داده‌ها انتخاب می‌شوند. برای انتخاب QSAR به عنوان روش مطلوب، بهتر است تنوع گونه‌ها زیاد باشد و ترکیبات انتخابی از نظر ساختاری تا حدود زیادی به مولکول اصلی شباهت داشته باشند. هر چه سری داده‌ها بزرگتر و متنوع‌تر باشد مدل حاصل از آن معتبرتر بوده و قدرت پیش‌بینی بالاتری خواهد داشت.

## ۲-۲-۲- رسم و بهینه سازی ساختار هندسی

مرحله‌ی بعدی در مطالعات QSAR، فراهم کردن ساختار هندسی مولکولها و بهینه سازی آنها می‌باشد، بنابراین باید ساختار هندسی ترکیب بهینه<sup>۱</sup> شود و مختصاتی از ساختار ترکیب بدست آید که در آن مختصات، انرژی مولکول حداقل باشد. به این منظور از شیمی محاسباتی<sup>۲</sup> استفاده می‌شود. شیمی محاسباتی ساختارها و واکنشهای شیمیایی را شبیه سازی می‌کند و این امر به شیمیدانان اجازه می‌دهد که به مطالعه‌ی روشهای شیمیایی با اجرای محاسبات توسط کامپیوتر بپردازند، بدون اینکه از روشهای تجربی و آزمایشگاهی استفاده کنند [۱۸ و ۲۳]. روشهای محاسباتی در شیمی به چند دسته تقسیم می‌شوند:

۱- روشهای مکانیک کوانتومی، که خود نیز به سه دسته تقسیم می‌شود:

(a) روشهای آغازین<sup>۳</sup>

(b) روشهای نیمه تجربی<sup>۴</sup>

(c) روشهای تابعی چگالی<sup>۵</sup>

۲- روشهای مکانیک مولکولی

در روشهای آغازین از هیچ داده‌ی تجربی استفاده نمی‌شود و فقط ثابتهای فیزیکی مانند جرم الکترون، بار الکترون و ثابت پلانک استفاده می‌شود و تمام پارامترهای ساختاری از حل دقیق معادله شرودینگر با یکسری تقریب بدست می‌آیند. این روش چون کل الکترونها را در محاسبات وارد می‌کند، بیشتر منحصر به مولکولهای کوچک است و برای مولکولهای بزرگتر، محاسبات پیچیده تر و نیازمند ابر رایانه است [۲۴ و ۲۵].

---

1-Optimize

2-Computational Chemistry

3-Ab Initio Methods

4-Semi-empirical methods

5-Density functional theory(DFT)

روشهای نیمه تجربی مانند روشهایی که در برنامه هایی نظیر MOPAC و Hyperchem وارد شده- اند و محاسبات در این روشها بر پایه‌ی مکانیک کوانتومی است. در این روشها فقط الکترونهای لایه‌ی ظرفیت در محاسبه در نظر گرفته می‌شود. روشهای نیمه تجربی مختلفی مانند صرف نظر کردن کامل از انتگرالهای همپوشانی دیفرانسیلی (CNDO<sup>1</sup>)، صرف نظر از برخی انتگرال های همپوشانی در الکترونها (INDO<sup>2</sup>)، صرف نظر اصلاح شده از انتگرال همپوشانی دواتمی (MNDO<sup>3</sup>) ارائه شده است [۱۸]. روش AM1<sup>4</sup> همان روش MNDO اصلاح شده است که جزء دقیق ترین روشهاست و در این پروژه از آن استفاده شده است. در روشهای تابعی چگالی، بجای استفاده از تابع موج از دانسیته‌ی الکترونی استفاده می‌شود. این روش حجم محاسباتی کمی دارد و برای کارهای کاربردی نظیر تعیین ساختار و بدست آوردن طیف های NMR و IR استفاده می‌شود [۲۴].

در روشهای مکانیک مولکولی هیچ ماهیت کوانتومی وجود ندارد و نسبت به روش مکانیک کوانتومی دارای صحت کمتری است، زیرا روش مکانیک کوانتومی اثرات الکترونی اعمال شده روی مولکول را نیز در محاسبات وارد می‌کند در حالیکه مکانیک مولکولی الکترونها را در نظر نمی‌گیرد.

## ۲-۲-۳- استخراج توصیف کننده‌ها

پس از بهینه‌سازی ساختار مولکولها بایستی توصیف کننده‌های وابسته به ساختار مولکول استخراج شوند. توصیف کننده‌های مولکولی مقادیر عددی هستند که ساختار یا شکل مولکولی را توصیف می‌کنند و به پیش بینی فعالیت و خصوصیات مولکولها در آزمایشهای پیچیده کمک می‌کنند. برخی ویژگیهای لازم برای یک توصیفگر مولکولی مناسب عبارتند از [۲۶]:

• ساده بودن

<sup>1</sup>-Complete Neglect of Diffrential Overlap

<sup>2</sup>-Intermediate Neglect of Diffrential Overlap

<sup>3</sup>-Modified Neglect of Diatomic Overlap

<sup>4</sup>-Austian Method 1

- توانایی در تفسیر ساختار مولکول
- عدم همبستگی با دیگر توصیفگرها
- قابل کاربرد برای دامنه وسیعی از ساختارهای مولکولی
- مستقل بودن
- تمایز بین ایزومرهای مختلف مولکول
- وابستگی صحیح به اندازه مولکول

توصیف کننده‌های مولکولی نتیجه نهایی یک فرآیند منطقی و ریاضی هستند که اطلاعات شیمیایی مربوط به ساختار یک مولکول را به اعداد تبدیل می‌کنند. این اعداد می‌توانند برای تفسیر خواص مولکولی استفاده شوند و یا برای پیش بینی تعدادی از ویژگیهای مولکولی در یک مدل شرکت کنند. توصیفگرهای مورد استفاده در مطالعات QSAR، بر اساس ماهیت به دو دسته توصیفگرهای تئوری و توصیفگرهای تجربی تقسیم می‌شوند. توصیفگرهای تجربی براساس آزمایش اندازه گیری می‌شوند و برای تعداد محدودی از ترکیبات شیمیایی قابل کاربرد هستند در حالی که توصیفگرهای تئوری بدون نیاز به داده‌های تجربی و به کمک نرم‌افزارهای مناسب برای طیف وسیعی از ترکیبات قابل محاسبه هستند و باعث صرفه جویی در وقت، هزینه، مواد و تجهیزات می‌شوند. بنابراین توصیفگرهای تئوری در مطالعات QSAR از اهمیت زیادی برخوردار هستند [۱۸]. توصیفگرها بر اساس ابعاد نیز به چهار دسته، توصیفگرهای صفر بعدی، یک بعدی، دو بعدی و سه بعدی تقسیم می‌شوند.

توصیفگرهای مولکولی بدست آمده از فرمول شیمیایی، توصیف کننده‌های صفر بعدی نامیده می‌شوند. این دسته از توصیف کننده‌ها تعداد و نوع اتمها، جرم مولکول و سایر خصوصیات اتمی را تعیین می‌کنند و به توصیف کننده‌های ساختاری نیز معروفند. تعداد یک عنصر شیمیایی خاص در مولکول، تعداد گروه‌های عاملی خاص در مولکول، تعداد کل حلقه‌ها و وزن مولکولی مثالهایی از توصیف کننده‌های ساختاری هستند.

نمایش فهرست ویژگیهای زیرساختاری<sup>۱</sup> می تواند بعنوان نمایشی یک بعدی از مولکول در نظر گرفته شود. توصیف کننده های گروه های عاملی، قطعات اتم مرکزی، توصیف کننده های تجربی و خصوصیات مولکولی زیرگروه های این دسته می باشند.

نمایش دوبعدی از مولکول، چگونگی اتصال بر حسب حضور و ماهیت پیوندهای شیمیایی می باشد. توصیفگرهای مولکولی بدست آمده از الگوریتم های بکاررفته برای نمایش توپولوژیکی، توصیف کننده های دو بعدی نامیده می شوند. این دسته از توصیف کننده ها شامل زیرگروه های توپولوژیکی، توصیف کننده های BCUT<sup>۲</sup>، شمارش های مسیرهای مولکولی<sup>۳</sup>، ضرایب بار<sup>۴</sup> و توصیف کننده های خود همبستگی دو بعدی<sup>۵</sup> می شود.

به نمایش سه بعدی یک مولکول که نه تنها یک نمایش از ماهیت و اتصال اتمها، بلکه نمایشی از صورت بندی فضایی مولکول نیز می باشد نمایش هندسی<sup>۶</sup> می گویند. توصیف کننده های مشتق شده از این نمایش، توصیف کننده های سه بعدی نامیده می شوند و زیرگروه های آن عبارتند از شاخص های آروماتیکی<sup>۷</sup>، خصوصیات مولکولی راندیک<sup>۸</sup> و توصیف کننده های 3D-MORSE<sup>۹</sup>.

## ۲-۲-۴- انتخاب بهترین توصیف کننده ها

یکی از مراحل مهم در مطالعات QSAR انتخاب توصیفگرهای مناسب می باشد، زیرا توصیفگرهای نامناسب و اضافی فرایند برازش و مدلسازی را طولانی می کنند. به همین منظور باید در انتخاب توصیفگرها دقت لازم به عمل آید و توصیفگرهایی که دارای همبستگی بیشتری با متغیر وابسته

---

<sup>1</sup>-Substructure List Representation

<sup>2</sup>-Burden-CAS-University of Texas eigenvalues

<sup>3</sup>-Molecular Walk Count

<sup>4</sup>-Charge indices

<sup>5</sup>-2D autocorrelation

<sup>6</sup>-Geometrical Representation

<sup>7</sup>-Aromaticity indices

<sup>8</sup>-Randic molecular properties

<sup>9</sup>-3-D MOlecular Representation of Structures based on Electron diffraction

هستند انتخاب شوند. یک روش برای انتخاب توصیفگرهای مناسب، روش برازش مرحله ای<sup>۱</sup> است که ضریب تعیین<sup>۲</sup> را بین توصیفگر و فعالیت بیولوژیکی محاسبه می کند. ضریب تعیین میزان نزدیکی اطلاعات بدست آمده به خط رگرسیون را نشان می دهد. مقدار ضریب تعیین بین ۰ و ۱ تغییر می کند که توصیفگری با ضریب تعیین ۱ می تواند خواص موردنظر را به درستی توصیف کند و توصیفگری با ضریب تعیین صفر، هیچ ارتباطی با خاصیت دارویی ندارد. در نتیجه توصیفگرهایی با ضرایب تعیین بزرگ در برازش منحنی برای ساختن مدل جهت پیش گویی خواص استفاده می شوند [۱۸]. در ادامه چند روش خطی انتخاب متغیرها به اختصار بیان می شوند.

#### ۲-۲-۴-۱- روش ورود اجباری<sup>۳</sup>

فرایندی از انتخاب متغیرهاست که طی آن تمام توصیفگرها در یک مرحله وارد می شوند. این روش، روش مطلوبی نیست زیرا اثرات مثبت و منفی متغیرهای مستقل برای ورود در مدل نادیده گرفته می شوند.

#### ۲-۲-۴-۲- روش جلوبرنده<sup>۴</sup>

فرایند مرحله ای در انتخاب متغیرهاست که طی آن متغیرها یکی پس از دیگری وارد مدل می شوند متغیری که دارای بالاترین همبستگی با متغیر وابسته باشد برای ورود به مدل انتخاب می شود. این روند تا زمانیکه دیگر متغیری وجود نداشته باشد که دارای شرط ورود باشد و باعث بهبود مدل گردد، ادامه دارد.

---

<sup>1</sup>-Stepwise Regression

<sup>2</sup>-Determination Coefficient

<sup>3</sup>-Enter

<sup>4</sup>-Forward

## ۲-۲-۴-۳-روش عقب برنده<sup>۱</sup>

در این روش تمام توصیفگرها همزمان در ساخت مدل بکار می روند و سپس تغییری که کمترین همبستگی را با متغیر وابسته دارد حذف می شود. پس از حذف اولین متغیر، از بین متغیرهای باقیمانده در معادله تغییری که کوچکترین همبستگی را با متغیر وابسته دارد برای حذف در نظر گرفته می شود. این روند زمانی متوقف می شود که حذف تغییری که از شرط فرایند پیروی کند از ارزش مدل بکاهد.

## ۲-۲-۴-۴-روش مرحله ای<sup>۲</sup>

این روش مشابه روش انتخاب جلوبرنده است با این تفاوت که در این روش در هر مرحله برای متغیرهای وارد شده در مدل، آماره  $F$  (نسبت فیشر) محاسبه می شود و متغیر دارای کوچکترین آماره  $F$  حذف و مدل مناسب حاصل خواهد شد [۲۷].

## ۲-۲-۵-ساختن مدل مناسب

پس از ارزیابی توصیف کننده ها و انتخاب مناسب ترین آنها، هدف ساختن مدلی است که قادر باشد با استفاده از توصیفگرهای مولکولی، فعالیت موردنظر را پیش بینی کند. مدل یک رابطه ریاضی است که بیانگر رابطه بین متغیر وابسته و متغیر(های) مستقل می باشد و با روشهای آماری گوناگون می توان آن را بدست آورد. مدل های رگرسیون که شامل تعداد کمی از متغیرهای مستقل هستند بر سایر مدل ها برتری دارند، زیرا این مدل ها می توانند به آسانی تفسیر شوند و قدرت پیش بینی بالایی دارند.

---

<sup>1</sup>-Backward

<sup>2</sup>-Stepwise



علاوه بر رگرسیون خطی چندگانه (MLR)، آنالیز اجزای اصلی (PCA)<sup>۱</sup> و حداقل مربعات جزئی (PLS)<sup>۲</sup> نیز به عنوان روشهایی خطی و شبکه عصبی مصنوعی (ANN)<sup>۳</sup> به عنوان روش غیر خطی برای ساختن مدل QSAR استفاده می شوند.

## ۲-۲-۶- مدلسازی به کمک روش رگرسیون خطی چندگانه

رگرسیون خطی چندگانه راهی برای یافتن رابطه‌ی خطی میان چند متغیر مستقل و یک متغیر وابسته است. این روش تنها زمانی می تواند به کار رود که تعداد مولکولهای مورد بررسی بیشتر از تعداد توصیف‌کننده‌ها باشد و متغیرها نیز همبستگی زیادی با هم نداشته باشند. متغیر مستقل متغیری است که می توان آنرا بطور مستقل تغییر داد و مقدار آن توسط محقق تعیین می‌شود. متغیر وابسته (متغیر پاسخ) نیز متغیری است که یافتن وابستگی آماری میان آن و تعدادی از متغیرهای مستقل موردنظر است و معمولاً از طریق اندازه‌گیری های تجربی بدست می‌آیند. از طرفی، مقدار متغیر وابسته با آهنگ متغیر مستقل تغییر می‌کند، بطوریکه به ازای هر مقداری از متغیر مستقل، برای متغیر وابسته یک یا چند مقدار وجود دارد. مدل رگرسیون خطی چندگانه، ساده ترین و متداول ترین مدلی است که برای ارتباط چند متغیر بکار می‌رود، که توسط رابطه ی (۱-۲) نمایش داده می‌شود:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad \text{رابطه (۱-۲)}$$

در این معادله  $\beta_0$  و  $\beta_1$  و  $\beta_p$  ضرایب رگرسیون (ضرایب متغیرهای مستقل) و  $X_1$  و  $X_2$  و  $X_p$  متغیرهای مستقل (توصیف‌کننده‌های عددی) و  $Y_i$  متغیر وابسته‌ی موردنظر (فعالیت دارویی) می‌باشند.

مقدار ضریب متغیر مستقل در معادله، بیانی از میزان اهمیت آن متغیر در پیش بینی مقادیر متغیر وابسته است. روشی که غالباً برای بدست آوردن معادله‌ی خط بین متغیرهای مستقل و وابسته

<sup>1</sup>-Principle Component Analysis

<sup>2</sup>-Partial Least Squares

<sup>3</sup>-Artificial Neutral Network

بکار می رود، روش حداقل مربعات<sup>۱</sup> است و خطی که با این روش بدست می آید، دارای بهترین معادله برای مرتبط کردن متغیرهای مستقل و وابسته است. ویژگی چنین خطی این است که مجموع مربعات انحراف عمودی نقاط مشاهده شده  $(y_i)$  از این خط، کمتر از مجموع مربعات انحراف عمودی نقاط مشاهده شده  $(y_i)$  از هر خط مفروض دیگر است. برای بدست آوردن معادله‌ی رگرسیون خطی چندگانه، باید مجموع مربعات انحرافات مقادیر پیشگویی شده برای متغیر وابسته از مقادیر واقعی آن به کمترین حد ممکن کاهش یابد. به این معناکه باید مقدار  $(\sum e_i^2)$  که طبق معادله‌ی زیر تعریف شده است به کمترین حد ممکن کاهش یابد.

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 \quad \text{رابطه (۲-۲)}$$

که در آن  $(y_i)$  مقدار مشاهده شده‌ی متغیر وابسته و  $(\hat{y}_i)$  مقدار محاسبه شده‌ی آن بوسیله‌ی رگرسیون است. در اینجا، برای محاسبه‌ی ضرایب نیز از روش کمترین مجموع مربعات استفاده می‌شود.

## ۲-۲-۷- شبکه عصبی مصنوعی

نحوه‌ی یادگیری و ایجاد ارتباط میان مفاهیم و اشیاء یکی از رموز پیچیده‌ی دستگاه آفرینش است. مغز به عنوان یک سیستم پردازش اطلاعات با ساختار کاملاً پیچیده حدود دو درصد وزن بدن را تشکیل می‌دهد و کلیه‌ی اعمال آگاهانه و بسیاری از رفتارهای ناآگاهانه را کنترل می‌کند. این دستگاه که از تجمع و آرایش ویژه‌ی سلولهای تخصص یافته‌ای به نام نرون تشکیل شده است، با تأثیرپذیری و تأثیرگذاری متقابل بر دستگاه‌های دیگر بدن، ارتباط هدفمند جاندار را با محیط اطرافش ممکن می‌سازد. دریافت پیامهای بیرونی، تحلیل و پاسخ دهی به آنها منوط به چینش صحیح و عملکرد سریع نورونهاست [۲۸]. پیشرفت عمده شبکه‌های عصبی از سال ۱۹۴۳ با فعالیت‌های وارن مک کلوت<sup>۲</sup> و

<sup>1</sup>-Least Squares Method

<sup>2</sup>-Warren McCulloch

والتر پیتس<sup>۱</sup> آغاز شد. آنها نشان دادند که شبکه عصبی می تواند هر تابع حسابی یا منطقی را محاسبه کند [۲۹]. به دنبال آن دونالد هب<sup>۲</sup> مکانیسمی برای آموزش شبکه ارائه داد. اولین کاربرد عملی شبکه-های مصنوعی در سال ۱۹۵۸ بود که در آن از شبکه‌ای که روزنبلات<sup>۳</sup> پیشنهاد داده بود، استفاده شد [۳۰]. روزنبلات یک شبکه چند لایه را ارائه کرد و توانایی آنرا در انجام شناسایی الگو اثبات نمود. این موفقیت زود هنگام باعث پیدایش زمینه وسیعی در تحقیقات شبکه عصبی شد. تقریباً در همان زمان برنارد ویدرو<sup>۴</sup> الگوریتم آموزشی جدیدی ارائه داد که از لحاظ ساختار به شبکه روزنبلات شبیه بود. قاعده آموزش ویدرو هاف امروزه نیز در شبکه‌های عصبی استفاده می‌شود. اما شبکه ویدرو هاف و روزنبلات دارای محدودیت‌هایی بود که اولین بار توسط ماروین مینسکی<sup>۵</sup> مطرح شد. محققان تحت تاثیر نظریه مینسکی بر این عقیده بودند که تحقیق در مورد شبکه‌های عصبی پایان یافته است و از آنجایی که کامپیوترهای دیجیتال قوی در دسترس نبود، نزدیک به یک دهه تحقیق در مورد شبکه - های عصبی معلق ماند. اما در سال ۱۹۸۶ با ظهور کامپیوترهایی با توان محاسباتی سریع، مجدد تحقیق در مورد شبکه‌های عصبی آغاز شد. در این سال الگوریتم آموزشی پس انتشار توسط جیمز مک کلو<sup>۶</sup> ارائه شد [۳۱ و ۳۲] که نقطه عطفی در تحقیقات شبکه‌های عصبی به شمار می‌آید. از سال ۱۹۸۶ تاکنون محققان زیادی در مورد شبکه‌های عصبی تحقیق نموده‌اند که نتیجه آن صدها مقاله در این زمینه است.

امروزه از شبکه‌های عصبی به عنوان یک ابزار کارآمد در زمینه‌های مختلف علمی از جمله صنایع الکترونیک، پزشکی، اکتشاف نفت و گاز، رباتیک، شیمی و داروسازی استفاده می‌شود [۳۲]. برای آشنایی با شبکه های عصبی مصنوعی، ابتدا باید سیستم عصبی طبیعی را به اختصار مورد بررسی قرار داد.

---

1-Walter Pitts

2-Donald Hebb

3-Rosenblatt

4- Bernard Widrow

5- Marvin Minsky

6-james McClelland

## ۲-۷-۱- سیستم عصبی طبیعی

همانطور که در بخش قبل بیان شد مغز بیش از دو درصد از وزن بدن را تشکیل می‌دهد و بیش از بیست درصد کل اکسیژن بدن را مصرف می‌کند. مغز از ۱۰۰ تریلیون نرون که با بیش از  $10^{16}$  ارتباط به هم متصل شده‌اند تشکیل شده است. نرون‌ها ساده‌ترین واحد ساختاری سیستم‌های عصبی هستند. بیشتر نرون‌ها از سه قسمت اصلی تشکیل شده‌اند:

۱- بدنه ی سلول که شامل هسته و قسمت‌های اساسی دیگر می باشد.

۲- دندریت<sup>۱</sup>

۳- آکسون<sup>۲</sup>

دندریت‌ها به عنوان مناطق دریافت سیگنال‌های الکتریکی، شبکه‌هایی تشکیل یافته از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بی شمار می باشند. به همین علت آنها را شبکه‌های دریافت‌کننده درخت‌مانند نیز می‌نامند. دندریت‌ها سیگنال‌های الکتریکی را به هسته‌ی سلول منتقل می‌کنند. بدنه‌ی سلول، انرژی لازم را برای فعالیت نرون فراهم نموده و بر روی سیگنال‌های دریافتی عمل می‌کند.

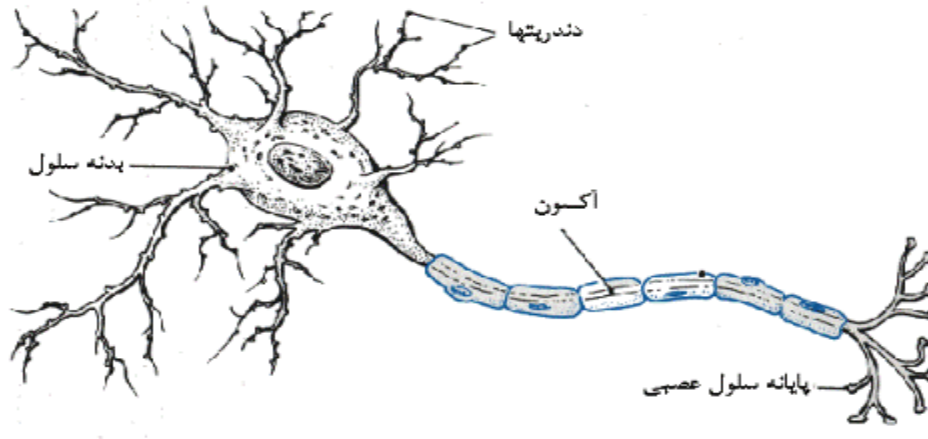
آکسون‌ها بر عکس دندریت‌ها از سطحی هموارتر و تعداد شاخه‌های کمتری برخوردار هستند. آنها طول بیشتری دارند و سیگنال‌الکتروشیمیایی دریافتی از هسته‌ی سلول را به نرون‌های دیگر منتقل می‌کنند. هر آکسون توسط یک فاصله‌ی کوچک موسوم به سیناپس از نرون‌های مجاور جدا می‌شود. سیناپس‌ها واحدهای ساختاری کوچکی هستند که ارتباط بین نرون‌ها را برقرار می‌سازند. چگونگی برقراری این ارتباط به میزان مواد انتقال‌دهنده‌ی نرونی که در انتهای آکسون‌ها ذخیره شده‌اند بستگی دارد. وقتی که یک پتانسیل تحریک به انتهای یک آکسون می‌رسد، موجب آزاد شدن یک ماده‌ی شیمیایی به نام انتقال‌دهنده‌ی نرونی از انتهای آکسون می‌شود و پس از نفوذ در

---

<sup>1</sup>-Dendrite

<sup>2</sup>-Axon

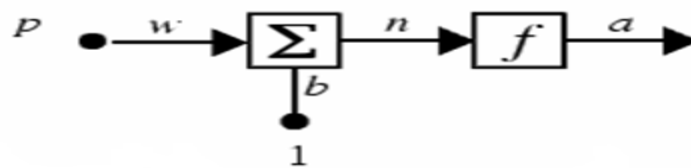
سیناپس ها، گیرنده‌های سلول‌های مجاور را فعال می‌کند [۳۳]. شکل (۲-۲) شماتیکی از ساختار کلی یک نرون عصبی طبیعی می‌باشد.



شکل (۲-۲) - شماتیک سلول عصبی طبیعی [۳۳]

#### ۴-۲-۷-۲- نرون مصنوعی

هر نرون مصنوعی نیز مانند یک نرون طبیعی از سه جزء لایه‌ی ورودی، لایه‌ی میانی و لایه‌ی خارجی تشکیل شده است. لایه‌ی ورودی سیگنال‌ها را از محیط خارج دریافت می‌کند. لایه‌ی میانی عملیات محاسباتی را روی سیگنال‌های ورودی انجام می‌دهد. لایه‌ی خارجی سیگنالها را به خارج یا نرون‌های دیگر انتقال می‌دهد. عمل لایه‌ی ورودی، لایه‌ی میانی و لایه‌ی خارجی در یک شبکه‌ی عصبی شبیه عمل دندریت، جسم سلولی و آکسون در یک نرون طبیعی است. ساختار و عملکرد یک نرون مصنوعی ساده در شکل (۳-۲) نمایش داده شده است [۱۸]. در این نرون، چند ورودی کمی در ضرایب مخصوص که وزن ( $w$ ) نام دارد ضرب می‌شوند و وارد نرون می‌گردند.



شکل (۳-۲) - مدل نرون مصنوعی

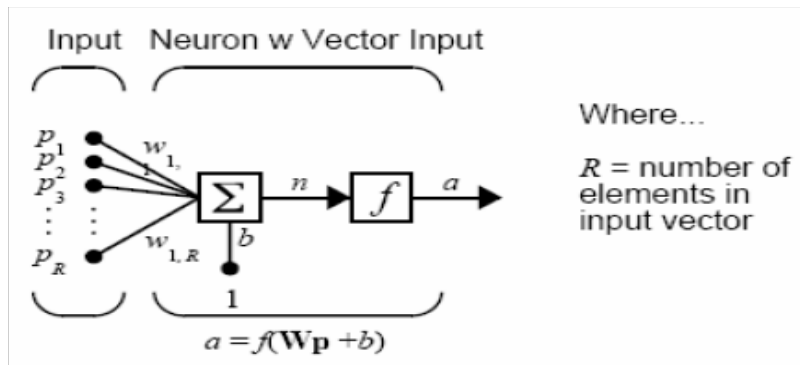
<sup>1</sup>- Weight

در این شکل از علامت  $p$  برای نشان دادن یک سیگنال ورودی استفاده می شود. در واقع در این مدل یک سیگنال ورودی پس از تقویت (یا تضعیف) شدن با وزن  $w$ ، با مقدار  $n=wp$  به تابع انتقال نرون،  $f$ ، اعمال می شود. ورودی بایاس<sup>۱</sup> یک مقدار ثابت، ۱، است که در جمله بایاس ضرب شده و سپس با  $pw$  جمع می شود. این حاصل جمع، ورودی خالص  $n$  برای تابع محرک  $f$  را تشکیل می دهد.

$$n = wp + b \quad \text{رابطه (۳-۲)}$$

$b$  و  $w$  پارامتر تنظیم شونده در نرون ها می باشند و ایده‌ی اصلی شبکه عصبی این است که با تغییر مقادیر  $b$  و  $w$ ، شبکه یک رفتار یا تصمیم را اتخاذ کند. در جعبه ابزار مورد استفاده در نرم افزار مطلب<sup>۲</sup>، بایاس در نظر گرفته شده اما استفاده از آن اختیاری می باشد [۳۴].

در حالتی که طبق شکل (۴-۲) ورودی نرون به صورت یک بردار باشد، خروجی نرون از طریق روابط (۴-۲) و (۵-۲) به دست می آید [۳۵ و ۳۶].



شکل (۴-۲) - نرون با یک بردار ورودی

$$n = w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b \quad \text{رابطه (۴-۲)}$$

$$a = f(w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b) \quad \text{رابطه (۵-۲)}$$

<sup>1</sup>-Bias

<sup>2</sup>-MATLAB

## ۲-۲-۷-۳- توابع انتقال نرون

سه تابع انتقال رایج در شبکه های عصبی، که در این پایان نامه برای بهینه سازی شبکه به کار رفته است عبارتند از:

➤ تابع انتقال خطی

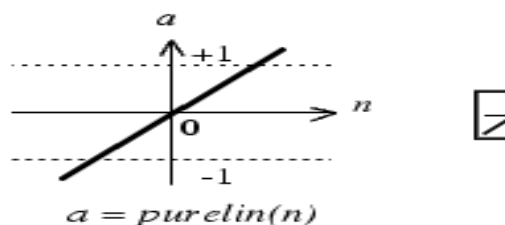
نرون هایی که از این تابع انتقال استفاده می کنند معمولاً برای تقریب خطی در فیلترهای خطی به کار می روند. این تابع همان مقدار ورودی را به عنوان خروجی برمی گرداند (شکل ۲-۵ الف).

➤ تابع انتقال لگاریتم زیگموئید (logsig)

از این تابع انتقال در شبکه های پس انتشار استفاده می شود. این تابع انتقال مقادیر ورودی را در محدوده  $-\infty$  تا  $+\infty$  دریافت کرده و خروجی بین ۰ و ۱ تولید می نماید (شکل ۲-۵ ب).

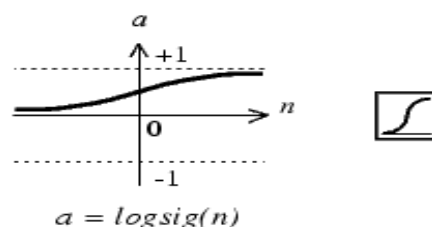
➤ تابع انتقال تانژانت هایپربولیک (tansig)

این تابع انتقال مقادیر ورودی را در محدوده  $-\infty$  تا  $+\infty$  دریافت کرده و خروجی بین ۱- و ۱- تولید می نماید (شکل ۲-۵ ج).



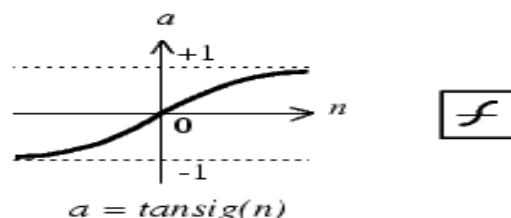
Linear Transfer Function

(الف)



Log-Sigmoid Transfer Function

(ب)



Tan-Sigmoid Transfer Function

(ج)

شکل (۲-۵) - برخی از توابع انتقال مورد استفاده در شبکه های عصبی

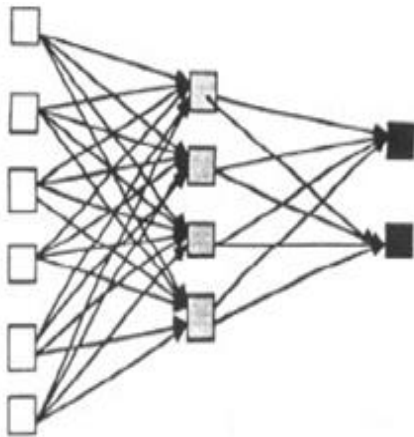
## ۲-۲-۷-۴- انواع شبکه های عصبی از نظر برگشت پذیری

در این بخش انواع شبکه های عصبی را از نظر برگشت پذیری مورد بررسی قرار می دهیم:

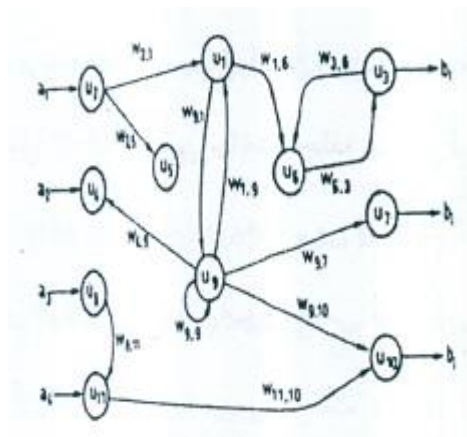
➤ شبکه های پیشخور<sup>۱</sup>

➤ شبکه های برگشتی<sup>۲</sup>

شبکه های پیشخور، شبکه هایی هستند که مسیر پاسخ در آن ها همواره رو به جلو پردازش می شود و به نرون های لایه قبل باز نمی گردد. در این نوع شبکه ها به سیگنال اجازه می دهد که در مسیر یکطرفه از ورودی تا خروجی عبور کند. بنابراین بازخوردی<sup>۳</sup> وجود ندارد بدین معنی که خروجی هر لایه تأثیری بر همان لایه ندارد. شکل (۲-۶ الف) نمونه ای از شبکه پیشخور را ارائه می دهد.



الف



ب

شکل (۲-۶) - انواع شبکه های عصبی: الف) شبکه عصبی پیشخور ب) شبکه عصبی برگشتی [۳۵]

ولی طبق شکل (۲-۶ ب) در شبکه های برگشتی حداقل یک سیگنال برگشتی از یک نرون به همان نرون یا نرون های دیگر همان لایه و یا نرون های لایه قبل وجود دارد. به عبارتی در این نوع شبکه ها بازخورد وجود دارد [۳۵ و ۳۶].

<sup>1</sup>-Feed Forward

<sup>2</sup> - Recurrent

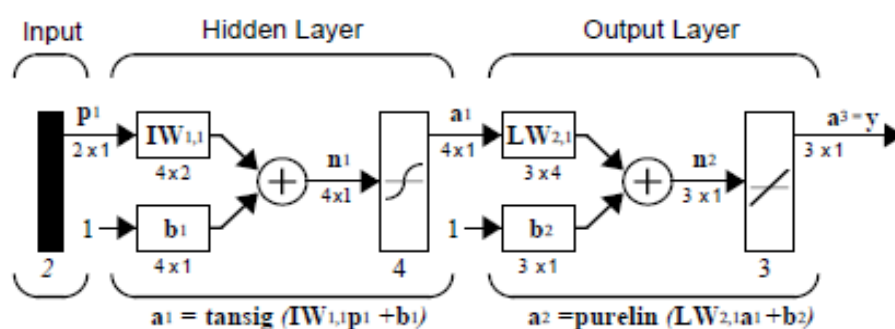
<sup>3</sup>-Feedback



## ۲-۲-۷-۵- معماری شبکه‌های پیشخور

در شبکه‌های پیشخور، نرون‌ها در چند لایه سازمان‌دهی می‌شوند. اولین و آخرین لایه به ترتیب لایه ورودی<sup>۱</sup> و لایه خروجی<sup>۲</sup> نام دارند، در حالی که لایه‌های بین این دو لایه، لایه‌های پنهان<sup>۳</sup> نامیده می‌شوند. در این نوع شبکه‌ها ابتدا اطلاعات ورودی از محیط خارج به نرون‌های لایه ورودی ارسال می‌شود. در این لایه هیچ پردازشی بر روی ورودی‌ها صورت نمی‌گیرد و در واقع لایه ورودی نقش انتقال-دهنده را بر عهده دارد. خروجی این نرون‌ها به لایه بعدی که همان لایه پنهان است، منتقل می‌شود. در نهایت خروجی شبکه از طریق نرون‌های لایه خروجی به محیط خارج ارسال می‌گردد.

طبق شکل (۲-۷)، شبکه‌های پیشخور اغلب از یک یا چند لایه مخفی از نرون‌های سیگموئیدی و از یک لایه پایانی خطی استفاده می‌کنند. شبکه چند لایه از نرون‌ها با یک تابع انتقال غیرخطی به شبکه اجازه می‌دهد که توانایی یادگیری رابطه خطی و غیرخطی را بین ورودی‌ها و خروجی‌ها داشته باشد. لایه خروجی خطی به شبکه این امکان را می‌دهد که خروجی در هر محدوده دلخواهی از  $-\infty$  تا  $+\infty$  باشد. البته اگر بخواهیم خروجی در دامنه محدودی قرار گیرد، می‌توان از توابع محدودسازی مانند توابع سیگموئیدی در لایه خروجی استفاده کرد [۳۷].



$P$ : بردار ورودی‌ها،  $w$ : بردار وزن‌ها،  $f$ : تابع انتقال،  $b$ : بایاس و  $n$ : ورودی تابع انتقال و  $a$ : خروجی تابع انتقال

شکل (۲-۷) - شبکه پیشخور  $\text{tansig/pureline}$  [۳۷]

- 1- Input layer
- 2- Output layer
- 3- Hidden layer

## ۲-۲-۷-۶- مراحل آموزش در شبکه‌های پیشخور

به طور کلی مراحل که در فرایند آموزش دنبال می‌شوند، شامل ایجاد شبکه و تعیین مجموعه داده‌های آموزش، مقداردهی آغازین به وزن‌ها، آموزش شبکه و شبیه‌سازی<sup>۱</sup> پاسخ شبکه می‌باشد [۳۷].

### ۲-۲-۷-۶-۱- ایجاد شبکه

اولین گام در فرایند آموزش با شبکه‌های پیشخور ایجاد شبکه است. در نرم‌افزار مطلب تابع `newff` یک شبکه پیشخور ایجاد می‌کند. این شبکه حداقل به چهار پارامتر نیاز دارد تا خروجی شبکه را برگرداند. این پارامترها شامل پارامتر ماتریس بردار ورودی، ماتریس مقادیر مطلوب خروجی -ها، نوع تابع انتقال و نوع تابع آموزش مورد استفاده، می‌باشند.

### ۲-۲-۷-۶-۲- مقدار دهی آغازین به وزن‌ها

قبل از آموزش یک شبکه پیشخور، مقادیر وزن‌ها و بایاس‌ها باید تعیین گردد. تابع `newff` به صورت اتوماتیک به وزن‌ها مقدار می‌دهد ولی اگر بخواهیم که مقادیر وزن‌ها و بایاس‌ها بر اساس تابع مقداردهی اولیه<sup>۲</sup> که خود تعیین کرده‌ایم، به روز شود، از دستور `init` استفاده می‌کنیم.

### ۲-۲-۷-۶-۳- آموزش شبکه

پس از مقدار دهی به وزن‌ها و بایاس‌ها نوبت به آموزش شبکه می‌رسد. فرایند آموزش به یک سری مثال‌ها از رفتار مورد انتظار شبکه نیاز دارد که شامل ورودی شبکه و مقادیر هدف آن ورودی می‌باشد.

---

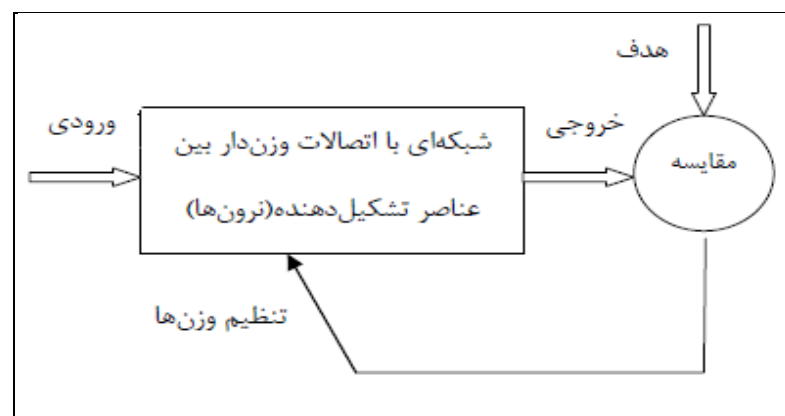
<sup>1</sup>-Simulation

<sup>2</sup>-Initialization function

در طول آموزش وزن‌ها و بایاس‌ها تنظیم می‌شوند تا تابع کارایی حداقل شود. تابع کارایی پیش‌فرض برای شبکه‌های پیش‌خور، میانگین مربع خطا<sup>۱</sup> (mse) می‌باشد.

## ۲-۲-۷-۶-۴- روش‌های آموزش

برای آموزش شبکه‌های عصبی آموزش نظارت شده<sup>۲</sup> و آموزش نظارت نشده<sup>۳</sup> مورد استفاده قرار می‌گیرند. در شیوه آموزش نظارت شده، مجموعه‌ای از زوج داده‌ها که به داده‌های آموزش معروفند، در اختیار شبکه گذاشته می‌شود. هر داده آموزشی شامل ورودی و خروجی متناظر با آن ورودی می‌باشد. در این شیوه یادگیری، پس از اعمال ورودی‌ها به شبکه، خروجی شبکه با خروجی مطلوب مقایسه می‌گردد (شکل (۲-۸)) و پس از محاسبه‌ی خطای یادگیری، از آن برای اصلاح پارامترهای شبکه (وزن‌ها و بایاس‌ها) استفاده می‌گردد، به گونه‌ای که اگر دفعه بعد همین ورودی به شبکه اعمال گردد، خروجی شبکه به خروجی مطلوب نزدیک‌تر باشد [۳۲ و ۳۷]. در حالی که در آموزش نظارت نشده، تنها سیگنالی که از محیط به شبکه اعمال می‌گردد، بردار ورودی شبکه است، بنابراین ناظری وجود نخواهد داشت تا خروجی مطلوب را برای آموزش به شبکه اعمال نماید. شبکه‌هایی را که از این نوع یادگیری بهره می‌برند، اصطلاحاً شبکه‌های خود سازمان‌ده<sup>۴</sup> می‌گویند [۳۷].



شکل (۲-۸) - شیوه آموزش نظارت شده

<sup>1</sup>-Mean Square Error  
<sup>2</sup> - Supervised training  
<sup>3</sup>- Unsupervised training  
<sup>4</sup>-Self organized

## ۲-۷-۷- آموزش شبکه‌های پیشخور با تکنیک پس انتشار<sup>۱</sup>

پس‌انتشار خطا یک روش متداول آموزش با ناظر برای شبکه‌های پیشخور است یعنی برای به دست آوردن ارتباط بین متغیرهای ورودی و خروجی در یادگیری به الگوی آموزشی نیاز است بنابراین همانطور که در شکل (۲-۹) مشاهده می‌شود، باید داده‌های آموزشی (زوج ورودی-هدف) به شبکه ارائه شود. دو روش مختلف برای ارائه الگوهای آموزشی و پیاده‌سازی الگوریتم وجود دارد:

➤ آموزش گام به گام<sup>۲</sup>

➤ آموزش دسته‌ای<sup>۳</sup>

در آموزش گام به گام وزن‌ها و بایاس‌ها بعد از اعمال هر ورودی به روز<sup>۴</sup> می‌شوند در حالی که در روش دسته‌ای پس از اعمال تمام ورودی‌ها (اعضای مجموعه آموزشی) عملیات به روزرسانی وزن‌ها انجام می‌شود. بدین صورت که شیب‌های محاسبه شده برای هر ورودی با هم جمع می‌شوند تا در نهایت وزن‌ها و بایاس‌ها از طریق آن به روز شود. در این تحقیق شیوه آموزش دسته‌ای مورد استفاده قرار گرفته است [۳۶].

به طور کلی آموزش به کمک تکنیک پس‌انتشار بر طبق مراحل زیر انجام می‌شود [۳۷]:

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

---

<sup>1</sup> - Back propagation

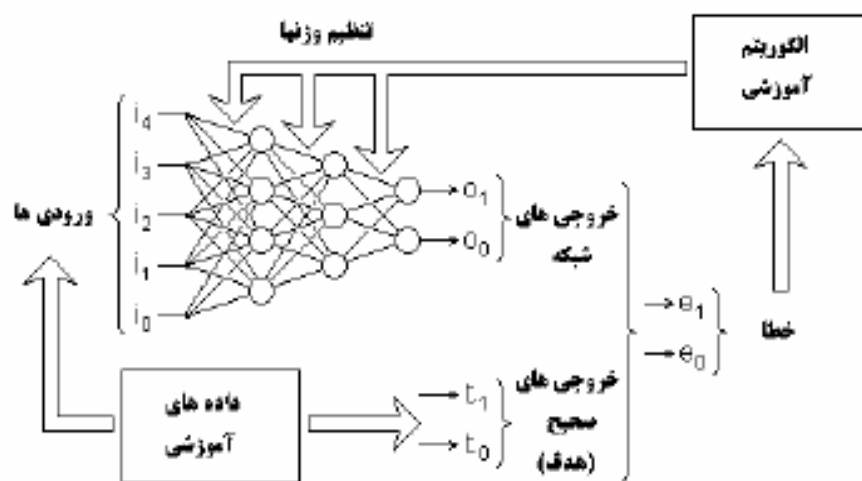
<sup>2</sup> -Incremental training

<sup>3</sup> -Batch training

<sup>4</sup> -Update

مراحل فوق تا زمانی تکرار می‌شود که به حداکثر تکرار<sup>۱</sup> مجاز رسیده باشد یا مقدار تابع کارایی از مقداری که تعیین شده کمتر باشد.

به دلیل فرایندی که در طی این تکنیک رخ می‌دهد، این روش انتشار به عقب نامیده می‌شود زیرا ابتدا وزن‌های ارتباطی بین لایه خروجی و لایه پنهان و سپس وزن‌های ارتباطی بین لایه پنهان و لایه ورودی اصلاح می‌گردند. شکل (۹-۲) نیز گویای این مطلب است.



شکل (۹-۲) - نحوه عملکرد تکنیک پس‌انتشار در یک شبکه عصبی پیشخور

متداول‌ترین الگوریتمی که در الگوی آموزشی پس‌انتشار به کار می‌رود، الگوریتم آموزشی نزول گرادیانی<sup>۲</sup> است. در این الگوریتم آموزشی پارامترهای شبکه از قبیل وزن‌ها و بایاس‌ها در جهتی که تابع کارایی حداقل شود، تغییر می‌کند. به روز رسانی وزن‌ها در گام آموزشی (n+1) به صورت زیر است:

$$w_{ji,n+1} = w_{ji,n} + \alpha \Delta w_{ji,n} \quad \text{رابطه (۶-۲)}$$

$$\Delta w_{ji,n} = - \frac{\partial E}{\partial w_{ji,n}} \quad \text{رابطه (۷-۲)}$$

<sup>1</sup>Epoch

<sup>2</sup>-Gradient descent

که در روابط فوق  $W_{ji,n}$  وزن ارتباطی بین نرون  $i$  ام و نرون  $j$  ام در گام آموزشی  $n$  ام،  $\alpha$  سرعت یادگیری و  $E$  تابع خطاست. سرعت یادگیری اندازه گام‌های برداشته شده در اصلاح وزن‌ها را نشان می‌دهد. سرعت یادگیری زیاد باعث نوسان‌های زیاد در حین آموزش شده در حالی که مقدار کم آن باعث آهستگی سرعت همگرایی می‌شود. علامت منفی در رابطه نشان می‌دهد که حرکت به سمتی است که تابع خطا حداقل شود به بیان دیگر هدف از تصحیح وزن‌ها کاهش مقدار خطاست.

نزول گرادیانی با وجود متداول بودنش دارای دو ضعف عمده و اساسی است. اول سرعت همگرایی پایین [۳۸] و دوم ساکن شدن و متوقف شدن پارامترهای شبکه به دلیل به دام افتادن آنها در نقاط مینیمم محلی<sup>۱</sup> [۳۹ و ۴۰] قبل از پایان یادگیری کامل شبکه می‌باشد.

الگوریتم دیگری که در الگوی آموزشی پس‌انتشار استفاده می‌شود، الگوریتم آموزشی لونبرگ مارکوارت<sup>۲</sup> است این الگوریتم آموزشی که تقریبی از روش نیوتن است، اولین بار توسط لونبرگ در سال ۱۹۹۴ پیشنهاد شد. در این الگوریتم ماتریس هسین که مشتقات مرتبه دوم خطا نسبت به پارامترهای شبکه است، به صورت زیر تقریب زده می‌شود:

$$H(x) = \nabla^2 F(x) \approx J^T(x)J(x) \quad \text{رابطه (۲-۸)}$$

$$g(x) = \nabla F(x) \approx J^T(x)E(W) \quad \text{رابطه (۲-۹)}$$

که  $E(W)$  بردار خطای شبکه و  $J$  ماتریس ژاکوبین که شامل مشتقات مرتبه اول خطای شبکه نسبت به پارامترهای شبکه است و به فرم زیر نوشته می‌شود:

<sup>1</sup> - Local minimum

<sup>2</sup> - Levenberg-Marquardt

$$J(x) = \begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \dots & \dots & \frac{\partial e_1(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \dots & \dots & \frac{\partial e_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{\partial e_n(x)}{\partial x_1} & \frac{\partial e_n(x)}{\partial x_2} & \dots & \dots & \frac{\partial e_n(x)}{\partial x_n} \end{bmatrix}_{n \times n} \quad \text{رابطه (۲-۱۰)}$$

طبق این الگوریتم، تغییر پارامترها در هر گام آموزشی برابر است با [۳۸ و ۴۱]:

$$x_{k+1} = x_k - [J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)E(x_k) \quad \text{رابطه (۲-۱۱)}$$

این الگوریتم حالت بینابین روش نزول گرادیانی و روش گوس-نیوتن است که در آن ماتریس  $I$  یکانی،  $x_k$  پارامتر شبکه (وزن ها و بایاس ها) در گام آموزشی  $k$  ام،  $E(x_k)$  بردار خطا در گام آموزشی  $k$  ام و  $\mu_k$  ضریب یادگیری مارکوات و عددی غیر منفی می باشد. هر قدر  $\mu_k$  بزرگتر شود، نحوه آموزش به نزول گرادیانی با سرعت یادگیری کوچک ( $\frac{1}{\mu_k}$ ) نزدیک تر می شود:

$$x_{k+1} = x_k - \frac{1}{\mu_k} J^T(x_k)E(x_k) = x_k - \frac{1}{\mu_k} g(x_k) \quad \text{رابطه (۲-۱۲)}$$

و هر چه  $\mu_k$  کوچک تر انتخاب شود، نحوه آموزش به گوس-نیوتن نزدیک تر می شود:

$$x_{k+1} = x_k - [J^T(x_k)J(x_k)]^{-1} J^T(x_k)E(x_k) \quad \text{رابطه (۲-۱۳)}$$

این ضریب پس از هر گام موفق کاهش می یابد و تنها در صورتی افزایش می یابد که گام آزمایشی تابع کارایی را افزایش دهد. به این ترتیب تابع کارایی همیشه در حال کاهش است. مراحل که در حین آموزش با این الگوریتم دنبال می شود، به شرح زیر است:

۱- ارائه ورودی به شبکه و تعیین مقدار اولیه وزن ها و  $\mu$

۲- محاسبه خروجی شبکه و خطای مربوط به آن

۳- محاسبه ماتریس ژاکوبین

۴- به روز کردن وزن‌ها بر طبق رابطه (۲-۱۱)

۵- محاسبه مجدد خطای شبکه

اگر خطای محاسبه شده در این حالت کمتر از حالت اول باشد، مقدار  $\mu$  کاهش داده می‌شود و سپس با توجه به مقدار جدید  $\mu$  و مقادیر جدید وزن‌ها به مرحله (۱) برمی‌گردد و چنانچه خطا افزایش یافته باشد، نتایج مربوط به آن دور نادیده گرفته می‌شود (تغییرات مربوط به وزن‌ها اعمال نمی‌شود) و  $\mu$  در فاکتوری بزرگتر ضرب می‌شود و به مرحله (۴) برمی‌گردد. تکرار مراحل این الگوریتم تا زمانی ادامه دارد که تابع کارایی به کمترین مقدار خود برسد و یا به حداکثر تعداد چرخه‌های آموزش رسیده باشد.

## ۲-۲-۷-۸- بهبود تعمیم<sup>۱</sup>

یکی از مشکلات در آموزش شبکه‌های عصبی، بیش‌برازش<sup>۲</sup> شبکه است. اگر شبکه بیش از حد آموزش داده شود، شروع به خاطر سپردن داده‌ها می‌کند و از پیدا کردن ارتباط موجود در داده‌ها دور می‌شود. به عبارتی در بیش‌برازش، خطا بر روی مجموعه آموزش به مقدار قابل قبول می‌رسد ولی با ارائه داده‌های جدید به عنوان ورودی، خطای شبکه بسیار بالا می‌باشد [۴۲ و ۴۳]. در اصطلاح گفته می‌شود شبکه تعمیم برای حالت‌های جدید را یاد نگرفته است. یک روش برای ارتقای عمومیت شبکه، بکارگیری شبکه‌های بزرگ برای ایجاد یک انطباق مناسب است. اگر ما از شبکه کوچکی استفاده کنیم، این شبکه توانایی لازم را برای انطباق داده‌ها نخواهد داشت. متأسفانه دانستن این مورد که از ابتدا، اندازه شبکه برای یک کاربرد خاص چه میزان می‌بایست بزرگ باشد، مشکل است.

---

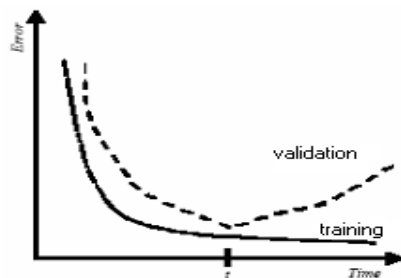
<sup>1</sup>- Improving generalization

<sup>2</sup>-Over fitting



در جعبه ابزارمطلب برای جلوگیری از بیش برآزش و ارتقایعمومیت روش‌های مختلفی از جمله توقف زودرس<sup>۱</sup> و تنظیم<sup>۲</sup> وجود دارد [۴۴].

در توقف زودرس برای آموزش شبکه، داده‌ها به‌طور تصادفی به سه زیر مجموعه آموزش، ارزیابی و تست تقسیم می‌شوند، درحین آموزش شبکه، زمانی که شبکه سعی در تطابق بیش از حد دارد، طبق شکل (۲-۱۰) خطای مجموعه ارزیابی شروع به افزایش می‌کند که در این هنگام باید آموزش متوقف شود [۳۶].



شکل (۲-۱۰) - تغییرات خطای سری آموزش و سری ارزیابی [۳۶]

تابع کارایی معمول مورد استفاده در شبکه‌های feedforward مجموع مربعات خطاها می‌باشد.

$$F = MSE = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - t_i)^2 \quad \text{رابطه (۲-۱۴)}$$

روش دیگر برای ارتقای تعمیم شبکه، تنظیم نامیده می‌شود که با اصلاح تابع کارایی از طریق افزایش یک عبارت که شامل میانگین مربعات وزن‌ها و بایاس‌های شبکه است، انجام می‌گیرد.

$$F = \beta MSE + \alpha MSW \quad \text{رابطه (۲-۱۵)}$$

$$MSW = \frac{1}{N} \sum_{j=1}^N (w_j)^2 \quad \text{رابطه (۲-۱۶)}$$

<sup>1</sup>-Early Stopping

<sup>2</sup>-Regularization

$\alpha$  و  $\beta$  پارامترهای تابع کارایی، MSE و MSW به ترتیب میانگین مربعات خطاها و میانگین مربعات وزن‌ها و N تعداد ترکیبات است. استفاده از این تابع کارایی اصلاح شده سببی می‌شود که شبکه، وزن-ها و بایاس‌ها را کوچک‌تر پیدا داشته باشد و این پاسخ‌بخش به همراه موازنه می‌کند.

در الگوریتم آموزشی تنظیم بایزین، اصلاح وزن‌ها و بایاس‌ها همانند الگوریتم لونیبرگ مارکوارت انجام می‌شود با این تفاوت که تابع خطا، همان تابع کارایی اصلاح شده است یعنی تابع خطا ترکیبی از میانگین مربع خطاها و میانگین مربع وزن‌هاست [۴۳] و هدف از آموزش به روش پس انتشار، کاهش این تابع کارایی اصلاح شده است. مراحل طی شده حین آموزش شبکه با الگوریتم تنظیم بایزین این چنین است:

۱- تعیین مقدار اولیه وزن‌ها،  $\alpha$  و  $\beta$

۲- اجرای الگوریتم لونیبرگ مارکوارت جهت کاهش تابع کارایی (رابطه ۲-۱۵)

۳- محاسبه تعداد مؤثر پارامترهای تابع کارایی ( $\gamma$ ) طبق رابطه ۲-۱۷

$$\gamma = N - 2\alpha \text{tr}(H)^{-1} \quad \text{رابطه (۲-۱۷)}$$

که در این رابطه N تعداد کل پارامترها و  $\text{tr}(H)^{-1}$  اثر ماتریس معکوس هسیان است.

۴- تنظیم مجدد پارامترهای تابع کارایی با استفاده از روابط (۲-۱۸) و (۲-۱۹)

$$\alpha = \frac{\gamma}{2E_w} \quad \text{رابطه (۲-۱۸)}$$

$$\beta = \frac{n - \gamma}{2E_D} \quad \text{رابطه (۲-۱۹)}$$

۵- تکرار مراحل ۲ تا ۴ برای رسیدن به همگرایی

شرایط پایان آموزش و توقف این الگوریتم همانند الگوریتم لونیبرگ مارکوارت است.

## ۲-۲-۸- ارزیابی مدل

برای اینکه بتوان دقت و صحت نتایج ارائه شده توسط مدل را ارزیابی کرد از شاخص‌های کمی استفاده می‌شود و با استفاده از این شاخص‌ها می‌توان از مناسب بودن مدل اطمینان حاصل کرد. معمولاً برای ارزیابی شبکه، داده‌ها به سه دسته‌ی آموزش، ارزیابی (اعتبار) و تست تقسیم می‌شوند. شبکه ابتدا بوسیله‌ی نمونه‌های آموزشی تمرین داده و وزنهای آن انتخاب می‌شوند. سپس کیفیت شبکه بوسیله‌ی نمونه‌های ارزیابی، سنجیده می‌شوند. در صورت اینکه پاسخ مناسبی یافت نشود مرحله‌ی آموزش را مجدداً تکرار کرده و مرحله‌ی ارزیابی را اجرا می‌کنند و اینکار را تا زمان پاسخ مناسب به نمونه‌های ارزیابی ادامه می‌دهند. سپس به منظور سنجش قابلیت انطباق‌پذیری شبکه، آنرا بوسیله‌ی نمونه‌های تست می‌سنجند [۴۵].

شاخص‌های ارزیابی مختلفی که در این تحقیق استفاده شده اند در اینجا مورد بررسی قرار می‌گیرند:

✓ مجموع مربعات کل ( $SST^1$ ): طبق رابطه‌ی (۲-۲۰) بصورت مجموع مربعات تفاضل بین مقادیر

واقعی متغیر وابسته و مقدار میانگین تعریف می‌شود:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{رابطه (۲-۲۰)}$$

✓ مجموع مربعات خطا ( $SSE^2$ ): طبق رابطه‌ی (۲-۲۱) بصورت مجموع مربعات تفاضل بین

مقادیر تجربی متغیر وابسته برای نمونه  $i$  ام ( $y_i$ ) و مقادیر محاسبه شده با مدل برای نمونه  $i$  ام

( $\hat{y}_i$ ) تعریف می‌شود:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{رابطه (۲-۲۱)}$$

✓ ضریب همبستگی<sup>۳</sup>: مقیاسی از میزان همبستگی میان دو یا چند متغیر است و مقدار آن بین

-۱ تا +۱ تغییر می‌کند. هر چه مقدار ضریب همبستگی بیشتر باشد ارتباط خطی بین متغیر

1- Sum Square Total

2-Sum Square Error

3-Correlation coefficient

وابسته و متغیرهای مستقل بیشتر می شود. ضریب همبستگی با رابطه ی (۲۲-۲) تعریف

می شود:

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad \text{رابطه (۲۲-۲)}$$

✓ ضریب تعیین<sup>۱</sup>: شاخصی برای بیان دقت خط رگرسیون برآورد شده است و مقدار آن بین

صفر و یک تغییر می کند و مبین واریانس مشترک بین متغیر وابسته و مستقل است یعنی

نسبتی از واریانس متغیر وابسته که توسط متغیرهای مستقل به حساب آمده معین می شود و

توسط رابطه ی زیر تعریف می شود:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{رابطه (۲۳-۲)}$$

طبق این رابطه اگر  $R^2$  برابر ۱ باشد یعنی  $y_i = \hat{y}_i$  باشد، مدل و داده ها با هم برازش کامل دارند.

✓ مجموع مربع باقیمانده ها ( $PRESS$ ): بصورت مجموع مربعات تفاضل بین مقدار تجربی ( $y_i$ ) و

مقدار پیش بینی شده ( $\hat{y}_i$ ) تعریف می شود:

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{رابطه (۲۴-۲)}$$

✓ میانگین مربعات خطاها ( $MSE$ )<sup>۳</sup>: به شکل زیر محاسبه می شود:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad \text{رابطه (۲۵-۲)}$$

✓ خطای نسبی پیش بینی ( $REP$ )<sup>۴</sup>:

$$REP(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{رابطه (۲۶-۲)}$$

<sup>1</sup>- Determination coefficient

<sup>2</sup>- Predictive Residual Sum of Squares

<sup>3</sup>- Mean square error

<sup>4</sup>- Relative Error of Prediction

## ۲-۲-۹- تکنیکهای سنجش اعتبار

این تکنیکها برای بررسی قدرت پیش بینی مدلها استفاده می‌شوند، یعنی توانایی آنها را برای انجام پیش بینی‌های مطمئن برای موارد جدیدی که پاسخ آنها ناشناخته است می‌سنجند. یکی از شرطهای لازم برای اعتبار یک مدل رگرسیون، آن است که ضریب تعیین ( $R^2$ ) یا ضریب همبستگی چند متغیره ( $R$ ) آن تا حد ممکن به یک نزدیک باشد و خطای استاندارد ( $SE$ ) آن کوچک باشد، ولی این شرط برای اعتبارسنجی مدل کافی نیست، زیرا این پارامترها ارتباطی به توانایی مدل برای ایجاد پیش‌بینی‌های مطمئن بر روی داده‌های بعدی ندارند. در اینجا به بررسی تعدادی از تکنیک‌های سنجش اعتبار می‌پردازیم:

### ۲-۲-۹-۱- ارزیابی متقاطع<sup>۱</sup>

رایج ترین تکنیک سنجش اعتبار است که در آن هر بار یک یا یک گروه کوچک از داده ها را کنار می گذاریم، سپس برای داده‌های باقیمانده مدلی بدست آورده و پاسخ را برای داده‌هایی که کنار گذاشته‌ایم از روی این مدل بدست می‌آوریم.

### ۲-۲-۹-۲- اعتبارسنجی بیرونی<sup>۲</sup>

تکنیکی است که در آن برای انجام یک بررسی اضافی بر روی قدرت پیش بینی مدل بدست آمده از سری آموزش و ارزیابی، یک سری تست<sup>۳</sup> نیز در نظر گرفته می‌شود که سری تست نباید در روند آموزش و ارزیابی دخالتی داشته باشد.

---

<sup>1</sup>-Cross validation

<sup>2</sup>-External validation

<sup>3</sup>-Test set

## ۲-۲-۹-۳- تقسیم مجموعه به آموزش و ارزیابی

این تکنیک بر این اساس است که مجموعه داده ها را به یک سری آموزش<sup>۱</sup> و یک سری ارزیابی<sup>۲</sup> تقسیم می کنیم. مدل از سری آموزش ساخته می شود و قدرت تعمیم و پیش گویی آن توسط سری ارزیابی بررسی می شود. در این تکنیک، تقسیم بندی داده ها به دو سری به صورت تصادفی صورت می گیرد.

## ۲-۲-۹-۴- آزمون Y-تصادفی<sup>۳</sup>

این تکنیک برای مطالعه‌ی همبستگی شانس‌ی مدل طراحی شده است. وقتی در مدلی، همبستگی متغیرهای مستقل با متغیرهای پاسخ بصورت تصادفی باشد، دارای همبستگی شانس‌ی هستند. در این آزمون، مقادیر پاسخ در محدوده‌ی پاسخ ها بصورت تصادفی تغییر داده می شود. سپس همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از یکی از شاخص های آماری (معمولاً R) مورد بررسی قرار می گیرد که اختلاف زیاد بین شاخص های آماری بدست آمده از این روش و شاخص آماری بدست آمده از مدل اصلی نشان دهنده‌ی عدم وجود همبستگی شانس‌ی است. این فرایند را چندین بار باید تکرار کرد [۲۶].

## ۲-۳- نرم افزارهای مورد استفاده در تحقیق

در این تحقیق از بسته ی نرم افزاری متنوعی در مراحل مدل سازی استفاده کرده ایم که هر یک به اختصار معرفی می شوند.

---

<sup>1</sup>-Training set

<sup>2</sup>-Validation set

<sup>3</sup>-Y-Randomization

## ۲-۳-۱- نرم افزار Hyperchem

از بسته‌ی نرم افزاری Hyperchem [۲۴] برای رسم ساختار مولکولها و بهینه سازی ساختار آنها توسط روشهای کوانتومی مکانیکی استفاده می شود. به کمک این نرم افزار می توان زوایای پیوندی، طول پیوندها، زوایای پیچشی، بار اتمها و... را در مولکول بدست آورد. همچنین به کمک این نرم افزار می توان توصیفگرهایی مثل حجم مولی و قطبش پذیری را محاسبه کرد.

## ۲-۳-۲- نرم افزار Dragon

این نرم افزار توسط گروه QSAR و کمومتریکس میلانو طراحی شده است [۴۶] و برای محاسبه‌ی توصیف‌کننده های مولکولی استفاده می شود که این توصیف‌کننده‌ها برای ارزیابی روابط ساختار-فعالیت بکار می رود. این نرم افزار قادر است ۱۴۸۱ توصیفگر مختلف را محاسبه کند که به ۱۸ دسته‌ی اصلی تقسیم می شوند که در جدول (۱-۲) ذکر شده اند.

جدول (۱-۲)- 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 آروماتیسسته	۱۷	توصیفگرهای تجربی
۹	پروفایل های مولکولی راندیک	۱۸	خصوصیات مولکولی

برای اجرای این کاربر، فایل های ساختار مولکولی که قبلاً توسط سایر نرم افزارهای مدلسازی مولکولی مانند Hyperchem ایجاد شده اند نیاز است. Dragon یک فایل خروجی کامل ایجاد می کند که به سادگی بوسیله هر نرم افزار آنالیز همبستگی قابل کاربرد است [۴۷].

## ۲-۳-۳- نرم افزار SPSS

عمومی ترین نرم افزار آماری است که قادر به تجزیه و تحلیل کلی اطلاعات است و دارای توانایی های بسیاری است که در زیر ذکر شده است [۴۸]:

تهیه ی انواع جداول مانند جداول فراوانی، فراوانی تجمعی و .....

محاسبه ی انواع توابع ریاضی مانند لگاریتم، قدرمطلق و .....

تهیه ی انواع طرح های آماری

ایجاد داده های تصادفی و پیوسته

قابلیت مبادله ی اطلاعات با نرم افزارهای دیگر

انجام رگرسیون تک متغیره و چندمتغیره

## ۲-۳-۴- نرم افزار MATLAB

این نرم افزار از جامع ترین و پرکاربردترین نرم افزارهای علمی- محاسباتی است [۴۹]. این نرم افزار امکان حل مسائل تخصصی، مخصوصاً مسائل ماتریسی را فراهم می کند. از کاربردهای این نرم افزار، استفاده از آن در آنالیز آماری و مدل سازی داده های شیمیایی است، که در این تحقیق نیز مدل سازی به روش شبکه عصبی مصنوعی توسط این نرم افزار انجام شده است. در کاربرد این نرم افزار، اطلاعات شیمیایی به صورت ماتریس به نرم افزار داده می شوند. سپس در محیط این



برنامه، با استفاده از فرامین و امکانات ویژه ی موجود، مدل سازی غیرخطی فعالیت شیمیایی ترکیبات

با ساختار آنها انجام می شود.

برای رسم نمودارهای سه بعدی در این تحقیق از نرم افزار sigma plot استفاده شده است.

مدلسازی فعالیت دارویی مشتقات بنزیمیدازول  
به عنوان داروهای ضد انگل

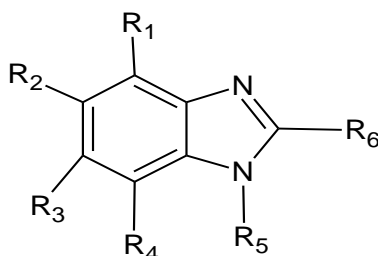
## ۳-۱- محاسبات و مدل‌سازی فعالیت دارویی مشتقات بنزیمیدازول به

### عنوان داروهای ضد انگل

در این تحقیق، از شبکه‌ی عصبی مصنوعی و رگرسیون خطی چندگانه برای مطالعه‌ی رابطه ساختار-فعالیت بمنظور پیش‌بینی فعالیت ضدانگلی یک سری جدید از مشتقات بنزیمیدازول بر ضد انگل گیاردیا اینتستینالیس (G.I) استفاده شده است و توانایی مدل‌های حاصله از این دو روش در پیش‌گویی قدرت بازدارندگی این مشتقات مورد ارزیابی قرار گرفته است.

### ۳-۱-۱- سری داده‌ها

سری داده‌ها که فعالیت دارویی ۴۲ ترکیب از مشتقات بنزیمیدازول (شکل ۳-۱) را در بر دارد، به همراه جزئیات ساختاری آنها در جدول (۳-۱) ارائه شده است.



شکل (۳-۱) - ساختار عمومی مشتقات بنزیمیدازول

در این جدول، استخلافها و مقادیر عددی  $pIC_{50}$  بر حسب مولار برای هر ترکیب مورد بررسی ذکر شده است [۵۳-۵۰]. نصف حداکثر غلظت بازدارندگی ( $IC_{50}$ ) میزان تاثیر یک دارو در بازدارندگی فعالیت بیولوژیکی یا بیوشیمی است. این کمیت نشان می‌دهد که چه مقدار از یک داروی خاص یا ماده‌ی دیگر (بازدارنده) نیاز است تا یک فرآیند ویژه‌ی بیولوژیکی را به نصف کاهش دهد. گاهی آنرا به صورت تابع لگاریتمی و بصورت تابع  $pIC_{50}$  (که برابر با  $-\log IC_{50}$  است) تبدیل می‌کنند که مقادیر

1 - Half maximal inhibitory concentration

بزرگتر آن بیانی از توان بیشتر دارو است. مطابق تعریفمدیریت غذا و دارو،  $IC_{50}$  نمایانگر غلظتی از دارو می باشد که جهت ۵۰٪ اثر بازدارندگی در محیط کشت مصنوعی مورد نیاز است. همانطور که در جدول (۱-۳) مشخص است نوع استخلافها و نحوه ی قرارگیری آنها اطراف اسکلت اصلی مولکول، اثر زیادی روی فعالیت دارویی ترکیب دارد. در این تحقیق مدل‌هایی ارائه شده اند که می‌توانند قدرت بازدارندگی هر ترکیب مورد بررسی را بدون نیاز به سنتز و آزمایش ترکیب موردنظر پیش‌بینی کنند.

جدول (۱-۳) ساختار و مقادیر  $pIC_{50}$  ترکیبات مورد بررسی

NO	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	$pIC_{50}$
۱ <sup>tr</sup>	H	Br	H	H	H	-CF <sub>3</sub>	۶/۱۹
۲ <sup>t</sup>	H	Br	Br	H	H	-CF <sub>3</sub>	۶/۹۲
۳ <sup>tr</sup>	Br	H	Br	H	H	-CF <sub>3</sub>	۶/۶۰
۴ <sup>v</sup>	Br	Br	Br	H	H	-CF <sub>3</sub>	۶/۸۵
۵ <sup>tr</sup>	Br	Br	Br	Br	H	-CF <sub>3</sub>	۷/۲۵
۶ <sup>t</sup>	H	Cl	Cl	H	H	-C <sub>2</sub> F <sub>5</sub>	۶/۲۵
۷ <sup>tr</sup>	H	NO <sub>2</sub>	NO <sub>2</sub>	H	H	-CF <sub>3</sub>	۶/۶۱
۸ <sup>v</sup>	Br	Br	Br	Br	H	-C <sub>2</sub> F <sub>5</sub>	۷/۶۳
۹ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	۷/۶۰
۱۰ <sup>tr</sup>	H	H	Cl	H	Me	-CONH <sub>2</sub>	۷/۱۱
۱۱ <sup>t</sup>	H	H	Cl	H	Me	-CONHCH <sub>3</sub>	۷/۱۴
۱۲ <sup>tr</sup>	H	H	Cl	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۷/۳۹
۱۳ <sup>tr</sup>	H	H	Cl	H	Me	-COOCH <sub>2</sub> CH <sub>3</sub>	۷/۳۱
۱۴ <sup>tr</sup>	H	H	Cl	H	Me	-CH(OH)CH <sub>3</sub>	۷/۱۴
۱۵ <sup>v</sup>	H	H	Cl	H	Me	-COCH <sub>3</sub>	۷/۱۷
۱۶ <sup>t</sup>	H	Cl	H	H	Me	-CONH <sub>2</sub>	۶/۶۳
۱۷ <sup>tr</sup>	H	Cl	H	H	Me	-CONHCH <sub>3</sub>	۶/۴۴
۱۸ <sup>tr</sup>	H	Cl	H	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۶/۶۱

t: test set, tr: train set, v: validation set

NO	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	pIC <sub>50</sub>
۱۹ <sup>v</sup>	H	Cl	H	H	Me	-CO <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	۷/۴۰
۲۰ <sup>tr</sup>	H	Cl	H	H	Me	-CH(OH)CH <sub>3</sub>	۷/۴۲
۲۱ <sup>t</sup>	H	Cl	H	H	Me	-COCH <sub>3</sub>	۷/۲۹
۲۲ <sup>tr</sup>	H	Cl	Cl	H	Me	-CONH <sub>2</sub>	۶/۳۳
۲۳ <sup>tr</sup>	H	Cl	Cl	H	Me	-CONHCH <sub>3</sub>	۶/۸۲
۲۴ <sup>v</sup>	H	Cl	Cl	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۷/۱۳
۲۵ <sup>tr</sup>	H	Cl	Cl	H	Me	-CO <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	۷/۵۶
۲۶ <sup>t</sup>	H	Cl	Cl	H	Me	-CH(OH)CH <sub>3</sub>	۷/۶۹
۲۷ <sup>tr</sup>	H	Cl	Cl	H	Me	-COCH <sub>3</sub>	۷/۴۵
۲۸ <sup>tr</sup>	H	H	H	H	Me	-CONH <sub>2</sub>	۷/۰۲
۲۹ <sup>v</sup>	H	H	H	H	Me	-CONHCH <sub>3</sub>	۶/۲۱
۳۰ <sup>tr</sup>	H	H	H	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۶/۲۹
۳۱ <sup>t</sup>	H	H	H	H	Me	-CO <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	۷/۱۶
۳۲ <sup>tr</sup>	H	H	H	H	Me	-CH(OH)CH <sub>3</sub>	۷/۱۶
۳۳ <sup>tr</sup>	H	H	H	H	Me	-COCH <sub>3</sub>	۷/۰۶
۳۴ <sup>tr</sup>	H	H	Cl	H	H	-CF <sub>3</sub>	۷/۱۰
۳۵ <sup>v</sup>	H	H	H	H	Me	-CF <sub>3</sub>	۷/۳۷
۳۶ <sup>tr</sup>	H	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> OH	۷/۴۲
۳۷ <sup>t</sup>	H	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> OH	۶/۹۷
۳۸ <sup>tr</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(CH <sub>2</sub> ) <sub>3</sub>	۷/۶۹
۳۹ <sup>tr</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	۷/۷۴
۴۰ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	۷/۴۶
۴۱ <sup>v</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	۷/۱۴
۴۲ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	۷/۶۹

t:test set, tr:train set, v:validation set

### ۳-۱-۲- رسم و بهینه سازی ساختار هندسی مولکولها

در مطالعات QSAR بعد از جمع‌آوری داده‌ها، رسم و بهینه‌سازی ساختار مولکولها انجام می‌شود. به این منظور، ابتدا ساختار مولکولی ترکیبات در نرم افزار Hyperchem 7.0 رسم شده و سپس ساختارنهایی آنها به روش نیمه تجربی و با استفاده از روش AM1 بهینه گردید. بهینه سازی تا زمانیکه گرادیان انرژی به  $0/001$  کیلوکالری بر مول رسید ادامه یافت.

### ۳-۱-۳- محاسبه توصیفگرهای مولکولی

توصیفگرهای مولکولی مقادیر عددی هستند که ویژگیهای ساختاری مولکول را توصیف می‌کنند. در این مرحله، ساختارهای بهینه شده در Hyperchem بعنوان ورودی به نرم‌افزار Dragon وارد گردید. به وسیله این نرم‌افزار توصیفگرهای مربوط به این ترکیبات شامل ۱۴۸۱ توصیفگر در ۱۸ گروه مختلف، محاسبه شد.

### ۳-۱-۴- انتخاب بهترین توصیف‌کننده‌ها

این مرحله از مهمترین مراحل در QSAR است که در آن توصیف‌کننده‌های بدست آمده از نرم-افزار Dragon وارد نرم‌افزار SPSS شدند. سپس به منظور انتخاب توصیفگرهای مهمی که ارتباط بیشتری بین ساختار مولکول و فعالیت دارویی ایجاد کرده اند روش کاهش متغیر بکار برده شد. در این روش، ابتدا توصیفگرهای دارای مقادیر یکسان از مجموعه‌ی توصیفگرها حذف شدند، سپس با توجه به این که نباید متغیرها در روش رگرسیون خطی چندگانه با یکدیگر همبستگی داشته باشند، از میان جفت توصیفگری که ضریب همبستگی بالاتر از  $0/90$  داشتند یکی حذف گردید. این کار برای هر گروه از توصیفگرها توسط نرم افزار SPSS انجام شد. در مرحله‌ی بعد، توسط نرم‌افزار SPSS، رگرسیون مرحله به مرحله برای تمام مولکولها و توصیف‌کننده‌های باقیمانده انجام گردید. در

رگرسیون مرحله به مرحله، متغیرها یکی پس از دیگری وارد مدل می‌شدند. ابتدا متغیری که بیشترین همبستگی را با متغیر وابسته داشت وارد مدل شد. با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آنها سطح معنی داری خود را از دست داده بودند، قبل از ورود متغیر جدید از مدل خارج می‌شدند. بدین منظور، داده‌های pIC<sub>50</sub> به عنوان متغیر وابسته و توصیف-کننده ها به عنوان متغیر مستقل در نظر گرفته شدند و بعد از انجام رگرسیون مرحله ای، تنها ۱۴ توصیفگر با داده های pIC<sub>50</sub> مولکولهای مورد بررسی ارتباط معنی داری داشتند، که این توصیفگرها به همراه کلاس مربوطه در جدول (۲-۳) آمده است.

جدول (۲-۳) - کل توصیفگرهای انتخاب شده بوسیله ی رگرسیون خطی چندگانه مرحله به مرحله

No	Symbol	Class	Meaning
۱	RDF070e	RDF	Radial Distribution Function -7.0
۲	RDF075e	RDF	Radial Distribution Function -7.5
۳	MATS8p	2D autocorrelation	Moran autocorrelation -lag8
۴	BEHm8	BCUT	Highest eigenvalue n.8 of Burden matrix
۵	R3u_A	GETAWAY	R maximal autocorrelation of lag 3
۶	MATS7e	2D autocorrelation	Moran autocorrelation -lag 7
۷	R4u	GETAWAY	R autocorrelation of lag4
۸	RDF110m	RDF	Radial Distribution Function -11.0
۹	G2v	WHIM	2st component symmetry directional WHIM index
۱۰	MATS3e	2D autocorrelation	Moran autocorrelation -lag 3
۱۱	R4e_A	GETAWAY	R maximal autocorrelation of lag 4
۱۲	jGL5	Galveztopological.charge indices	Mean topological charge index of order 5
۱۳	H8m	GETAWAY	H autocorrelation of lag8
۱۴	R7V_A	GETAWAY	R maximal autocorrelation of lag 7

e:weighted by atomic sanderson electronegativities , p:weighted by atomic polarizabilities  
u:unweighted, m:weighted by atomic masses

### ۳-۱-۵- بدست آوردن مدل نهایی به روش رگرسیون خطی چند گانه (MLR)

با استفاده از رگرسیون خطی چند گانه مرحله به مرحله از میان تعداد زیادی توصیفگر، ۱۴ توصیفگر (مدل) معنی دار تشخیص داده شد. برای انتخاب بهترین مدل از میان مدل‌های ایجاد شده، معمولاً مدلی که تعداد توصیفگر کمتر، ضریب همبستگی بالاتر، خطای کمتر و قدرت پیش بینی بیشتری

داشته باشد انتخاب می‌شود. برای این کار داده‌ها به طور تصادفی به سه سری آموزش، ارزیابی و تست تقسیم شدند. سری تست و ارزیابی هر کدام ۸ ترکیب و سری آموزش ۲۶ ترکیب را شامل می‌شوند. سپس، توسط سری آموزش، ابتدا با ۲ توصیفگر اول انتخاب شده، به روش اجباری مدل ساخته شد و به وسیله معادله 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، ۵، ۶، ۷ تا ۱۴ توصیفگر به طور مجزا سری ارزیابی مورد بررسی قرار گرفت. میانگین مربعات خطا (mse) و ضریب همبستگی (R) محاسبه شده برای سری ارزیابی در جدول (۳-۳) آمده است. مدل ۵ و ۷ مقدار R و MSE بسیار نزدیک به هم دارند اما با توجه به این که در شبکه عصبی که در بخش (۳-۱-۷-۲) ذکر می‌شود ۷ توصیفگر به عنوان بهترین توصیفگرها انتخاب شده‌اند در اینجا نیز جهت مقایسه‌ی بهتر دو روش، با در نظر گرفتن مقدار mse (جدول (۳-۳)) مدل ۷ با ۷ توصیفگر، که کمترین مقدار خطا را در سری ارزیابی ایجاد نموده، انتخاب گردید.

جدول (۳-۳) - مقادیر mse و R محاسبه شده برای سری ارزیابی برای مدل های ۱۴-۲

شماره مدل	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
mse	۰/۱۲۸	۰/۱۷۲	۰/۰۹۲۶	۰/۰۳۸۷	۰/۰۴۵۳	۰/۰۳۸۱	۰/۰۷۸۵	۰/۱۴۲	۰/۱۹۴	۰/۰۸۷۷	۰/۰۷۰۳	۰/۰۴۳۷	۰/۰۴۶۲
R	۰/۴۹۸	۰/۳۶۶	۰/۶۶۹	۰/۸۸۳	۰/۸۵۹	۰/۸۷۷	۰/۷۶۸	۰/۶۳۰	۰/۵۳۸	۰/۷۵۸	۰/۷۹۴	۰/۸۷۲	۰/۸۶۹

به این ترتیب، معادله ی خطی نهایی به صورت زیر بدست آمد:

$$\text{PIC}_{50} = 4.26 + 0.199\text{RDF070e} - 0.183\text{RDF075e} + \text{رابطه (۳-۱)}$$

$$0.280\text{MATS8p} + 0.458\text{BEHm8} + 3.82\text{R3u\_A} + 0.211\text{MATS7e} + 0.694\text{R4u}$$

مقادیر mse و R برای سری تست نیز برای مدل ۷ بترتیب ۰/۸۴۱ و ۰/۴۷۶ بدست آمده است.



جدول (۳-۴) ضریب رگرسیون، گروه، نام کامل توصیف کننده ها و اثر متوسط آنها در مدل منتخب را نشان می دهد. اثر متوسط یک متغیر مستقل با استفاده از رابطه (۳-۲) محاسبه گردید:

$$\text{MeanEffect} = \sum x_n \frac{\beta_x}{\sum \text{PIC50}} \quad \text{رابطه (۳-۲)}$$

که  $\beta_x$  ضریب متغیر مستقل  $x$  در مدل MLR،  $x_n$  مقدار متغیر مستقل  $x$  مورد نظر برای ترکیب  $n$  ام و  $\text{PIC}_{50}$  غلظت بازداری تجربی می باشد.

جدول (۳-۴) - توصیف کننده های وارد شده در مدل MLR منتخب

شماره	علامت	ضریب رگرسیون	اثر متوسط	گروه	نام کامل
۱	RDF070e	۰/۱۹۹	۰/۱۳۸	RDF	Radial Distribution Function -7.0
۲	RDF075e	-۰/۱۸۳	-۰/۰۹۴	RDF	Radial Distribution Function -7.5
۳	MATS8p	۰/۲۸۰	-۰/۰۰۴	2D autocorrelation	Moran autocorrelation -lag 8
۴	BEHm8	۰/۴۵۸	۰/۱۲۸	BCUT	Highest eigenvalue n.8 of Burden matrix
۵	R3u_A	۳/۸۲	۰/۱۲۷	GETAWAY	R maximal autocorrelation of lag 3
۶	MATS7e	۰/۲۱۱	۰/۰۰۴	2D autocorrelation	Moran autocorrelation -lag 7
۷	R4u	۰/۶۹۴	۰/۱۰۰	GETAWAY	R autocorrelation of lag4

به منظور بررسی همبستگی بین توصیفگرها، ماتریس همبستگی آنها با استفاده از مقادیر عددی شان به روش همبستگی دو متغیره<sup>۱</sup> بدست آمد که در جدول (۳-۵) نشان داده شده است. این ماتریس عدم وجود همبستگی بین توصیفگرهای منتخب در مدل را نشان می دهد. همانطور که ملاحظه می شود مقدار همبستگی بین پارامترها قابل قبول (کمتر از ۰/۹) می باشد.

جدول (۳-۵) - ماتریس همبستگی توصیفگرهای منتخب

	RDF070e	RDF075e	MATS8p	BEHm8	R3u_A	MATS7e	R4u
RDF070e	۱						
RDF075e	۰/۸۵۱	۱					
MATS8p	۰/۰۴۸	۰/۲۱۳	۱				

BEHm8	۰/۵۴۳	۰/۵۵۳	-۰/۲۹۷	۱			
R3u_A	-۰/۰۸۳	-۰/۰۵۳	-۰/۲۶۳	-۰/۰۲۸	۱		
MATS7e	-۰/۱۸۷	-۰/۲۸۲	-۰/۰۰۹	-۰/۰۶۵	-۰/۲۲۸	۱	
R4u	۰/۳۷۲	۰/۵۴۷	۰/۰۷۳	۰/۵۲۶	-۰/۲۲۹	-۰/۳۲۸	۱

مقادیر عددی توصیف کننده های وارد شده در مدل MLR نهایی در جدول (۳-۶) آورده شده است.

جدول (۳-۶) - مقادیر عددی توصیفگرهای وارد شده در مدل MLR نهایی

شماره ترکیب	RDF070e	RDF075e	MATS8p	BEHm8	R3u_A	MATS7e	R4u
1	0/55	1/32	0/00	1/58	0/26	-0/38	0/82
2	0/13	0/00	0/00	1/80	0/26	0/78	0/80
3	2/18	0/65	0/00	1/81	0/24	-0/53	0/78
4	0/66	0/00	0/00	1/93	0/23	0/64	0/78
5	1/85	0/00	0/00	2/26	0/22	0/51	0/75
6	1/57	1/75	-2/00	1/82	0/35	-0/74	0/91
7	0/69	1/97	0/90	1/80	0/21	-0/38	1/07
8	3/42	0/41	-2/00	2/51	0/34	-0/38	0/83
9	8/41	6/50	0/00	2/38	0/24	-0/55	1/69
10	4/42	1/37	0/00	1/80	0/17	0/70	0/95
11	4/06	2/58	0/00	1/87	0/20	0/52	0/81
12	3/83	3/52	0/00	2/35	0/18	0/44	1/44
13	9/32	6/05	1/00	1/96	0/27	0/41	0/71
14	2/62	2/13	0/00	1/92	0/28	0/39	1/16
15	4/86	2/18	0/00	1/83	0/26	0/52	0/92
16	4/13	1/54	0/00	1/84	0/16	0/25	0/95
17	3/52	3/59	0/00	1/95	0/20	0/30	0/80
18	3/01	4/31	0/00	2/24	0/18	0/30	1/44
19	10/49	6/82	1/00	2/00	0/27	0/25	0/70
20	4/01	1/90	0/00	1/98	0/33	0/14	1/15
21	5/64	0/72	0/00	1/89	0/19	0/18	0/83
22	4/89	2/06	-2/00	1/94	0/16	0/81	0/87
23	4/83	3/83	1/00	2/24	0/19	0/65	0/80
24	7/43	6/45	3/0	2/35	0/14	0/56	1/27
25	10/11	7/40	0/30	2/27	0/27	0/61	0/70
26	3/69	0/68	-2/00	2/12	0/33	0/33	1/14
27	5/46	0/86	-2/00	2/05	0/18	0/50	0/82

28	3/11	1/94	1/00	1/32	0/17	-0/18	0/90
29	3/4	3/45	0/00	1/4	0/19	0/01	0/83
30	5/88	6/78	0/00	1/99	0/15	0/05	1/30

ادامه جدول (۳-۶)

شماره ترکیب	RDF070e	RDF075e	MATS8p	BEHm8	R3u_A	MATS7e	R4u
31	8/63	7/65	0/00	1/87	0/27	-0/08	0/72
32	3/45	3/18	0/90	1/34	0/33	0/01	1/17
33	3/63	1/13	1/00	1/32	0/19	-0/03	0/84
34	0/65	0/41	0/00	1/73	0/26	1/57	0/83
35	3/96	0/83	0/00	1/58	0/23	-0/05	0/98
36	5/49	5/81	0/00	1/98	0/37	-0/04	0/72
37	5/52	5/79	0/00	2/06	0/37	0/01	0/71
38	8/60	7/02	0/00	2/35	0/24	-0/44	1/70
39	11/27	8/83	0/00	2/42	0/21	-0/30	1/45
40	10/01	9/18	0/00	2/63	0/20	-0/35	1/44
41	8/86	10/56	0/00	2/39	0/21	-0/09	1/73
42	8/84	10/31	0/00	2/57	0/21	-0/07	1/72

### ۳-۱-۶- مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN)

از آنجا که ممکن است ارتباط بین متغیرهای مستقل و متغیر وابسته (بازدارندگی ترکیبات) خطی نباشد، لذا شبکه عصبی مصنوعی برای مدل سازی غیر خطی استفاده گردید. شبکه عصبی انتخاب شده در این مطالعه، یک شبکه‌ی پیشخور با الگوریتم آموزشی پس‌انتشار می باشد که الگوریتم آن با استفاده از نرم‌افزار MATLAB نسخه‌ی 2009Rb نوشته شد. در روش شبکه عصبی مصنوعی، تابع کارایی شبکه، میانگین مربع خطا (mse) می‌باشد. همانند روش MLR داده‌ها بصورت تصادفی و با نسبت ۲۰:۲۰:۶۰ به سه سری آموزش، ارزیابی و تست تقسیم شدند. سری آموزش برای آموزش شبکه و سری ارزیابی برای بررسی قدرت تعمیم پذیری مدل بدست آمده از سری آموزش بکار گرفته شد.

در طی فرایند آموزش و ارزیابی، میانگین مربع خطا طبق رابطه (۳-۳) محاسبه گردید و معیار بهینه-سازی، به حداقل رساندن مقدار mse در نظر گرفته شد.

$$mse = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad \text{رابطه (۳-۳)}$$

در این رابطه  $\hat{y}_i$  مقدار پیش‌بینی شده توسط شبکه،  $y_i$  مقدار تجربی فعالیت دارویی و  $n$  تعداد کل ترکیبات سری ارزیابی است.

### ۳-۱-۷- بهینه سازی پارامترهای شبکه عصبی

برای بدست آوردن بهترین مدل و کمترین خطا، پارامترهای شبکه (تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه، تعداد گره ها در لایه ی پنهان و تعداد دوره‌های آموزش) بهینه شدند.

#### ۳-۱-۷-۱- انتخاب تعداد لایه های شبکه

هر شبکه دارای حداقل یک لایه ورودی، یک لایه خروجی و تعدادی لایه ی پنهان می باشد. متأسفانه هیچ راهنمای مناسبی برای انتخاب تعداد لایه های پنهان وجود ندارد و ساختار شبکه‌های عصبی، اغلب به روش سعی و خطا ایجاد می شود. با این وجود در بیشتر موارد یک لایه پنهان مناسب به نظر می رسد [۵۴ و ۵۵]. بنابراین در این تحقیق نیز یک لایه پنهان در نظر گرفته شده است.

#### ۳-۱-۷-۲- انتخاب نوع الگوریتم آموزشی ، نوع تابع انتقال، تعداد متغیرهای

ورودی و تعداد گره های لایه ی پنهان

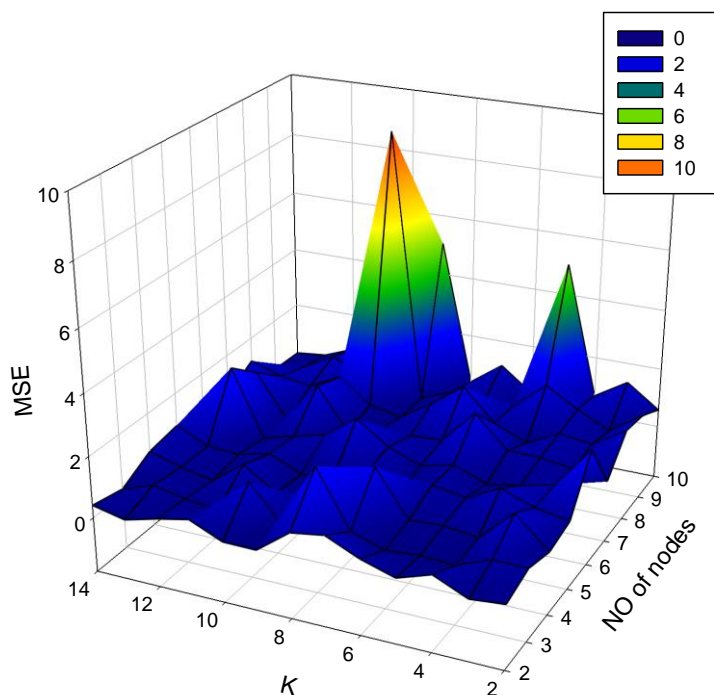
شبکه عصبی با استفاده از نرم‌افزار مطلب طراحی و بهینه شد. پس از طراحی اولیه شبکه، برای اینکه شبکه قادر به پیش بینی صحیحی باشد، باید آموزش ببیند. در جعبه ابزار<sup>۱</sup> نرم‌افزار مطلب، الگوریتم‌های آموزشی مختلفی وجود دارد که در این تحقیق دو الگوریتم آموزشی لونبرگ - مارکوارت و تنظیم بایزین مورد بررسی قرار گرفت. با توجه به اینکه خروجی‌های شبکه عصبی داده‌های کمی  $\log(1/IC_{50})$  می‌باشند، لذا تابع انتقال خطی برای لایه‌ی خروجی انتخاب گردید.

برای بهینه‌سازی و شناسایی بهترین پارامترهای شبکه، شبکه‌هایی با ورودی‌های از ۲ تا ۱۴ توصیفگر ایجاد شدند. هر شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تنظیم بایزین با تعداد گره-های متفاوت در لایه پنهان از ۲ تا ۱۰ آموزش داده شد. در حالی که برای به دست آوردن بهترین تابع انتقال در لایه‌ی پنهان مدل‌های شبکه عصبی، از توابع لگاریتم زیگموئید و تانژانت هایپربولیک به عنوان تابع فعالیت استفاده گردید. در تمام این شبکه‌ها تعداد دوره‌های آموزش و پارامتر  $\mu$  ثابت و به ترتیب ۲۰ و ۰/۰۰۵ در نظر گرفته شدند. در روند بهینه‌سازی شبکه، به حداقل رساندن میانگین مربع خطا به عنوان معیار انتخاب گردید. نتایج محاسبات در جداول (۳-۷)، (۳-۸)، (۳-۹) و (۳-۱۰) 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 در این جداول و شکل‌ها مشاهده می‌شود، شبکه‌های با تابع آموزش trainbr و تابع انتقال logsig با تعداد ۷ توصیفگر و تعداد گره ۵ نتایج بهتری نشان می‌دهد.

جدول (۳-۷) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی 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	۱۲	۱۳	۱۴	
	۲	۰/۲۸۷۸	۰/۱۸۱۳	۰/۶۷۹۹	۰/۳۱۴۵	۰/۶۵۴۶	۱/۱۴۹۶	۰/۹۸۲۱	۰/۱۷۷۳	۰/۱۹۰۸	۰/۶۶۹۰	۰/۴۴۷۵	۰/۱۷۲۵	۰/۴۱۱۵
	۳	۰/۵۶۱۳	۰/۵۸۳۹	۰/۱۴۸۸	۰/۲۰۴۲	۰/۱۲۰۰	۰/۴۷۲۶	۲/۳۱۲۳	۰/۱۷۱۲	۱/۳۶۱۱	۰/۱۹۸۲	۰/۳۳۶۹	۰/۵۵۷۳	۰/۱۹۶۰
	۴	۰/۲۵۹۷	۱/۲۵۰۹	۰/۲۱۷۰	۰/۳۶۸۲	۰/۲۹۸۰	۱/۷۹۱۴	۰/۴۳۱۷	۰/۳۳۰۳	۰/۲۴۷۳	۰/۹۶۲۴	۰/۱۸۹۱	۰/۴۲۰۱	۰/۷۲۵۰
	۵	۰/۴۵۴۸	۰/۶۵۷۷	۱/۰۴۶۹	۰/۵۰۰۹	۰/۶۹۳۱	۰/۱۷۳۵	۰/۶۲۶۳	۰/۲۴۴۴	۰/۱۲۹۹	۰/۶۰۶۶	۰/۳۵۰۳	۰/۴۲۱۴	۰/۹۰۸۲
	۶	۲/۰۱۹۲	۰/۶۴۵۲	۰/۳۹۶۶	۰/۲۰۹۲	۰/۶۳۰۸	۰/۳۶۶۹	۰/۷۸۸۱	۱/۵۲۶۷	۰/۷۵۲۷	۰/۴۹۴۲	۰/۱۹۷۴	۲/۴۲۳۶	۰/۳۱۱۶
	۷	۰/۲۴۴۸	۰/۴۳۷۰	۰/۵۹۶۰	۰/۶۱۴۹	۱/۶۵۰۶	۰/۵۳۴۸	۰/۴۵۹۶	۰/۴۶۰۵	۰/۱۴۲۵	۱/۵۵۶۴	۰/۱۱۹۹	۱/۶۱۰۲	۰/۶۴۳۸
	۸	۱/۲۴۷۶	۰/۵۵۷۷	۰/۷۱۳۳	۰/۱۵۱۸	۰/۶۳۱۴	۱/۶۸۳۳	۰/۸۹۴۱	۰/۲۰۱۱	۰/۴۲۷۵	۱/۲۲۷۰	۰/۶۸۸۱	۰/۸۴۰۲	۱/۱۹۰۵
	۹	۱/۰۸۳۰	۰/۳۱۹۶	۰/۸۴۹۲	۱/۴۷۸۶	۰/۵۵۹۸	۰/۴۱۴۶	۰/۶۳۴۹	۰/۲۰۱۹	۹/۱۰۶۱	۰/۷۳۹۰	۱/۰۹۰۵	۰/۵۲۶۹	۰/۲۶۴۳
	۱۰	۰/۵۷۸۱	۱/۳۳۹۵	۰/۶۶۳۷	۵/۰۱۴۸	۰/۳۳۷۷	۱/۱۳۵۶	۰/۳۴۹۰	۵/۰۵۶۴	۱/۱۴۲۷	۰/۷۰۱۹	۰/۷۵۱۴	۰/۳۳۶۸	۰/۲۲۳۳

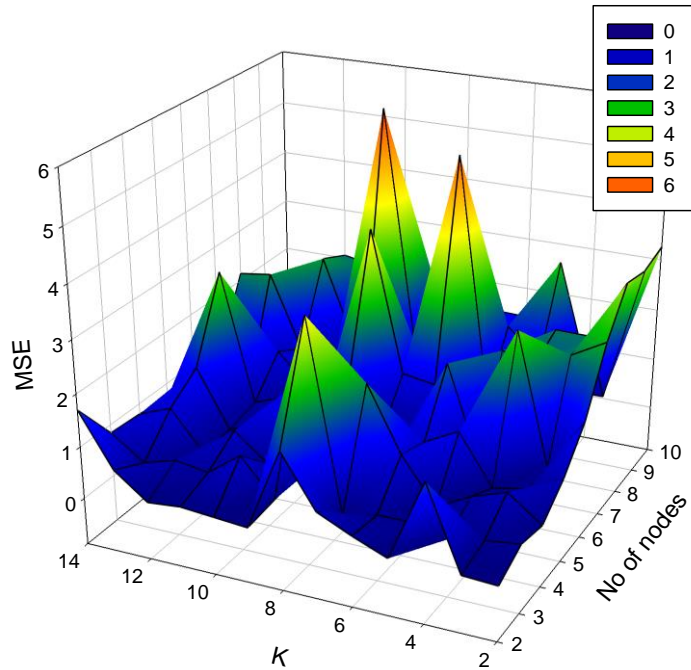


شکل (۳-۲) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش لونیبرگ- ماکوارت و تابع انتقال لگاریتم سیگموئیدی

جدول (۳-۸) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی برای شبکه با الگوریتم آموزشی لونیبرگ - ماکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)

	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
۲	۰/۱۴۳۶	۰/۱۸۱۲	۱/۷۳۱۰	۰/۱۸۴۴	۰/۴۴۹۵	۰/۷۵۱۹	۱/۷۳۰۷	۰/۱۵۸۴	۰/۲۱۰۷	۰/۲۶۴۴	۰/۲۰۵۹	۰/۶۹۰۲	۱/۷۰۱۶
۳	۰/۴۱۹۲	۰/۳۳۲۴	۰/۲۰۶۸	۰/۳۷۷۵	۰/۳۶۹۰	۰/۲۱۴۷	۳/۷۶۳۱	۰/۲۲۷۵	۰/۹۳۲۵	۰/۴۷۱۹	۰/۵۷۲۵	۰/۳۸۹۰	۰/۸۴۴۲
۴	۰/۲۸۶۰	۰/۷۶۷۰	۰/۵۴۲۵	۰/۲۰۴۰	۱/۰۶۴۰	۲/۲۷۱۴	۰/۸۷۳۵	۰/۲۹۳۴	۰/۳۴۴۰	۰/۸۵۸۶	۰/۲۹۰۴	۱/۰۳۱۵	۰/۸۰۰۵
۵	۰/۷۱۶۴	۰/۳۷۴۹	۰/۲۴۸۳	۱/۲۰۴۴	۰/۸۷۳۸	۰/۳۹۲۶	۱/۷۴۵۹	۱/۵۹۶۷	۱/۴۲۷۹	۰/۵۷۶۵	۰/۳۲۵۴	۱/۴۳۷۲	۰/۷۱۲۳
۶	۱/۳۶۰۱	۰/۳۳۳۴	۲/۸۴۷۹	۰/۱۳۷۷	۱/۹۸۰۲	۰/۲۸۶۴	۰/۲۸۲۹	۱/۲۲۸۵	۰/۳۰۰۷	۰/۷۰۱۳	۰/۴۶۳۸	۲/۹۵۵۸	۰/۲۷۸۵
۷	۲/۴۷۵۲	۲/۱۱۴۶	۱/۸۵۱۶	۱/۹۶۵۱	۱/۲۲۱۶	۱/۰۴۷۷	۱/۱۴۹۸	۳/۸۵۷۴	۰/۴۲۰۱	۰/۵۰۷۹	۱/۱۱۷۳	۰/۲۴۹۷	۰/۷۳۵۳
۸	۳/۲۳۲۸	۲/۱۰۰۹	۲/۱۱۱۲	۰/۲۳۸۳	۰/۴۶۳۸	۵/۱۲۴۸	۰/۹۰۱۹	۰/۱۱۵۸	۰/۳۵۲۱	۰/۴۴۴۴	۰/۷۹۱۸	۲/۲۲۵۴	۲/۱۴۰۹
۹	۳/۱۱۰۳	۰/۵۰۵۰	۰/۴۹۸۵	۰/۹۹۲۵	۱/۵۸۰۷	۰/۸۴۰۵	۰/۱۹۶۷	۱/۰۱۳۲	۵/۴۷۱۳	۰/۷۱۷۲	۲/۳۱۷۷	۰/۳۸۴۱	۰/۵۷۸۸
۱۰	۳/۲۴۶۲	۱/۸۵۱۴	۰/۴۶۶۶	۲/۶۳۴۱	۱/۴۱۵	۱/۴۳۵۸	۰/۷۵۷۲	۱/۸۲۱۳	۱/۶۸۱۴	۱/۹۴۱۸	۲/۰۵۸۵	۰/۷۵۲۰	۱/۴۲۸۸



شکل (۳-۳) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برایشبکه‌هایی با تابع آموزش لونبرگ-

ماکوارت و تابع انتقال تانژانت سیگموئید

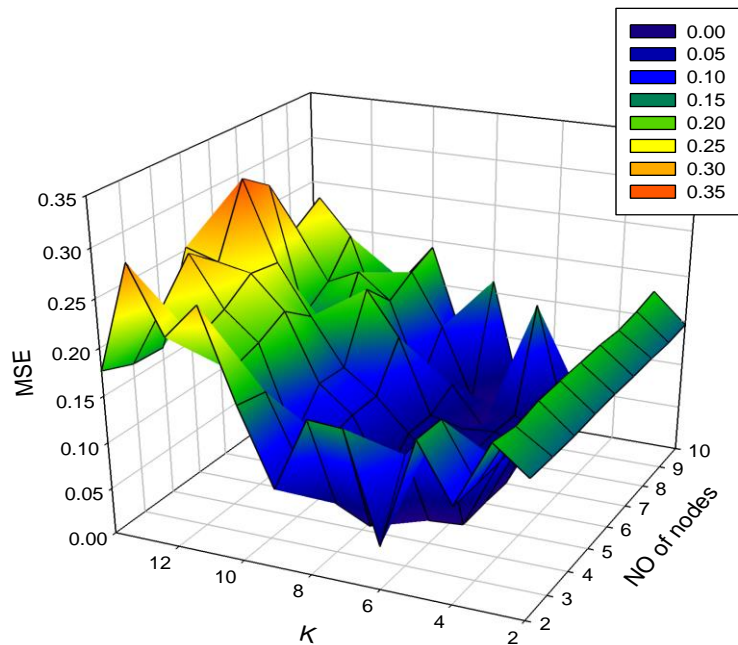
جدول (۳-۹) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع

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

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)

		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
تعداد گره‌های لایه پنهان	۲	۰/۱۴۶۶	۰/۱۷۴۹	۰/۱۰۴۲	۰/۱۶۴۳	۰/۰۴۶۵	۰/۱۶۱۹	۰/۱۶۴۳	۰/۰۸۵۰	۰/۱۶۵۸	۰/۲۶۰۲	۰/۲۲۲۶	۰/۲۹۱۱	۰/۱۷۶۴
	۳	۰/۱۴۵۴	۰/۱۶۹۳	۰/۱۰۳۵	۰/۱۶۳۴	۰/۱۲۰۰	۰/۰۳۵۸	۰/۰۵۳۴	۰/۰۸۳۶	۰/۱۵۹۴	۰/۱۷۵۶	۰/۲۰۱۶	۰/۲۲۷۹	۰/۱۶۴۲
	۴	۰/۱۴۴۸	۰/۱۷۴۳	۰/۱۰۴۳	۰/۰۲۸۵	۰/۰۲۵۶	۰/۰۹۳۸	۰/۰۵۹۶	۰/۰۹۲۹	۰/۰۹۰۴	۰/۱۸۵۳	۰/۲۱۴۸	۰/۲۶۷۵	۰/۱۶۳۶
	۵	۰/۱۴۶۲	۰/۱۷۴۴	۰/۱۰۴۳	۰/۰۲۹۶	۰/۰۴۹۰	۰/۰۱۶۷	۰/۰۳۴۸	۰/۰۹۹۵	۰/۱۲۸۱	۰/۲۰۰۵	۰/۲۳۵۳	۰/۲۴۲۱	۰/۲۵۲۸
	۶	۰/۱۴۶۵	۰/۱۷۳۷	۰/۱۱۷۷	۰/۰۲۸۶	۰/۰۵۶۲	۰/۰۳۰۴	۰/۰۶۶۱	۰/۱۰۰۱	۰/۱۲۶۳	۰/۱۷۷۲	۰/۲۳۰۱	۰/۳۱۴۰	۰/۲۰۱۵
	۷	۰/۱۴۵۴	۰/۱۷۳۳	۰/۱۰۴۸	۰/۰۴۴۲	۰/۰۵۸۱	۰/۰۵۵۰	۰/۰۶۶۱	۰/۱۴۶۹	۰/۱۹۴۹	۰/۱۳۳۷	۰/۱۵۶۲	۰/۲۹۲۴	۰/۲۵۷۶
	۸	۰/۱۴۵۶	۰/۱۷۰۶	۰/۱۰۴۸	۰/۰۳۷۳	۰/۰۵۵۹	۰/۰۳۶۲	۰/۰۷۹۱	۰/۰۷۷۵	۰/۱۶۴۷	۰/۱۹۱۹	۰/۱۷۵۷	۰/۲۳۱۷	۰/۱۶۴۹
	۹	۰/۱۴۴۲	۰/۱۷۳۲	۰/۱۰۴۷	۰/۰۳۷۴	۰/۱۶۳۸	۰/۰۳۶۶	۰/۰۴۴۲	۰/۱۲۲۲	۰/۱۸۰۳	۰/۱۷۶۸	۰/۲۱۱۴	۰/۲۵۰۳	۰/۱۸۹۷
	۱۰	۰/۱۴۶۰	۰/۱۷۸۳	۰/۱۰۴۶	۰/۰۳۶۴	۰/۰۴۶۱	۰/۰۳۰۷	۰/۱۶۳۳	۰/۰۹۱۷	۰/۱۹۲۵	۰/۱۴۹۰	۰/۱۵۵۲	۰/۱۸۲۵	۰/۱۶۵۱



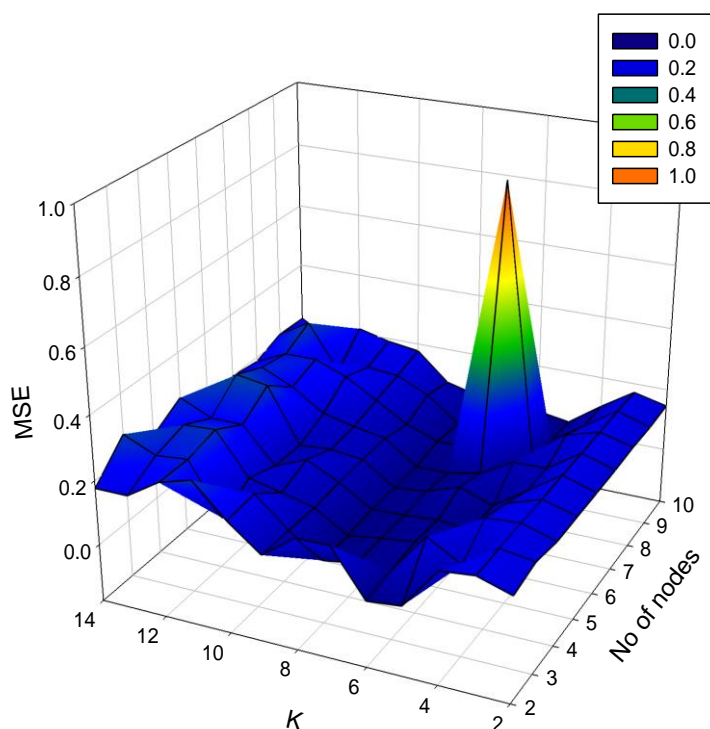


شکل (۳-۴) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

جدول (۳-۱۰) - مقادیر میانگین مربعات خطای حاصل از سری ارزیابی 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)

		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
تعداد گره های لایه پنهان	۲	۰/۱۴۴۱	۰/۱۷۷۲	۰/۱۶۶۱	۰/۰۳۵۸	۰/۰۲۴۹	۰/۱۶۶۶	۰/۱۶۶۰	۰/۰۸۹۵	۰/۱۶۹۶	۰/۲۶۹۴	۰/۲۴۰۳	۰/۱۷۶۷	۰/۱۷۶۵
	۳	۰/۱۶۲۵	۰/۱۷۷۳	۰/۱۰۳۹	۰/۱۶۴۰	۰/۰۳۶۷	۰/۰۳۴۰	۰/۰۱۸۴	۰/۰۸۷۲	۰/۱۷۷۳	۰/۰۷۸۷	۰/۲۳۸۸	۰/۲۲۹۲	۰/۲۷۳۶
	۴	۰/۱۴۳۸	۰/۱۷۶۲	۰/۱۶۸۱	۰/۰۳۱۱	۰/۰۴۷۴	۰/۰۳۳۷	۰/۰۱۸۳	۰/۰۸۲۲	۰/۱۰۷۹	۰/۱۳۱۲	۰/۲۶۲۴	۰/۲۵۶۲	۰/۲۰۱۵
	۵	۰/۱۴۴۶	۰/۱۷۳۲	۰/۱۳۸۰	۰/۰۳۱۵	۰/۰۴۸۴	۰/۰۳۰۷	۰/۰۲۶۵	۰/۰۹۰۳	۰/۱۴۱۰	۰/۰۹۸۹	۰/۳۱۰۹	۰/۲۱۰۷	۰/۲۵۶۱
	۶	۰/۱۴۲۵	۰/۱۷۵۷	۰/۱۰۴۰	۰/۰۳۲۸	۰/۰۷۰۱	۰/۰۳۴۰	۰/۰۲۲۹	۰/۰۹۲۷	۰/۱۰۴۶	۰/۰۹۷۱	۰/۲۷۰۹	۰/۲۸۶۳	۰/۱۷۲۵
	۷	۰/۱۴۲۰	۰/۱۷۴۰	۰/۱۰۳۸	۰/۱۰۰۹	۰/۰۴۳۰	۰/۰۲۲۰	۰/۰۱۹۹	۰/۰۹۲۸	۰/۱۱۱۵	۰/۱۳۴۵	۰/۲۰۲۳	۰/۲۴۷۵	۰/۱۹۹۱
	۸	۰/۱۴۴۹	۰/۱۷۲۸	۰/۱۴۲۳	۰/۰۴۴۰	۰/۰۸۹۸	۰/۰۳۶۲	۰/۰۳۵۶	۰/۰۷۲۴	۰/۱۳۳۰	۰/۱۹۲۳	۰/۲۳۱۱	۰/۲۳۱۰	۰/۱۸۵۸
	۹	۰/۱۴۱۴	۰/۱۷۳۹	۰/۱۰۴۱	۰/۰۳۳۸	۰/۰۴۹۹	۰/۰۳۳۰	۰/۰۲۸۲	۰/۱۰۶۰	۰/۱۵۲۳	۰/۱۶۱۸	۰/۱۲۹۴	۰/۲۷۵۸	۰/۲۱۸۸
	۱۰	۰/۱۴۵۹	۰/۱۷۱۰	۰/۱۰۳۹	۰/۰۲۹۹	۰/۰۴۶۴	۰/۰۳۴۲	۰/۰۶۰۴	۰/۰۸۶۵	۰/۱۷۵۹	۰/۱۹۰۲	۰/۲۱۵۲	۰/۱۵۷۸	۰/۲۱۴۷



شکل (۳-۵) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئید

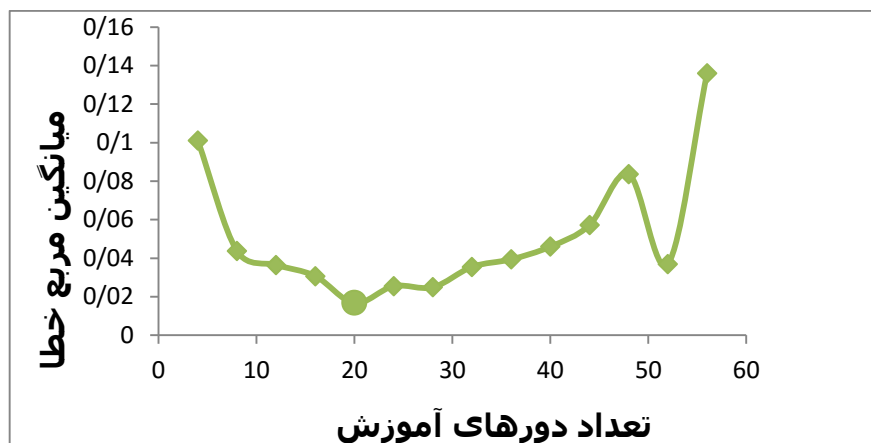
### ۳-۱-۷-۳ - بهینه سازی تعداد دوره‌های آموزشی

برای جلوگیری از خطا و آموزش بیش از حد شبکه، بهینه سازی تعداد دوره‌های آموزشی صورت می‌گیرد. بدین منظور، شبکه با تابع آموزش **trainbr** و تابع انتقال **logsig** با تعداد ورودی ۷ توصیفگر و مقدار بهینه‌ی تعداد نرون‌ها (۵) در تعداد دوره‌های متفاوت با فواصل یکسان آموزش داده شد و میانگین مربع خطا برای سری ارزیابی ثبت گردید (جدول (۳-۱۱)). سپس منحنی میانگین مربع خطای شبکه بر حسب تعداد دوره‌ها رسم گردید (شکل (۳-۷)).

جدول (۳-۱۱) - مقادیر میانگین مربع خطا در تعداد دوره‌های آموزشی متفاوت

تعداد دوره‌های آموزش	میانگین مربع خطا	تعداد دوره‌های آموزش	میانگین مربع خطا
۴	۰/۱۰۱۲	۳۲	۰/۰۳۵۵
۸	۰/۰۴۳۸	۳۶	۰/۰۳۹۴
۱۲	۰/۰۳۶۵	۴۰	۰/۰۴۶۰
۱۶	۰/۰۳۰۷	۴۴	۰/۰۵۷۲

۲۰	۰/۰۱۶۷	۴۸	۰/۰۸۳۷
۲۴	۰/۰۲۵۵	۵۲	۰/۱۰۳۷
۲۸	۰/۰۲۴۹	۵۶	۰/۱۳۶۱



شکل (۳-۶) - نمودار میانگین مربع خطا بر حسب تعداد دوره‌های آموزش متفاوت

همانطور که در جدول (۳-۱۱) و شکل (۳-۶) ملاحظه می‌شود در تعداد دور ۲۰، میانگین مربع خطا کمتر است و این تعداد دور به عنوان تعداد دوره‌های بهینه برای یادگیری شبکه انتخاب شد.

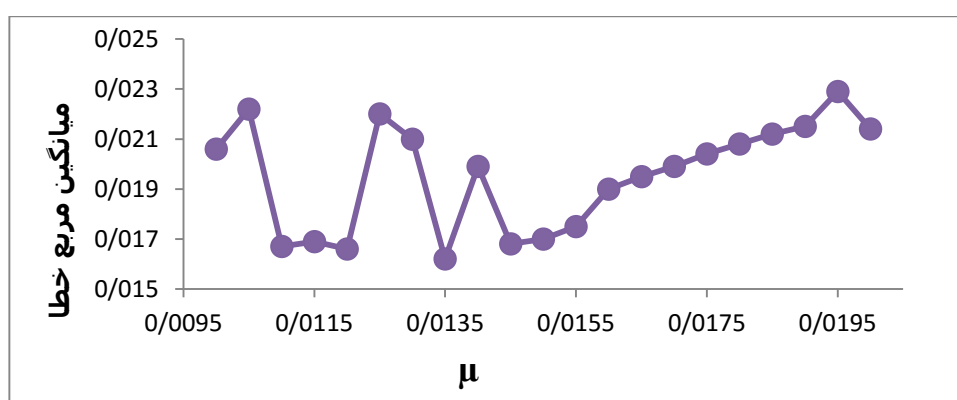
### ۳-۱-۷-۴ - بهینه سازی پارامتر $\mu$

جهت بهینه سازی مقدار  $\mu$  (پارامتر لرنیگ)، شبکه با ۷ متغیر ورودی، تابع آموزش **trainbr**، تابع انتقال **logsig** و تعداد ۵ نرون در لایه‌ی مخفی در نظر گرفته شد و مقدار  $\mu$  از ۰/۰۰۰۵ تا ۰/۰۲ با گام‌های ۰/۰۰۰۵ تغییر داده شد. سپس برای هر مورد مقدار mse سری های ارزیابی محاسبه گردید. نقطه‌ای که کمترین خطا را برای سری ارزیابی داشته باشد به عنوان مقدار بهینه انتخاب می‌گردد. مقدار بهینه  $\mu$ ، ۰/۰۱۳۵ به دست آمد که در رنج ۰/۰۱ تا ۰/۰۲ قرار دارد و به علت زیاد بودن داده‌ها، مقادیر mse برای مقادیر متفاوت  $\mu$  فقط در این رنج، در جدول (۳-۱۲) و شکل (۳-۷) نشان داده شده است.

جدول (۳-۱۲) - مقادیر خطای شبکه با مقادیر  $\mu$

مقادیر $\mu$	میانگین مربع خطا	مقادیر $\mu$	میانگین مربع خطا
۰/۰۱	۰/۰۲۰۶	۰/۰۱۵۵	۰/۰۱۷۵

۰/۰۱۰۵	۰/۰۲۲۲	۰/۰۱۶	۰/۰۱۹۰
۰/۰۱۱	۰/۰۱۶۷	۰/۰۱۶۵	۰/۰۱۹۵
۰/۰۱۱۵	۰/۰۱۶۹	۰/۰۱۷	۰/۰۱۹۹
۰/۰۱۲	۰/۰۱۶۶	۰/۰۱۷۵	۰/۰۲۰۴
۰/۰۱۲۵	۰/۰۲۲۰	۰/۰۱۸	۰/۰۲۰۸
۰/۰۱۳	۰/۰۲۱۰	۰/۰۱۸۵	۰/۰۲۱۲
۰/۰۱۳۵	۰/۰۱۶۲	۰/۰۱۹	۰/۰۲۱۵
۰/۰۱۴	۰/۰۱۹۹	۰/۰۱۹۵	۰/۰۲۲۹
۰/۰۱۴۵	۰/۰۱۶۸	۰/۰۲	۰/۰۲۱۴
۰/۰۱۵	۰/۰۱۷۰		



شکل (۳-۷) - نمودار میانگین مربع خطا بر حسب مقادیر پارامتر  $\mu$

### ۳-۱-۷-۵ - مشخصات و هندسه‌ی شبکه عصبی بهینه

مشخصات شبکه بهینه در جدول (۳-۱۳) ذکر شده است.

جدول (۳-۱۳) - مشخصات شبکه بهینه

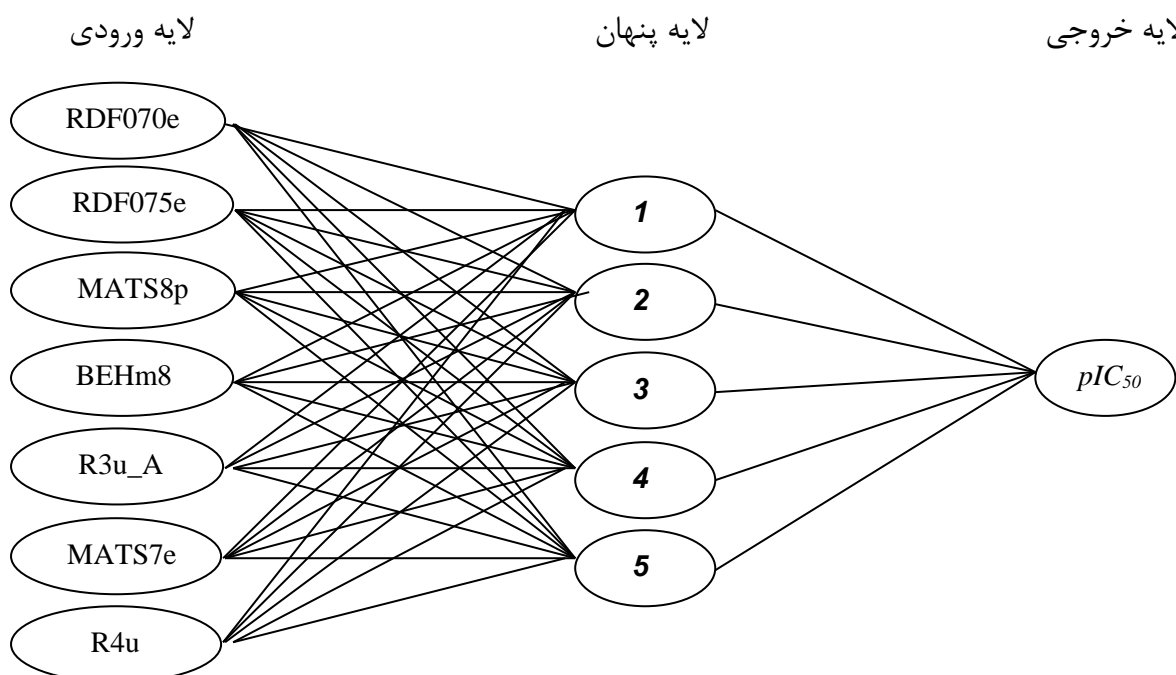
Trainbr	تابع آموزش
Logsig	تابع انتقال لایه پنهان
۵	تعداد نرون لایه پنهان
۷	تعداد متغیرهای ورودی
۲۰	تعداد دوره‌های آموزش
۰/۰۱۳۵	پارامتر $\mu$

طبق نتایج بدست آمده، ۷ توصیفگر بعنوان توصیفگرهایی که بیشترین ارتباط را با فعالیت بازدارندگی دارند، انتخاب شدند که در جدول (۳-۱۴) به همراه طبقه و معنی شان ذکر شده اند.

جدول (۳-۱۴) - توصیفگرهای مورد استفاده در شبکه عصبی

No	Symbol	Class	Meaning
۱	RDF070e	RDF	Radial Distribution Function -7.0
۲	RDF075e	RDF	Radial Distribution Function -7.5
۳	MATS8p	2D autocorrelation	Moran autocorrelation -lag 8
۴	BEHm8	BCUT	Highest eigenvalue n.8 of Burden matrix
۵	R3u_A	GETAWAY	R maximal autocorrelation of lag 3
۶	MATS7e	2D autocorrelation	Moran autocorrelation -lag 7
۷	R4u	GETAWAY	R autocorrelation of lag4

با توجه به نتایج حاصله از بهینه‌سازی شبکه‌ی عصبی پیش‌خور سه‌لایه با الگوی آزمایشی پس‌انتشار با الگوریتم تنظیم بایزین و تابع انتقال لگاریتم زیگموئید انتخاب شد و ساختار آن در شکل (۳-۸) نشان داده شده است.



### ۳-۲- ارزیابی مدل‌های خطی و غیرخطی ارائه شده و مقایسه مدل‌ها

مرحله‌ی اساسی در مطالعات QSAR ارزیابی قدرت پیش‌بینی مدل‌های ارائه شده می‌باشد. قدرت پیش‌بینی مدل، توانایی آن مدل در ارائه‌ی خروجی‌های رضایت‌بخش برای آن دسته از مولکول‌هایی که در سری آموزش وجود ندارند می‌باشد. در اینجا چندین روش به منظور ارزیابی توانایی مدل‌های ارائه شده (که دارای کمترین مقدار mse بودند) در پیش‌بینی داده‌های  $pIC_{50}$  ذکر شده است.

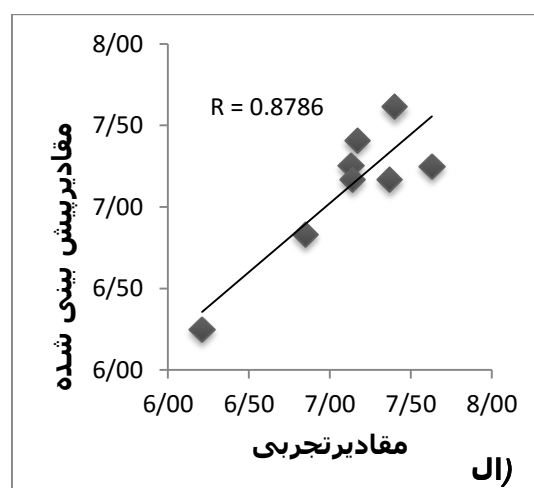
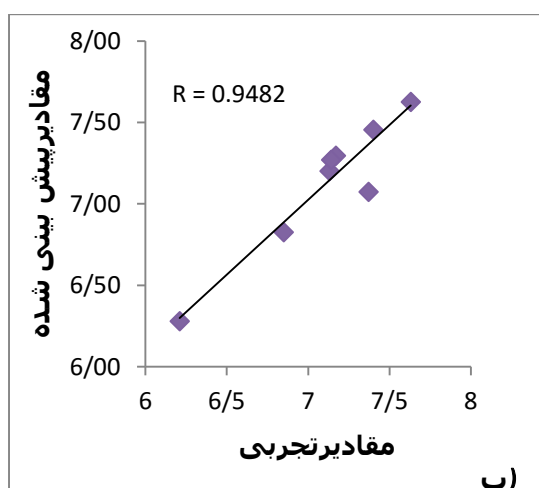
### ۳-۲-۱- ارزیابی مدل‌ها با استفاده از سری ارزیابی

بدین منظور، شبکه‌ی عصبی با ۲۶ مولکول سری آموزش، آموزش داده شد. معیار بهینه‌سازی شبکه، کمترین میانگین مربعات خطا در سری ارزیابی که شامل ۸ مولکول که بطور تصادفی انتخاب شده‌اند، می‌باشد. در مدل رگرسیون خطی چندگانه، از ضرائب مدل برتر برای هر ترکیب سری ارزیابی استفاده شد. نتایج حاصله برای هر دو مدل رگرسیون خطی چندگانه و شبکه عصبی مصنوعی در جدول (۳-۱۵) و ترسیمی از مقادیر پیش‌بینی شده بر حسب مقادیر تجربی برای سری ارزیابی در شکل (۳-۹) آورده شده است.

جدول (۳-۱۵) - نتایج حاصل از مدل رگرسیون خطی چندگانه و شبکه عصبی برای سری ارزیابی

$pIC_{50}$	شماره
------------	-------

ترکیب	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۴	۶/۸۵	۶/۸۳	۶/۸۳	-۰/۲۹	-۰/۲۹
۸	۷/۶۳	۷/۲۵	۷/۶۳	-۴/۹۸	۰/۰۰
۱۵	۷/۱۷	۷/۴۱	۷/۳۰	۳/۳۵	۱/۸۱
۱۹	۷/۴۰	۷/۶۱	۷/۴۶	۲/۸۴	۰/۸۱
۲۴	۷/۱۳	۷/۲۵	۷/۲۰	۱/۶۸	۰/۹۸
۲۹	۶/۲۱	۶/۲۵	۶/۲۸	۰/۶۴	۱/۱۳
۳۵	۷/۳۷	۷/۱۷	۷/۰۷	-۲/۷۱	-۴/۰۷
۴۱	۷/۱۴	۷/۱۷	۷/۲۷	۰/۴۲	۱/۸۲



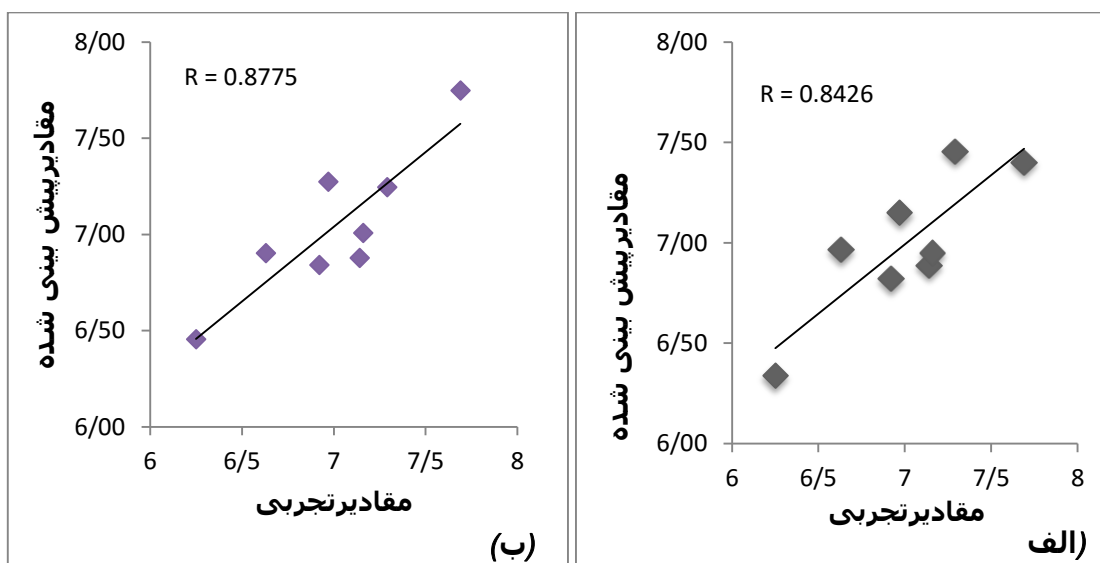
شکل (۳-۹) - نمودار مقادیر پیش بینی شده  $IC_{50}$  بر حسب مقادیر تجربی برای سری ارزیابی توسط الف) روش MLR (ب) شبکه عصبی مصنوعی

### ۳-۲-۲ - ارزیابی مدل ها با استفاده از سری تست

بعد از انتخاب شبکه عصبی بهینه با استفاده از سری ارزیابی، قدرت پیش بینی شبکه، بوسیله ۸ داده که در سری تست قرار دارند مورد بررسی قرار گرفت. در رگرسیون خطی چندگانه نیز از ضرائب مدل برتر برای هر ترکیب سری تست استفاده شد. نتایج حاصل در جدول (۳-۱۶) و شکل (۳-۱۰) نشان داده شده است که بیانگر برتری مدل شبکه عصبی بر مدل رگرسیون خطی چندگانه می باشد.

جدول (۳-۱۶) - نتایج حاصل از مدل رگرسیون خطی چندگانه و شبکه عصبی برای سری تست

شماره ترکیب	pIC <sub>50</sub>				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۲	۶/۹۲	۶/۸۲	۶/۸۴	-۱/۴۵	-۱/۱۶
۶	۶/۲۵	۶/۳۴	۶/۴۶	۱/۴۴	۳/۳۶
۱۱	۷/۱۴	۶/۸۹	۶/۸۸	-۳/۵۰	-۳/۶۴
۱۶	۶/۶۳	۶/۹۷	۶/۹۰	۵/۱۳	۴/۰۷
۲۱	۷/۲۹	۷/۴۶	۷/۲۵	۲/۳۳	-۰/۵۵
۲۶	۷/۶۹	۷/۴۰	۷/۷۵	-۳/۷۷	۰/۷۸
۳۱	۷/۱۶	۶/۹۵	۷/۰۱	-۲/۹۳	-۲/۰۹
۳۷	۶/۹۷	۷/۱۵	۷/۲۷	۲/۵۸	۴/۳۰



شکل (۳-۱۰) - نمودار مقادیر پیش بینی شده pIC<sub>50</sub> بر حسب مقادیر تجربی برای سری تست توسط الف) روش MLR ب) شبکه عصبی مصنوعی

### ۳-۲-۳- ارزیابی مدلها به روش رد مرحله‌ای تک تک داده‌ها

جهت ارزیابی بیشتر شبکه عصبی بهینه و مدل رگرسیون خطی چندگانه، روش رد مرحله‌ای تک تک مورد استفاده قرار گرفت. در این روش در شبکه عصبی، ابتدا یک مولکول از داده‌ها کنارگذاشته شد و شبکه عصبی در شرایط بهینه با ۴۱ ترکیب باقیمانده آموزش داده شد. سپس شبکه



آموزش داده شده برای پیش بینی  $pIC_{50}$  ترکیب کنار گذاشته شده، بکار گرفته شد. در مدل رگرسیون خطی چندگانه، نیز ابتدا یک مولکول از داده ها کنار گذاشته شد و با ۴۱ مولکول باقیمانده، به روش ورود اجباری<sup>۱</sup>، برای پیش بینی  $pIC_{50}$  ترکیب کنار گذاشته شده، ضرائب مدل به دست آمد. این کار برای تک تک داده ها انجام شد. نتایج حاصل در جدول (۳-۱۷) و شکل (۳-۱۱) نشان داده شده است. نتایج نشان می دهد که مدل های ارائه شده در این تحقیق، صحت بالایی دارند و از این مدل ها می توان برای پیش بینی  $pIC_{50}$  هر ترکیب جدید با خطای قابل قبول بهره برد.

جدول (۳-۱۷) - نتایج حاصل از ارزیابی مدل رگرسیون خطی چندگانه و شبکه عصبی با استفاده از روش رد مرحله ای تک تک

شماره ترکیب	$pIC_{50}$				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۱	۶/۱۹	۶/۴۰	۶/۴۲	۳/۳۹	۳/۷۲
۲	۶/۹۲	۶/۸۴	۶/۸۴	-۱/۱۶	-۱/۱۶
۳	۶/۶۰	۶/۸۵	۶/۷۹	۳/۷۷	۲/۸۸
۴	۶/۸۵	۶/۸۷	۶/۸۳	۰/۲۹	-۰/۲۹
۵	۷/۲۵	۷/۱۹	۷/۱۰	-۰/۸۳	-۲/۰۷
۶	۶/۲۵	۶/۵۷	۶/۱۹	۵/۱۲	-۰/۹۶
۷	۶/۶۱	۶/۵۹	۶/۴۷	-۰/۳۰	-۲/۱۲
۸	۷/۶۳	۷/۲۵	۷/۶۸	-۴/۹۸	۰/۶۶
۹	۷/۶۰	۷/۸۸	۷/۷۹	۳/۶۸	۲/۵۰
۱۰	۷/۱۱	۷/۱۵	۷/۰۷	۰/۵۶	-۰/۵۶
۱۱	۷/۱۴	۶/۹۵	۶/۸۳	-۲/۶۶	-۴/۳۴
۱۲	۷/۳۹	۷/۲۳	۷/۰۷	-۲/۱۷	-۴/۳۳
۱۳	۷/۳۱	۷/۵۷	۷/۴۳	۳/۵۶	۱/۶۴

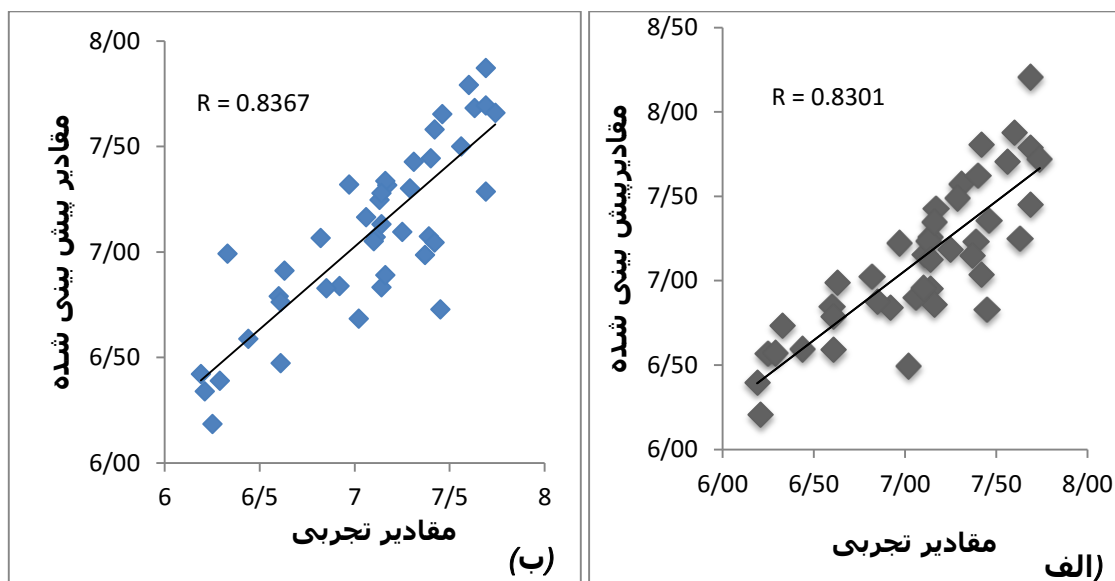
1-Enter

۱۴	۷/۱۴	۷/۲۶	۷/۱۳	۱/۶۸	-۰/۱۴
۱۵	۷/۱۷	۷/۴۳	۷/۳۲	۳/۶۳	۲/۰۹
۱۶	۶/۶۳	۶/۹۹	۶/۹۱	۵/۴۳	۴/۲۲
۱۷	۶/۴۴	۶/۶۰	۶/۵۹	۲/۴۸	۲/۳۳
۱۸	۶/۶۱	۶/۷۹	۶/۷۶	۲/۷۲	۲/۲۷
۱۹	۷/۴۰	۷/۶۳	۷/۴۴	۳/۱۱	۰/۵۴
۲۰	۷/۴۲	۷/۸۱	۷/۵۸	۵/۲۶	۲/۱۶
۲۱	۷/۲۹	۷/۴۹	۷/۳۰	۲/۷۴	۰/۱۴

ادامه جدول (17-3)

شماره ترکیب	pIC <sub>50</sub>				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۲۲	۶/۳۳	۶/۷۴	۶/۹۹	۶/۴۸	۱۰/۴۳
۲۳	۶/۸۲	۷/۰۲	۷/۰۷	۲/۹۳	۳/۶۷
۲۴	۷/۱۳	۷/۲۴	۷/۲۵	۱/۵۴	۱/۶۸
۲۵	۷/۵۶	۷/۷۱	۷/۵۰	۱/۹۸	-۰/۷۹
۲۶	۷/۶۹	۷/۴۵	۷/۸۷	-۳/۱۲	۲/۳۴
۲۷	۷/۴۵	۶/۸۳	۶/۷۳	-۸/۳۲	-۹/۶۶
۲۸	۷/۰۲	۶/۴۹	۶/۶۸	-۷/۵۵	-۴/۸۴
۲۹	۶/۳۱	۶/۲۱	۶/۳۴	۰/۰۰	۲/۰۹
۳۰	۶/۲۹	۶/۵۷	۶/۳۹	۴/۴۵	۱/۵۹
۳۱	۷/۱۶	۶/۸۶	۶/۸۹	-۴/۱۹	-۳/۷۷
۳۲	۷/۱۶	۷/۳۵	۷/۳۴	۲/۶۵	۲/۵۱
۳۳	۷/۰۶	۶/۹۰	۷/۱۶	-۲/۲۷	۱/۴۲
۳۴	۷/۱۰	۶/۹۶	۷/۰۵	-۱/۹۷	-۰/۷۰

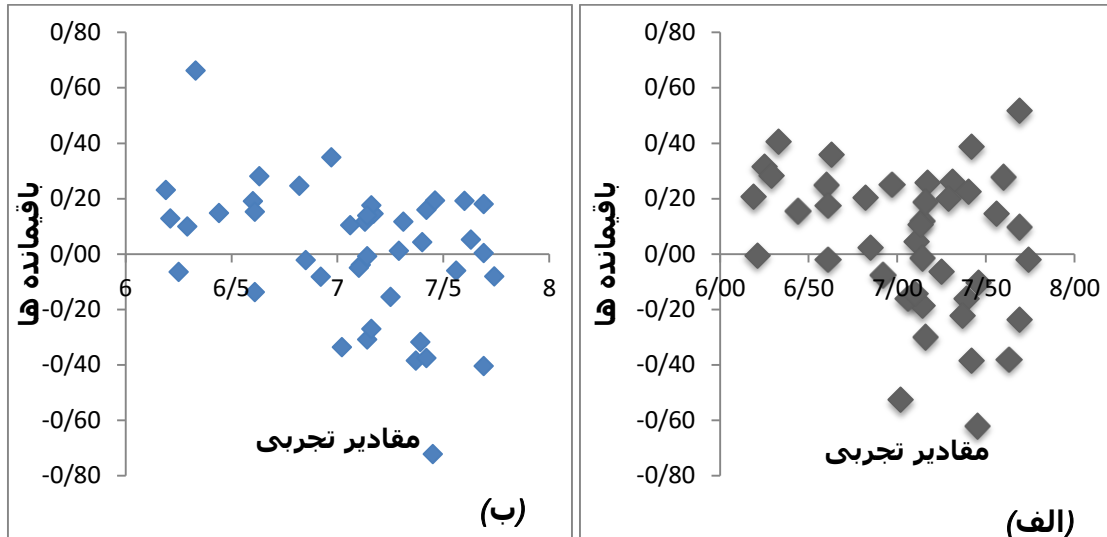
۳۵	۷/۳۷	۷/۱۵	۶/۹۹	-۲/۹۹	-۵/۱۶
۳۶	۷/۴۲	۷/۰۴	۷/۰۵	-۵/۱۲	-۴/۹۹
۳۷	۶/۹۷	۷/۲۲	۷/۳۲	۳/۵۹	۵/۰۲
۳۸	۷/۶۹	۷/۷۹	۷/۶۹	۱/۳۰	۰/۰۰
۳۹	۷/۷۴	۷/۷۲	۷/۶۶	-۰/۲۶	-۱/۰۳
۴۰	۷/۴۶	۷/۳۶	۷/۶۵	-۱/۳۴	۲/۵۵
۴۱	۷/۱۴	۷/۱۳	۷/۲۸	-۰/۱۴	۱/۹۶
۴۲	۷/۶۹	۸/۲۱	۷/۲۹	۶/۷۶	-۵/۲۰



شکل (۳-۱) - نمودار مقادیر پیش بینی شده  $pIC_{50}$  با استفاده از الف) روش MLR ب) شبکه عصبی مصنوعی بر حسب مقادیر واقعی به روش رد مرحله ای تک تک

جهت بررسی وجود یا عدم وجود خطای معین در این مدل، نمودار مقادیر باقیمانده‌ی محاسبه شده در ارزیابی مدل MLR و شبکه عصبی از مقادیر واقعی بر حسب مقادیر تجربی برای مولکول رسم

گردید که نتایج در شکل (۳-۱۲) نشان داده شده است. توزیع نسبتاً یکنواخت و اتفاقی داده‌ها حول محور صفر نشان می‌دهد که خطای سیستماتیکی در مدل وجود ندارد.



شکل (۳-۱۲) - نمودار باقیمانده‌ها برای کل داده‌ها (الف) با مدل MLR (ب) با مدل ANN

### ۳-۲-۴ - آزمون $Y$ - تصادفی<sup>۱</sup>

در این روش با استفاده از برنامه‌ای که در نرم‌افزار MATLAB نوشته و اجرا شد، مقادیر پاسخ به صورت تصادفی در محدوده‌ی پاسخ‌ها تغییر داده شد. سپس همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از شاخص  $R$  مورد بررسی قرار گرفت. این کار ده بار تکرار شد که مقادیر  $R$  برای سری تست و ارزیابی در جدول (۳-۱۸) ذکر شده است. مقادیر پایین  $R$  نشان دهنده‌ی عدم وجود همبستگی شانس‌ی در مدل بدست آمده از شبکه است. این تکنیک برای بررسی هرگونه ارتباط تصادفی بین داده‌ها انجام می‌شود.

جدول (۳-۱۸) - مقادیر ضریب همبستگی برای سری تست و سری ارزیابی بعد از ۱۰ بار تست  $Y$ -تصادفی

تکرار	سری ارزیابی	سری تست
۱	-۰/۳۲۶۴	-۰/۶۵۲۶
۲	-۰/۴۰۸۳	-۰/۲۹۹۹

<sup>1</sup>Y-randomization test

۳	-۰/۳۲۵۹	۰/۲۹۲۹
۴	۰/۱۴۴۶	۰/۱۶۲۱
۵	۰/۳۴۶۵	۰/۶۴۱۰
۶	-۰/۰۳۱۳	-۰/۱۳۲۴
۷	-۰/۱۸۳۸	۰/۲۷۶۵
۸	۰/۱۹۷۹	۰/۵۵۹۲
۹	۰/۴۱۷۰	-۰/۰۰۵۳
۱۰	۰/۲۸۸۴	۰/۳۳۷۰

### ۳-۲-۵- ارزیابی مدل‌های ارائه شده با استفاده از پارامترهای آماری

در این بخش جهت بررسی میزان کارایی مدل رگرسیون خطی چندگانه و شبکه عصبی مصنوعی ارائه شده چهار پارامتر آماری، MSE، REP، PRESS و میان پارامترهای آماری که در بخش (۳-۲-۸) ذکر شده اند انتخاب شده و مقادیر آنها محاسبه گردید. نتایج حاصل در جدول (۳-۱۹) نشان داده شده است.

جدول (۳-۱۹)- پارامترهای آماری محاسبه شده برای مدل رگرسیون خطی چندگانه و شبکه عصبی

پارامتر	سری ارزیابی (n=8)		سری تست (n=8)		کل داده ها (n=42)	
	ANN	MLR	ANN	MLR	ANN	MLR
MSE	۰/۰۱۶	۰/۰۳۸	۰/۰۳۹	۰/۰۴۷	۰/۰۶۰	۰/۰۶۷
PRESS	۰/۱۳۳	۰/۳۰۴۸	۰/۳۱۲	۰/۳۸۰	۲/۵۲۰	۲/۸۳۰
REP(%)	۱/۸۱۶	۲/۷۴۴	۲/۸۱۸	۳/۱۱۴	۳/۴۵۹	۳/۶۵۵
R	۰/۹۴۰	۰/۸۷۰	۰/۸۷۰	۰/۸۴۰	۰/۸۳۰	۰/۸۳۰

با توجه به نتایج رد مرحله ای تک تک ذکر شده برای مدل رگرسیون خطی چندگانه و شبکه عصبی مصنوعی ملاحظه می‌شود که هر دو مدل از قدرت پیش‌بینی نسبتاً بالایی برای مقادیر بازدارندگی مشتقات بنزیمیدازول برخوردارند، اما با توجه به نتایج حاصل از سری تست، شبکه عصبی نتایج بهتری نسبت به رگرسیون خطی چندگانه ارائه داده است.

### ۳-۲-۶- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی

مشارکت توصیفگرها، پارامتری است که سهم هر توصیفگر و در نتیجه، خواص ساختاری و ترکیبی مؤثر در میزان بازداری ترکیبات دارویی مورد بررسی، را نشان می‌دهد.

میزان مشارکت توصیفگرهای منتخب به روش زیر تعیین شد:

۱- توصیفگر مورد نظر به همراه اوزان مربوطه اش از شبکه‌ی بهینه شده حذف گردید.

۲- با استفاده از بقیه توصیفگرها مقدار متغیر وابسته ( $pIC_{50}$ ) برای هر ترکیب سری ارزیابی پیش بینی شد.

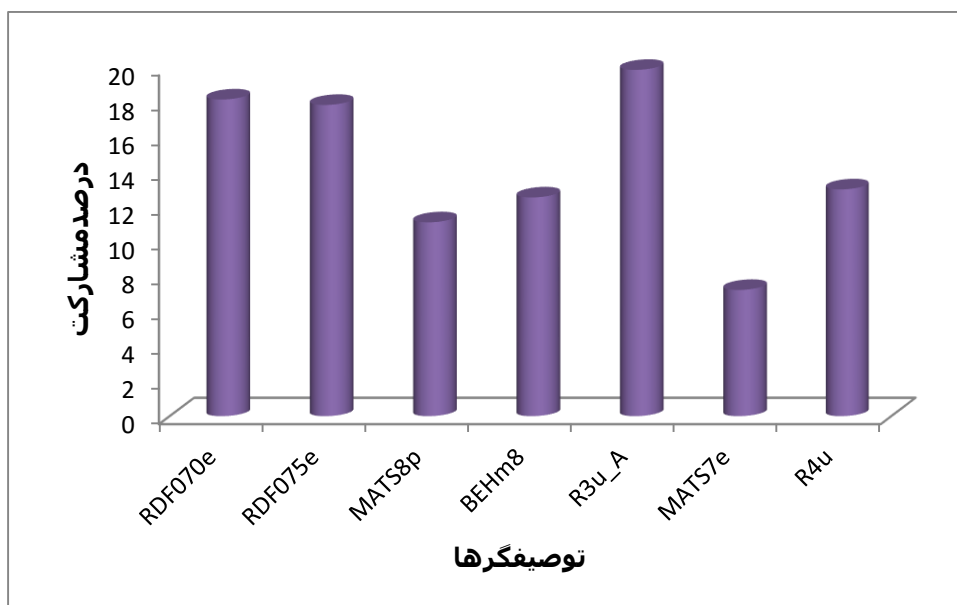
۳- میانگین خطای مطلق ترکیبات سری ارزیابی محاسبه گردید.

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

۵- سرانجام درصد مشارکت هر توصیفگر توسط رابطه (۳-۴) محاسبه شد [۵۶ و ۵۷].

$$c_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad \text{رابطه (۳-۴)}$$

در این رابطه  $c_i$  درصد مشارکت توصیفگر حذف شده  $i$ ام،  $N$  تعداد توصیفگرهای مدل و  $\Delta m_i$  میانگین خطای مطلق سری ارزیابی در غیاب توصیفگر  $i$ ام را نشان می‌دهد. درصد مشارکت توصیفگرهای منتخب در ترکیبات مورد بررسی در شکل (۳-۱۳) نشان داده شده است.



شکل (۳-۱۳) - درصد مشارکت توصیفگرهای حاضر در شبکه عصبی بهینه

همانطور که در نمودار مشاهده می شود، توصیفگرهای  $R3u\_A$ ،  $RDF070e$  و  $RDF075e$  بیشترین مشارکت را دارند.  $R3u\_A$  بیان کننده ی ویژگی های توپولوژیکی و هندسی مولکول است، بنابراین فعالیت بازداری این ترکیبات به هندسه ی مولکول وابستگی زیادی دارد.  $RDF070e$  و  $RDF075e$  نیز به نقش خاصیت اتمی (الکترونگاتیویته و قطبش پذیری و ...) ترکیب بازدارنده اشاره دارد.

### ۳-۲-۷- بررسی ارتباط توصیف کننده های ساختاری وارد شده در مدل و

#### فعالیت ضد انگلی

در این قسمت، یک بررسی اجمالی روی اثرات مختلف موجود در پارامتر بازدارندگی ترکیبات مورد مطالعه صورت می گیرد. بهترین مدل منتخب شامل هفت توصیفگر است که هر کدام بیانگر خصوصیات متفاوتی از مولکولهای مورد بررسی می باشد که در زیر در مورد ارتباط این توصیفگرها با بازدارندگی ترکیبات بحث شده است.

### ۳-۲-۷-۱- توصیف‌کننده‌های RDF<sup>۱</sup>

RDF یا تابع توزیع شعاعی مربوط به یک دسته اتمها (N)، معادل توزیع احتمال یافتن یک اتم در فضای کروی به شعاع r است و فرم کلی این تابع به صورت زیر است:

$$g(r) = f \sum_{i=1}^{N-1} \sum_{j>i}^N A_i A_j e^{B(r-r_{ij})^2} \quad \text{رابطه (۵-۳)}$$

در این رابطه، N تعداد اتمها،  $A_i$  و  $A_j$  خاصیت اتمی (الکترونگاتیویته، قطبش پذیری اتمی، جرم اتمی و...)،  $r_{ij}$  شعاع بین اتم  $i, j$ ،  $f$  فاکتور مقیاس،  $B$  فاکتور تسهیل‌کننده است که توزیع احتمال فاصله-ی اتمی را مشخص می‌کند و می‌توان آنرا به فاکتور دما برای تعریف انرژی جنبشی اتمی تعبیر کرد [۵۷].  $g(r)$  در نقاط گسسته‌ای با فواصل معین حساب می‌شود. با مجموعه‌ای از کدهای RDF با ویژگی‌های مختلف اتمی می‌توان ساختار سه بعدی یک مولکول را بطور واضح توصیف کرد. از میان این توصیفگرها، **RDF070e<sup>2</sup>** و **RDF075e** بعنوان اولین توصیفگرها در مدل برتر دیده می‌شوند. ضریب مثبت **RDF070e** و ضریب منفی **RDF075e** در مدل برتر نشان دهنده‌ی رابطه مولکولی مستقیم توصیفگر اولی و رابطه‌ی عکس توصیفگر دوم با مقادیر pIC<sub>50</sub> می‌باشد.

### ۳-۲-۷-۲- توصیفگرهای 2D-autocorrelation

توصیفگرهای دوبعدی خودارتباطی 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 ضریب گری<sup>۳</sup> نام دارد و بدین صورت محاسبه می‌شود:

$$C(d) = \frac{\frac{1}{2\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - w_j)^2}{\frac{1}{A-1} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad \text{رابطه (۶-۳)}$$

<sup>۱</sup>- Radial Distribution Function

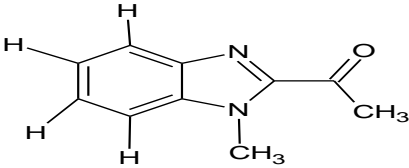
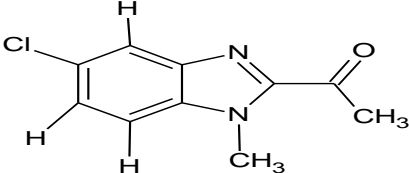
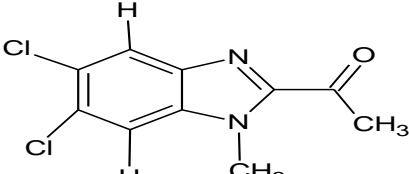
<sup>1</sup>-Radial Distribution Function weighted by atomic sanderson electronegativities

<sup>2</sup>-Geary Autocorrelation



که  $w$  یک ویژگی اتم،  $\bar{w}$  میانگین مقدار آن ویژگی روی مولکول،  $A$  تعداد اتمها و  $d$  فاصله‌ی توپولوژیکی است و  $\delta_{ij}$  نیز - که به تابع کرونگر<sup>۱</sup> معروف است، در حالتی که  $d=d_{ij}$  باشد برابر یک است، و در غیر اینصورت صفر است و  $\Delta$  هم مجموعه‌ی  $\delta$  هاست. مقدار این توصیف‌کننده که از جنس فاصله است، از صفر تا بی‌نهایت متغیر است [۲۶]. از میان این توصیفگرها **MATS7e**<sup>۲</sup> و **MATS8p**<sup>۲</sup> در مرحله‌ی انتخاب توصیفگرهای مؤثر در مدل ارائه شده در این فصل برگزیده شده‌اند. **MATS8p** چگونگی توزیع بار و قطبش پذیری در سطح اتمی را نشان می‌دهد و **MATS7e** که بر اساس الکترونگاتیویته‌ی ساندerson وزن دار شده است، نقش خاصیت فیزیکی الکترونگاتیویته‌ی ساندerson را در ترکیبات بیان می‌کند. با توجه به مثبت بودن ضریب رگرسیون برای این دو توصیفگر، افزایش مقدار آنها سبب افزایش مقدار  $pIC_{50}$  می‌شود. چند مثال از اثر توصیفگر **MATS7e** در جدول (۳-۲۰) ارائه شده است. همانطور که در جدول مشاهده می‌شود با افزایش تعداد گروه‌های الکترونگاتیو در مولکول، خاصیت دارویی ( $pIC_{50}$ ) ترکیب افزایش می‌یابد.

جدول (۳-۲۰) - مثال هایی از اثر توصیفگر **MATS7e** بر فعالیت دارویی

ترکیب	$pIC_{50}$
	۷/۰۶
	۷/۲۹
	۷/۴۵

3- Kronecker

4- Moran autocorrelation -lag8 weighted by polarizabilities

5- Moran autocorrelation -lag7 weighted by atomic sanderson electronegativities

### ۳-۲-۷-۳- توصیف‌کننده‌های BCUT

این توصیفگرها از دسته توصیفگرهای دوبعدی هستند و به عنوان مقدار ویژه یک ماتریس ارتباطی اصلاح شده<sup>۱</sup> (ماتریس بردن<sup>۲</sup>) می‌باشند. ماتریس بردن یک گراف تهی از هیدروژن را ارائه می‌دهد که در آن  $B_{ii}$ ، یعنی عناصر قطری در ارتباط با خاصیت اتمی مانند الکترونگاتیویته، حجم و اندروالس، جرم اتمی و عناصر غیر قطری در ارتباط با مرتبه پیوند دو اتم پیوند داده شده، هستند [۲۶ و ۵۶]. توصیفگر BEHm8<sup>۳</sup> که با توجه به وزن اتمی محاسبه شده است، از این گروه توصیف-کننده هاست و در مدل ارائه شده در این فصل ظاهر شده است. علامت مثبت این توصیفگر در معادله نشان می‌دهد که مقادیر بالاتر از این توصیفگر، باعث افزایش برهمکنش انگل-دارو و در نتیجه کاهش سرعت رشد انگل می‌شود.

### ۳-۲-۷-۴- توصیفگرهای GETAWAY<sup>۴</sup>

این توصیفگرها که بیان‌کننده ویژگی‌های توپولوژیکی و هندسی مولکول هستند، از ماتریس تأثیر مولکولی (MIM)<sup>۵</sup> محاسبه می‌شوند. این ماتریس به آسانی از مختصات فضایی اتم‌های مولکول در یک صورتبندی مشخص به دست می‌آید که با رابطه زیر بیان می‌گردد:

$$H = M (M^T M) \cdot M^T \quad \text{رابطه (۳-۷)}$$

که  $M$  ماتریس مختصات اتمی، بالانویس  $T$  به معنای ماتریس ترانپوز<sup>۶</sup> و  $H$  ماتریس تأثیر مولکولی است. عناصر قطری ( $h_{ii}$ ) ماتریس  $H$  لوریج<sup>۷</sup> نام دارند که هر یک بیانگر اثر یک اتم در ایجاد شکل

<sup>1</sup>- Modified connectivity matrix

<sup>2</sup>-Burden matrix

-highest eigenvalue n.8 of Burdan matrix weighthed by atomic masses

3

<sup>4</sup>- Geometrical, Topology and atom Weighted assembly

<sup>5</sup>-Molecular Influence Matrix

<sup>6</sup>- Transpose matrix

<sup>7</sup>-Leverage

کلی یک مولکول است. به عنوان مثال اتم‌های دورتر نسبت به اتم‌های نزدیکتر به مرکز اتم، دارای  $h_{ii}$  بزرگتر هستند.

اولین سری از این توصیفگرهای مولکولی، H-GETAWAY نامیده می‌شوند که تنها توسط اطلاعات به دست آمده توسط ماتریس نفوذ مولکولی محاسبه می‌شوند در حالی که سری R-GETAWAY، ترکیبی از اطلاعات ماتریس تأثیر – فاصله است که به صورت زیر تعریف می‌شود:

$$[R]_{ij} = \left[ \frac{\sqrt{h_{ii} h_{jj}}}{r_{ij}} \right]_{ij} \quad i \neq j \quad \text{رابطه (۳-۸)}$$

که  $h_{ii}$  و  $h_{jj}$  لوریج (قدرت نفوذ) دو اتم  $i$  و  $j$  و  $r_{ij}$  فاصله آن دو اتم است [۵۸].  $R3u^+$  (از دسته توصیفگرهای R-GETAWAY) از لوریج‌های دو اتم که در فاصله توپولوژیکی ۳ قرار دارند و با ویژگی خاصی از مولکول وزن‌دار نشده، محاسبه می‌شود.  $R4u^+$  نیز که از توصیفگرهای منتخب مدل است توصیفگری است که با ویژگی خاصی از اتمها محاسبه نگردیده است. این توصیفگرها از آن جهت مهم هستند که به ساختار سه بعدی مولکول حساس هستند و معمولاً برای مقایسه‌ی مولکولها و حتی کنفورمرها از نظر شکل، تقارن و اندازه اتم بکار می‌روند. علامت مثبت این دو توصیفگر در مدل برتر، نشان می‌دهد که با افزایش مقدار آنها، مقدار  $pIC_{50}$  افزایش می‌یابد.

### ۳-۳- نتیجه‌گیری نهایی

روابط بین مولکول‌های آلی و فعالیت دارویی آنها روابطی پیچیده و غیرخطی است، از این رو بهترین مدل برای بیان ارتباط ساختار- فعالیت مدلهایی هستند که ارتباط بین توصیفگرها و فعالیت دارویی را به صورت غیرخطی بیان کنند. به همین جهت در این تحقیق از شبکه عصبی مصنوعی به عنوان یک روش غیرخطی برای مطالعه ارتباط کمی ساختار – فعالیت مشتقات بنزیمیدازول استفاده شده

1- R maximal autocorrelation of lag3 unweighted  
2- R autocorrelation of lag4 unweighted

است و این اولین گزارش در مورد بکارگیری شبکه عصبی با الگوریتم آموزشی تنظیم بایزین در مورد ترکیبات مورد مطالعه می‌باشد. نتایج بیانگر آن است که شبکه می‌تواند یک رابطه رضایت‌بخش بین ساختار مولکولی و قدرت دارویی برقرار کند. بنابراین پژوهشگران می‌توانند با الگوبرداری از این روند، فعالیت دارویی ترکیبات جدیدی از این سری مشتقات با استخلاف‌های متفاوت را مورد بررسی قرار دهند و بدین ترتیب با صرف هزینه و وقت کمتر در مسیر تهیه داروهایی با اثر بخشی بهتر گام بردارند.

## فصل چهارم

مطالعه ارتباط کمی ساختار- فعالیت مشتقات  
بنزیمیدازول با استفاده از رگرسیون خطی چندگانه  
و شبکه عصبی مصنوعی

## ۴-۱- محاسبات و مدل سازی فعالیت دارویی مشتقات بنزیمیدازول به

### عنوان داروهای ضدانگل

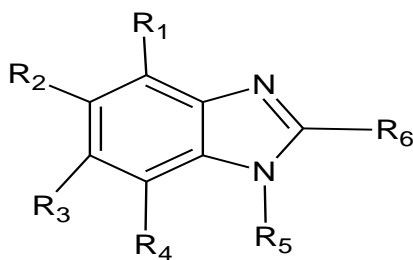
در راستای مطالعه‌ی ارتباط کمی ساختار- فعالیت ضدانگل، فعالیت دسته‌ی جدیدی از مشتقات بنزیمیدازول روی انگل تریکوموناس واژینالیس (T.V) توسط روش رگرسیون خطی چندگانه و شبکه عصبی مصنوعی پیش‌بینی شد و مدل‌های بدست آمده از هر دو روش توسط پارامترهای آماری مختلف مورد سنجش قرار گرفت.

### ۴-۱-۱- نرم افزارهای مورد استفاده

همانطور که در فصل قبل ذکر شد در این پروژه از نرم افزار Hyperchem 7.0 جهت رسم و بهینه سازی ساختار ترکیبات به روش نیمه تجربی AM1 استفاده گردید. جهت محاسبه‌ی توصیف-کننده‌های مولکولی نرم افزار Dragon 2.1 به کار گرفته شد. بدین منظور هر یک از ساختارها بصورت تک تک به این نرم افزار داده شد و برای هر یک از آنها تمامی توصیف کننده‌ها محاسبه گردید. به کمک نرم افزار SPSS13.0 و به روش رگرسیون مرحله‌ای، توصیف کننده‌های مولکولی مناسب انتخاب و مدل‌های آماری ایجاد گردیدند. برنامه‌ها و الگوریتم شبکه عصبی مصنوعی که در این پروژه بکار گرفته شدند نیز در محیط نرم افزار MATLAB، نسخه‌ی 2009Rb نوشته و اجرا شدند.

### ۴-۱-۲- سری داده‌ها

در این قسمت از تحقیق، غلظت بازدارنده (pIC<sub>50</sub>) ۴۹ ترکیب از مشتقات بنزیمیدازول با گروه‌های عاملی متفاوت به عنوان سری داده‌ها مورد استفاده قرار گرفت [۶، ۵۰، ۵۳ و ۵۹]. ساختار این ترکیبات به همراه مقادیر pIC<sub>50</sub> آنها در جدول (۴-۱) ذکر شده است. اسکلت مولکولی ترکیبات مورد بررسی در شکل (۴-۱) آورده شده است.



شکل (۴-۱)- اسکلت مولکولی ترکیبات مورد بررسی

جدول (۴-۱)- سری داده های بکار رفته در مدل سازی

NO	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	pIC <sub>50</sub>
۱ <sup>v</sup>	H	Br	H	H	H	-CF <sub>3</sub>	۵/۸۰
۲ <sup>tr</sup>	H	Br	Br	H	H	-CF <sub>3</sub>	۶/۶۵
۳ <sup>tr</sup>	Br	H	Br	H	H	-CF <sub>3</sub>	۶/۷۲
۴ <sup>t</sup>	Br	Br	Br	H	H	-CF <sub>3</sub>	۶/۵۶
۵ <sup>v</sup>	H	Cl	Cl	H	H	-C <sub>2</sub> F <sub>5</sub>	۶/۵۲
۶ <sup>tr</sup>	H	NO <sub>2</sub>	NO <sub>2</sub>	H	H	-CF <sub>3</sub>	۶/۲۴
۷ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	۶/۴۳
۸ <sup>t</sup>	H	H	SCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	H	H	-CF <sub>3</sub>	۶/۴۶
۹ <sup>tr</sup>	H	H	SOCH <sub>2</sub> CH <sub>2</sub> CH	H	H	-CF <sub>3</sub>	۵/۵۵
۱۰ <sup>tr</sup>	H	SCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	H	H	Me	-CF <sub>3</sub>	۶/۶۹
۱۱ <sup>v</sup>	H	H	Cl	H	Me	-CONH <sub>2</sub>	۶/۹۶
۱۲ <sup>tr</sup>	H	H	Cl	H	Me	-CONHCH <sub>3</sub>	۶/۹۷
۱۳ <sup>t</sup>	H	H	Cl	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۶/۶۲
۱۴ <sup>tr</sup>	H	H	Cl	H	Me	-COOCH <sub>2</sub> CH <sub>3</sub>	۷/۷۲

NO	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	pIC <sub>50</sub>
۱۵ <sup>tr</sup>	H	H	Cl	H	Me	-CH(OH)CH <sub>3</sub>	۶/۹۶
۱۶ <sup>v</sup>	H	Cl	Cl	H	Me	-COCH <sub>3</sub>	۶/۶۴
۱۷ <sup>tr</sup>	H	Cl	H	H	Me	-CONH <sub>2</sub>	۶/۷۳
۱۸ <sup>t</sup>	H	Cl	H	H	Me	-CONHCH <sub>3</sub>	۶/۴۴
۱۹ <sup>tr</sup>	H	Cl	H	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۶/۶۷
۲۰ <sup>tr</sup>	H	Cl	H	H	Me	-COOCH <sub>2</sub> CH <sub>3</sub>	۷/۵۶
۲۱ <sup>v</sup>	H	Cl	H	H	Me	-CH(OH)CH <sub>3</sub>	۷/۰۲
۲۲ <sup>tr</sup>	H	Cl	H	H	Me	-COCH <sub>3</sub>	۶/۸۸
۲۳ <sup>tr</sup>	H	Cl	Cl	H	Me	-CONH <sub>2</sub>	۶/۸۷
۲۴ <sup>t</sup>	H	Cl	Cl	H	Me	-CONHCH <sub>3</sub>	۶/۶۴
۲۵ <sup>tr</sup>	H	Cl	Cl	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۷/۱۲
۲۶ <sup>v</sup>	H	Cl	Cl	H	Me	-COOCH <sub>2</sub> CH <sub>3</sub>	۷/۵۳
۲۷ <sup>tr</sup>	H	Cl	Cl	H	Me	-CH(OH)CH <sub>3</sub>	۷/۳۹
۲۸ <sup>tr</sup>	H	Cl	Cl	H	Me	-COCH <sub>3</sub>	۷/۱۹
۲۹ <sup>tr</sup>	H	H	H	H	Me	-CONH <sub>2</sub>	۶/۷۸
۳۰ <sup>t</sup>	H	H	H	H	Me	-CONHCH <sub>3</sub>	۶/۹۶
۳۱ <sup>v</sup>	H	H	H	H	Me	-CON(CH <sub>3</sub> ) <sub>2</sub>	۶/۳۶
۳۲ <sup>tr</sup>	H	H	H	H	Me	-COOCH <sub>2</sub> CH <sub>3</sub>	۷/۰۶



NO	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	pIC <sub>50</sub>
۳۳ <sup>tr</sup>	H	H	H	H	Me	-CH(OH)CH <sub>3</sub>	۶/۷۷
۳۴ <sup>tr</sup>	H	H	H	H	Me	-COCH <sub>3</sub>	۶/۶۷
۳۵ <sup>t</sup>	H	Cl	Cl	H	H	-CF <sub>3</sub>	۶/۶۲
۳۶ <sup>tr</sup>	H	Cl	H	H	Me	-CF <sub>3</sub>	۶/۶۲
۳۷ <sup>v</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> OH	۶/۱۶
۳۸ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> OH	۶/۸۵
۳۹ <sup>tr</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	۷/۱۴
۴۰ <sup>t</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	۷/۳۹
۴۱ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> N(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	۷/۵۳
۴۲ <sup>v</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	۷/۰۷
۴۳ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	۸/۰۰
۴۴ <sup>t</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> (N-piperidyl)	۷/۱۴
۴۵ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> (N-piperidyl)	۷/۶۹
۴۶ <sup>v</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> (N-morpholinyl)	۷/۲۹
۴۷ <sup>tr</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> (N-morpholinyl)	۷/۲۲
۴۸ <sup>tr</sup>	Cl	H	Cl	H	H	-SCH <sub>2</sub> CH <sub>2</sub> (p-nitrophenyl)	۸/۰۰
۴۹ <sup>t</sup>	Br	H	Br	H	H	-SCH <sub>2</sub> CH <sub>2</sub> (p-nitrophenyl)	۸/۴۹

t:test set , tr:train set, v:validation set

## ۴-۱-۳- رسم و بهینه سازی ساختار مولکولها

با توجه به بررسی ارتباط کمی ساختار - فعالیت، ابتدا ساختار مولکولها در نرم افزار Hyperchem7 رسم و پس از اتصال اتم‌های هیدروژن، این ساختارها با روش کوانتومی AM1 از دسته روشهای نیمه تجربی بهینه شدند و بهینه سازی تا زمانیکه گرادیان انرژی به  $0/001$  کیلوکالری بر مول رسید ادامه یافت.

## ۴-۱-۴- محاسبه و انتخاب توصیفگرهای مهم با روش رگرسیون خطی

### چندگانه

پس از رسم و بهینه سازی ساختارها، ساختار بهینه شدهی مولکولها به نرم افزار Dragon 2.1 منتقل شدند و برای هر ترکیب ۱۴۸۱ توصیفگر محاسبه شد، اما با توجه به این که توصیفگرهای زیاد فرایند مدل سازی را پیچیده و طولانی می کنند و تأثیری در بهبود نتایج ندارند، مراحل زیر برای انتخاب توصیفگرهایی که بطور معناداری فعالیت دارویی را به ساختار مولکول ارتباط دهند، انجام گرفت:

الف- توصیفگرهایی که دارای مقادیر یکسان برای همه ترکیبات بودند، از فهرست توصیفگرها حذف شدند.

ب- با توجه به اینکه نبایستی متغیرهای ورودی در روش رگرسیون خطی چندگانه، با یکدیگر همبستگی داشته باشند، لذا از میان جفت توصیفگری که ضریب همبستگی بیشتر از  $0/90$  داشتند، یکی حذف گردید. این کار با استفاده از نرم افزار SPSS برای هر گروه از توصیفگرها انجام شد.

ج- در مرحله بعد، روش رگرسیون خطی چندگانه مرحله‌ای بر اساس انتخاب رو به جلو حذف رو به عقب جهت غربال کردن بهترین توصیفگرهای مورد نیاز با استفاده از نرم افزار SPSS 13.0 انجام شد که در این مرحله ۷ مدل با ۷ توصیفگر انتخاب شدند که بیشترین همبستگی را با فعالیت ضدانگلی

ترکیبات مورد مطالعه دارند. با توجه به پارامترهای آماری، معمولاً مدلی که تعداد توصیفگر کمتر و ضریب همبستگی (R) بیشتر و خطای کمتر دارد، به عنوان مدل برتر انتخاب می‌شود. برای این کار داده‌ها به طور تصادفی به سه سری آموزش، ارزیابی و تست تقسیم شدند. سری تست و ارزیابی هر کدام ۱۰ ترکیب و سری آموزش نیز ۲۹ ترکیب را شامل می‌شوند. سپس، توسط سری آموزش، ابتدا با ۲ توصیفگر اول انتخاب شده، به روش اجباری<sup>۱</sup> 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، ۵، ۶ و ۷ توصیفگر به طور مجزا سری ارزیابی مورد بررسی قرار گرفت. میانگین مربعات خطا (mse) و ضریب همبستگی (R) محاسبه شده برای سری ارزیابی در جدول (۲-۴) آمده است

جدول (۲-۴) - مقایسه آماره‌های چند مدل نهایی به دست آمده از روش MLR

مدل	۲	۳	۴	۵	۶	۷
R	۰/۸۸	۰/۸۵	۰/۹۲	۰/۹۰	۰/۸۹	۰/۸۷
MSE	۰/۱۰۹	۰/۱۲۰	۰/۰۹۰	۰/۰۸۴	۰/۰۸۵	۰/۱۰۳

همانطور که در جدول (۲-۴) مشاهده می‌شود مدل پنجم 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 بصورت زیر به دست آمد:

$$Y = 5.54 + 0.135X_{0sol} - 2.21M_{0r23P} - 4.17LDip - 0.383nCONR_{2ph} + 0.345nCOOR_{ph}$$

توصیف‌کننده‌های وارد شده در این مدل به همراه کلاس، معنی، ضریب رگرسیون و اثرمتوسط آنها، در جدول (۳-۴)، ماتریس همبستگی این توصیف‌کننده‌ها در جدول (۴-۴) و مقادیر عددی توصیفگرهای منتخب در جدول (۵-۴) آورده شده است.

<sup>1</sup>- Enter

جدول (۳-۴) - توصیف کننده های وارد شده در مدل MLR به همراه کلاس مربوطه

تعداد	نماد	کلاس	ضریب رگرسیون	اثرمتوسط	معانی
۱	X0sol	topological	۰/۱۳۵	۰/۲۳۸	Solvation connectivity index chi-0
۲	Mor23p	3D -MORSE	-۲/۲۱۰	۰/۱۰۱	3D -MORSE-signal 23
۳	LDip	charge	-۴/۱۷۰	-۰/۱۳۱	Local dipole index
۴	nCONR <sub>2</sub> ph	Functional groups	-۰/۳۸۳	-۰/۰۰۴	Number of tertiary amids (aromatic)
۵	nCOORph	Functional groups	۰/۳۴۵	۰/۰۱۱	Number of esters (aromatic)

جدول (۴-۴) - ماتریس همبستگی توصیف کننده های به کار رفته در مدل نهایی

	X0sol	Mor23p	LDip	nCONR <sub>2</sub> ph	NCOORph
X0sol	۱				
Mor23p	-۰/۰۱۴	۱			
LDip	-۰/۲۲۵	۰/۰۳۵	۱		
nCONR <sub>2</sub> ph	۰/۰۱۶	-۰/۰۵۰	۰/۱۹۳	۱	
nCOORph	-۰/۰۰۴	-۰/۲۱۱	-۰/۱۲۳	-۰/۰۸۹	۱

این ماتریس عدم وجود همبستگی بین توصیفگرهای مدل را نشان می دهد.

جدول (۴-۵) -مقادیر عددی توصیفگرهای منتخب در مدل MLR نهایی

شماره ترکیب	X0sol	Mor23P	LDIp	nCONR2ph	nCOONRph
۱	۸/۳۵	-۰/۳۱	۰/۲۶	۰/۰۰	۰/۰۰
۲	۱۰/۲۲	-۰/۲۹	۰/۲۵	۰/۰۰	۰/۰۰
۳	۱۰/۲۲	-۰/۳۳	۰/۲۴	۰/۰۰	۰/۰۰
۴	۱۲/۰۹	-۰/۳۱	۰/۲۵	۰/۰۰	۰/۰۰
۵	۹/۷۲	-۰/۳۴	۰/۲۲	۰/۰۰	۰/۰۰
۶	۱۱/۳۷	-۰/۲۷	۰/۲۸	۰/۰۰	۰/۰۰
۷	۱۴/۷۷	-۰/۱۶	۰/۲۰	۰/۰۰	۰/۰۰
۸	۹/۸۲	-۰/۳۱	۰/۲۸	۰/۰۰	۰/۰۰
۹	۱۰/۶۳	-۰/۲۶	۰/۴۹	۰/۰۰	۰/۰۰
۱۰	۱۰/۶۹	-۰/۳۶	۰/۲۵	۰/۰۰	۰/۰۰
۱۱	۱۰/۷۹	-۰/۳۱	۰/۲۴	۰/۰۰	۰/۰۰
۱۲	۱۱/۵	-۰/۳۱	۰/۲۱	۰/۰۰	۰/۰۰
۱۳	۱۲/۳۷	-۰/۲۸	۰/۱۹	۱/۰۰	۰/۰۰
۱۴	۱۲/۲۱	-۰/۳۴	۰/۲۰	۰/۰۰	۱/۰۰
۱۵	۱۰/۷۹	-۰/۳۶	۰/۱۹	۰/۰۰	۰/۰۰
۱۶	۱۰/۷۹	-۰/۳۲	۰/۲۰	۰/۰۰	۰/۰۰
۱۷	۱۰/۷۹	-۰/۲۹	۰/۲۳	۰/۰۰	۰/۰۰
۱۸	۱۱/۵	-۰/۳۰	۰/۲۰	۰/۰۰	۰/۰۰
۱۹	۱۲/۳۷	-۰/۳۰	۰/۱۸	۱/۰۰	۰/۰۰
۲۰	۱۲/۲۱	-۰/۳۷	۰/۱۹	۰/۰۰	۱/۰۰
۲۱	۱۰/۷۹	-۰/۳۸	۰/۱۸	۰/۰۰	۰/۰۰
۲۲	۱۰/۷۹	-۰/۳۴	۰/۱۹	۰/۰۰	۰/۰۰
۲۳	۱۲/۱۶	-۰/۳۱	۰/۲۳	۰/۰۰	۰/۰۰
۲۴	۱۲/۸۷	-۰/۳۲	۰/۲۰	۰/۰۰	۰/۰۰
۲۵	۱۳/۷۴	-۰/۳۶	۰/۱۸	۱/۰۰	۰/۰۰

شماره ترکیب	X0sol	Mor23P	LDIp	nCONR2ph	nCOONRph
۲۶	۱۳/۵۸	-۰/۳۷	۰/۱۹	۰/۰۰	۱/۰۰
۲۷	۱۲/۱۶	-۰/۳۷	۰/۱۷	۰/۰۰	۰/۰۰
۲۸	۱۲/۱۶	-۰/۳۳	۰/۱۹	۰/۰۰	۰/۰۰
۲۹	۹/۴۲	-۰/۳۳	۰/۲۴	۰/۰۰	۰/۰۰
۳۰	۱۰/۱۳	-۰/۳۴	۰/۲۲	۰/۰۰	۰/۰۰
۳۱	۱۱/۰۰	-۰/۳۷	۰/۲۰	۱/۰۰	۰/۰۰
۳۲	۱۰/۸۴	-۰/۳۹	۰/۲۱	۰/۰۰	۱/۰۰
۳۳	۹/۴۲	-۰/۳۹	۰/۱۹	۰/۰۰	۰/۰۰
۳۴	۹/۴۲	-۰/۳۸	۰/۲۱	۰/۰۰	۰/۰۰
۳۵	۹/۲۲	-۰/۲۹	۰/۲۴	۰/۰۰	۰/۰۰
۳۶	۸/۷۲	-۰/۳۳	۰/۲۱	۰/۰۰	۰/۰۰
۳۷	۱۲/۱۹	-۰/۲۱	۰/۲۲	۰/۰۰	۰/۰۰
۳۸	۱۳/۱۹	-۰/۱۷	۰/۲۲	۰/۰۰	۰/۰۰
۳۹	۱۳/۷۷	-۰/۱۷	۰/۱۹	۰/۰۰	۰/۰۰
۴۰	۱۵/۱۸	-۰/۳۵	۰/۲۰	۰/۰۰	۰/۰۰
۴۱	۱۶/۱۸	-۰/۳۳	۰/۲۱	۰/۰۰	۰/۰۰
۴۲	۱۴/۴۸	-۰/۲۳	۰/۱۹	۰/۰۰	۰/۰۰
۴۳	۱۵/۴۸	-۰/۲۱	۰/۱۹	۰/۰۰	۰/۰۰
۴۴	۱۵/۳۰	-۰/۳۴	۰/۲۰	۰/۰۰	۰/۰۰
۴۵	۱۶/۳۰	-۰/۳۳	۰/۲۰	۰/۰۰	۰/۰۰
۴۶	۱۵/۳۰	-۰/۲۰	۰/۱۸	۰/۰۰	۰/۰۰
۴۷	۱۶/۳۰	-۰/۲۵	۰/۱۹	۰/۰۰	۰/۰۰
۴۸	۱۷/۷۵	-۰/۴۹	۰/۲۴	۰/۰۰	۰/۰۰
۴۹	۱۸/۷۵	-۰/۵۳	۰/۲۴	۰/۰۰	۰/۰۰

#### ۴-۱-۵- مدل سازی با استفاده از شبکه عصبی مصنوعی (ANN)

بخاطر اینکه ممکن است مدل خطی بهترین مدل برای توصیف رفتار سری داده‌ها نباشد، شبکه عصبی مصنوعی برای مدل سازی غیرخطی داده‌ها مورد استفاده قرار گرفت. در این تحقیق برنامه-نویسی و اجرای یک شبکه سه لایه در محیط نرم افزار MATLAB نسخه 2009Rb صورت گرفت. تابع کارایی شبکه در اینجا میانگین مجذور خطای استاندارد (mse) می باشد. همانند روش MLR داده ها بصورت تصادفی و با نسبت ۶۰:۲۰:۲۰ به سه سری آموزش، ارزیابی و تست تقسیم شدند. سری آموزش برای آموزش شبکه و سری ارزیابی برای بررسی قدرت تعمیم پذیری مدل بدست آمده از سری آموزش بکار گرفته شد.

#### ۴-۱-۶- بهینه سازی پارامترهای شبکه عصبی

برای بدست آوردن معادله و کمترین خطا، پارامترهای شبکه (تعداد متغیرهای ورودی شبکه، تابع آموزش، تابع انتقال، تعداد گره‌ها در لایه پنهان،  $\mu$  و تعداد دورهای آموزش) همانند فصل قبل بهینه شدند.

#### ۴-۱-۶-۱- انتخاب تعداد ورودی‌های شبکه، نوع تابع انتقال، تابع آموزش و

#### تعداد نرون های لایه ی پنهان

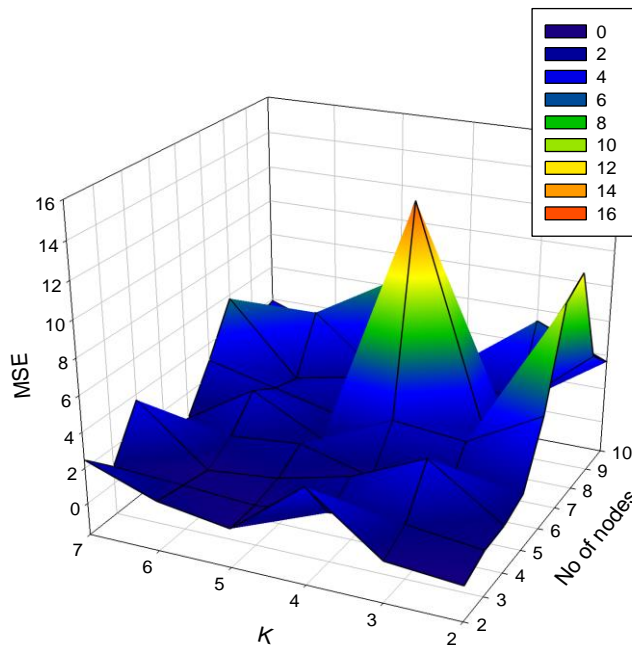
برای شناسایی و بهینه سازی بهترین پارامترهای شبکه، شبکه هایی با ۲ تا ۷ توصیف کننده ایجاد شدند. سپس هر شبکه با الگوریتم آموزشی لونیبرگ- مارکوارت و تنظیم بایزین با تعداد گره ۲ تا ۱۰ در لایه ی پنهان آموزش داده شد. در حالیکه برای بدست آوردن بهترین تابع انتقال در لایه ی پنهان مدل های شبکه عصبی نیز از توابع لگاریتم زیگموئید و تانژانت هایپربولیک به عنوان توابع انتقال استفاده شد. در روند بهینه سازی شبکه، به حداقل رساندن میانگین مربعات خطای شبکه به عنوان

معیار انتخاب گردید. در تمامی این شبکه ها تعداد دوره‌های آموزش و پارامتر  $\mu$  ثابت و به ترتیب ۲۰ و ۰/۰۰۵ در نظر گرفته شدند. همچنین از تابع انتقال خطی (purelin) در لایه‌ی خروجی استفاده گردید. نتایج بهینه سازی این پارامترها در جداول (۴-۶)، (۴-۷)، (۴-۸) و (۴-۹) گردآوری شده است. نمودار سه بعدی این جداول در شکل‌های (۴-۲)، (۴-۳)، (۴-۴) و (۴-۵) ارائه شده است. بر اساس این نتایج تابع آموزش تنظیم بایزین و تعداد متغیر ورودی ۵، تعداد ۶ گره در لایه‌ی مخفی و تابع لگاریتم زیگموئید به عنوان تابع انتقال لایه‌ی پنهان انتخاب شد.



جدول (۴-۶) - مقادیر میانگین مربعات خطاهای سری ارزیابی برای شبکه با الگوریتم آموزشی لونیبرگ-مارکوارت و تابع انتقال تانزانانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

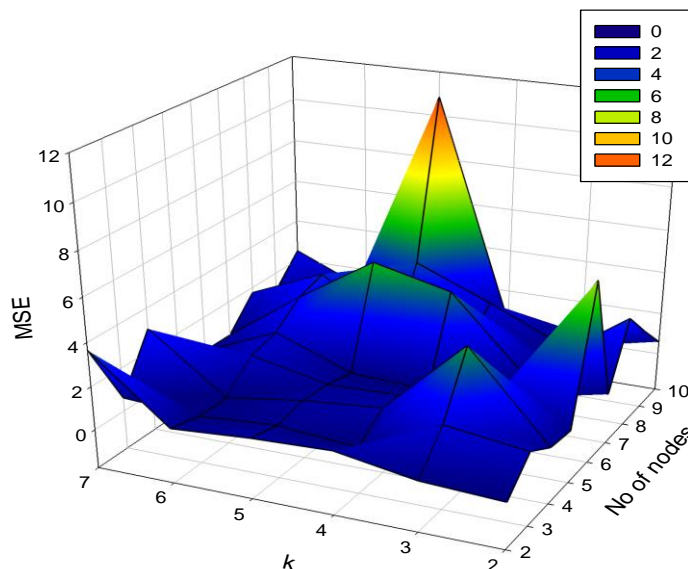
		تعداد توصیفگرها					
		۲	۳	۴	۵	۶	۷
تعداد گره ها در لایه پنهان	۲	۰/۱۶۰۷	۰/۵۰۸۷	۳/۵۸۹۳	۰/۳۹۰۵	۱/۰۱۳۴	۲/۴۶۳۶
	۳	۰/۷۶۴۱	۱/۲۳۶۸	۰/۹۷۶۱	۰/۴۷۶۵	۰/۷۲۹۶	۰/۹۴۶۴
	۴	۰/۸۴۸۷	۳/۶۴۳۹	۱/۸۰۹۸	۰/۸۳۳۴	۰/۴۴۳۳	۳/۶۹۳۰
	۵	۱/۳۹۴۶	۰/۴۱۷۱	۰/۵۹۶۴	۱/۶۸۸	۱/۳۲۷۷	۰/۵۵۳۳
	۶	۴/۵۸۱۶	۲/۳۵۰۸	۲/۷۶۹۲	۰/۸۵۲۸	۲/۹۳۲۹	۰/۵۱۶۷
	۷	۹/۷۳۷۵	۱/۸۸۰۱	۱۳/۹۰۲۵	۱/۱۰۷۸	۲/۰۶۷۹	۲/۹۸۷۳
	۸	۱۰/۴۷۶۹	۲/۰۸۰۸	۳/۳۵۱۰	۲/۴۸۴۱	۱/۶۵۳۱	۵/۷۸۸۵
	۹	۴/۹۳۷۰	۱/۴۰۰۰	۲/۷۳۲۳	۱/۵۴۱۸	۴/۷۱۷۳	۳/۴۱۳۲
	۱۰	۳/۶۲۶۹	۵/۳۵۳۶	۰/۷۶۳۹	۶/۶۲۴۸	۲/۴۹۵۴	۳/۸۸۷۲



شکل (۴-۲) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش لونیبرگ-مارکوارت و تابع انتقال تانزانانت سیگموئیدی

جدول (۴-۷) - مقادیر میانگین مربعات خطاهای سری ارزیابی برای شبکه با الگوریتم آموزشی لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

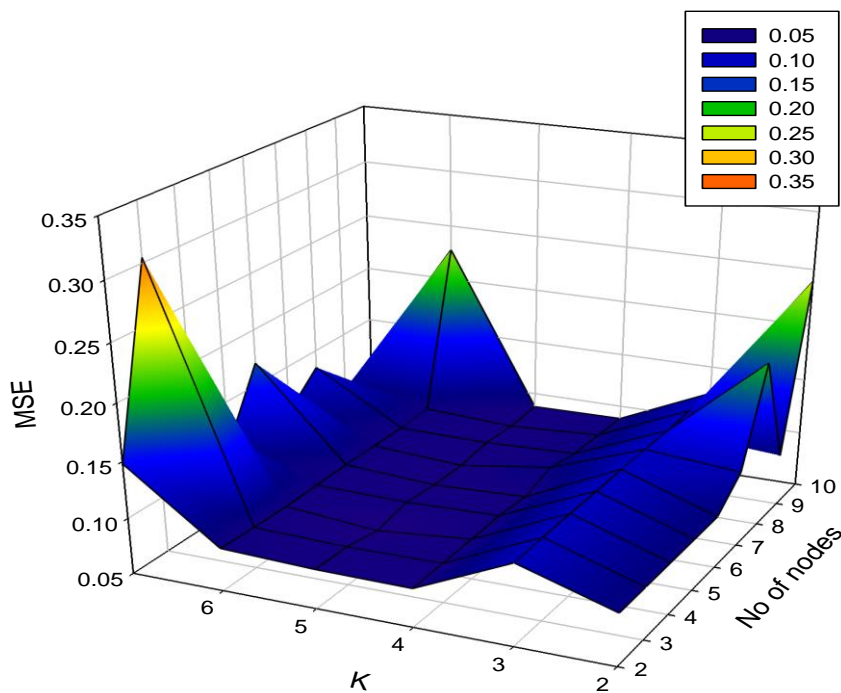
		تعداد توصیفگرها					
		۲	۳	۴	۵	۶	۷
تعداد گره ها در لایه پنهان	۲	۰/۳۲۴۷	۰/۵۱۵۳	۱/۱۳۳۲	۰/۹۹۶۱	۰/۷۴۱۳	۳/۶۵۰۲
	۳	۱/۵۶۸۶	۲/۰۵۳۲	۰/۳۳۵۵	۰/۶۵۳۰	۰/۴۱۰۳	۰/۶۱۱۳
	۴	۰/۷۹۵۳	۴/۶۵۵۰	۱/۲۴۱۱	۰/۹۱۸۵	۰/۳۹۴۳	۳/۰۳۳۷
	۵	۰/۶۳۶۸	۱/۱۲۱۱	۰/۸۹۰۰	۰/۸۱۲۸	۱/۵۷۲۷	۰/۳۴۹۴
	۶	۶/۵۳۵۲	۰/۵۹۵۲	۰/۳۱۷۳	۰/۵۰۹۹	۱/۸۶۳۳	۰/۳۷۵۸
	۷	۰/۵۴۷۹	۰/۶۴۹۱	۴/۱۷۹۶	۵/۰۲۱۴	۲/۸۱۲۷	۰/۳۵۵۲
	۸	۳/۳۱۲۷	۰/۷۱۳۵	۱/۲۳۷۱	۰/۵۶۸۱	۳/۲۰۷۳	۱/۷۴۵۲
	۹	۱/۹۱۷۷	۰/۵۵۸۳	۲/۱۲۹۱	۳/۵۹۰۶	۰/۶۵۰۹	۱/۱۲۰۹
	۱۰	۰/۶۰۶۵	۰/۴۰۸۱	۰/۴۳۳۳	۱۰/۸۶۶۱	۰/۹۵۲۱	۲/۴۶۲۱



شکل (۴-۳) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی

جدول (۴-۸) - مقادیر میانگین مربعات خطاهای سری ارزیابی برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

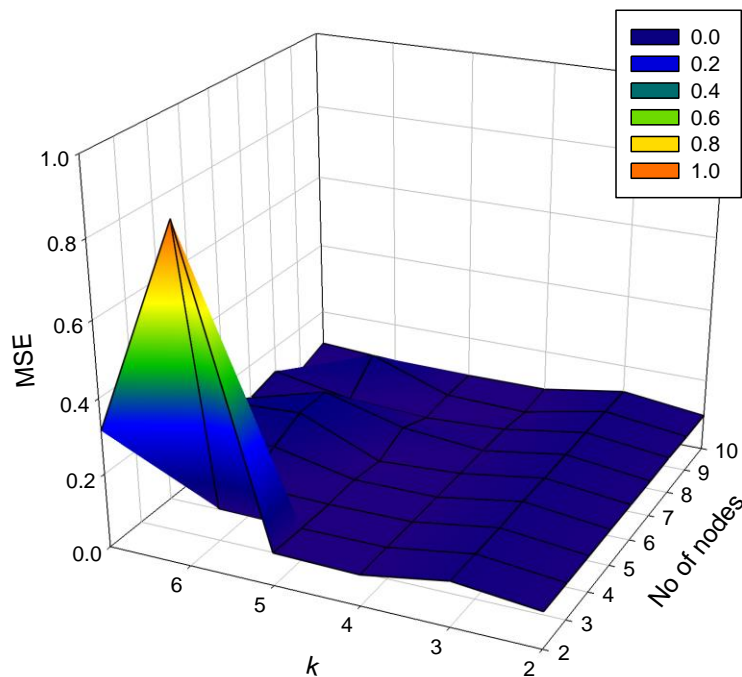
		تعداد توصیفگرها					
		۲	۳	۴	۵	۶	۷
تعداد گره ها در لایه پنهان	۲	۰/۰۹۵۹	۰/۱۲۲۰	۰/۰۸۴۲	۰/۰۸۵۰	۰/۰۸۸۱	۰/۱۴۸۱
	۳	۰/۰۹۵۹	۰/۱۲۳۱	۰/۰۸۵۰	۰/۰۸۴۳	۰/۰۸۶۷	۰/۳۰۴۴
	۴	۰/۰۹۵۹	۰/۱۲۳۷	۰/۰۸۵۹	۰/۰۷۸۶	۰/۰۸۶۱	۰/۱۵۶۹
	۵	۰/۰۹۵۹	۰/۱۲۳۸	۰/۰۸۷۴	۰/۰۸۵۲	۰/۰۸۶۹	۰/۱۰۹۱
	۶	۰/۰۹۶۰	۰/۱۲۴۵	۰/۰۸۵۳	۰/۰۸۲۳	۰/۰۸۷۶	۰/۱۶۹۹
	۷	۰/۱۱۷۱	۰/۱۲۴۳	۰/۰۹۳۹	۰/۰۸۲۳	۰/۰۸۷۶	۰/۱۰۷۲
	۸	۰/۱۹۷۲	۰/۱۲۳۸	۰/۰۸۹۰	۰/۰۸۶۰	۰/۰۸۷۳	۰/۱۳۳۹
	۹	۰/۰۹۶۱	۰/۱۲۵۳	۰/۰۸۸۹	۰/۰۸۴۷	۰/۰۸۸۹	۰/۰۹۶۴
	۱۰	۰/۲۳۹۶	۰/۱۲۵۱	۰/۰۸۷۹	۰/۰۸۷۰	۰/۲۲۷۸	۰/۱۱۰۵



شکل (۴-۴) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی

جدول (۴-۹) - مقادیر میانگین مربعات خطاهای سری ارزیابی 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 توصیفگرها					
		۲	۳	۴	۵	۶	۷
تعداد گره ها در لایه پنهان	۲	۰/۰۹۸۴	۰/۱۲۱۹	۰/۰۸۳۶	۰/۰۸۸۶	۰/۰۸۸۲	۰/۳۱۵۶
	۳	۰/۰۹۸۰	۰/۱۲۰۸	۰/۰۸۵۲	۰/۰۸۴۵	۰/۰۸۸۷	۰/۳۱۵۷
	۴	۰/۰۹۷۸	۰/۱۲۰۴	۰/۰۸۵۴	۰/۰۸۳۲	۰/۰۸۸۳	۰/۱۶۰۴
	۵	۰/۰۹۸۲	۰/۱۲۳۸	۰/۰۸۶۰	۰/۰۸۵۷	۰/۱۳۱۸	۰/۱۱۲۱
	۶	۰/۰۹۷۶	۰/۱۲۲۴	۰/۰۸۶۹	۰/۰۷۴۰	۰/۱۶۶۹	۰/۱۵۴۹
	۷	۰/۰۹۷۵	۰/۱۲۴۳	۰/۰۸۸۴	۰/۱۰۸۳	۰/۱۶۷۱	۰/۱۰۶۸
	۸	۰/۰۹۸۹	۰/۱۲۴۷	۰/۰۸۸۲	۰/۰۷۷۱	۰/۰۸۸۷	۰/۱۲۸۳
	۹	۰/۰۹۷۴	۰/۱۲۴۸	۰/۰۷۸۷	۰/۰۸۵۱	۰/۱۶۲۱	۰/۰۷۳۵
	۱۰	۰/۰۹۷۷	۰/۱۲۴۵	۰/۰۸۸۰	۰/۰۸۶۹	۰/۰۸۹۴	۰/۱۰۲۶



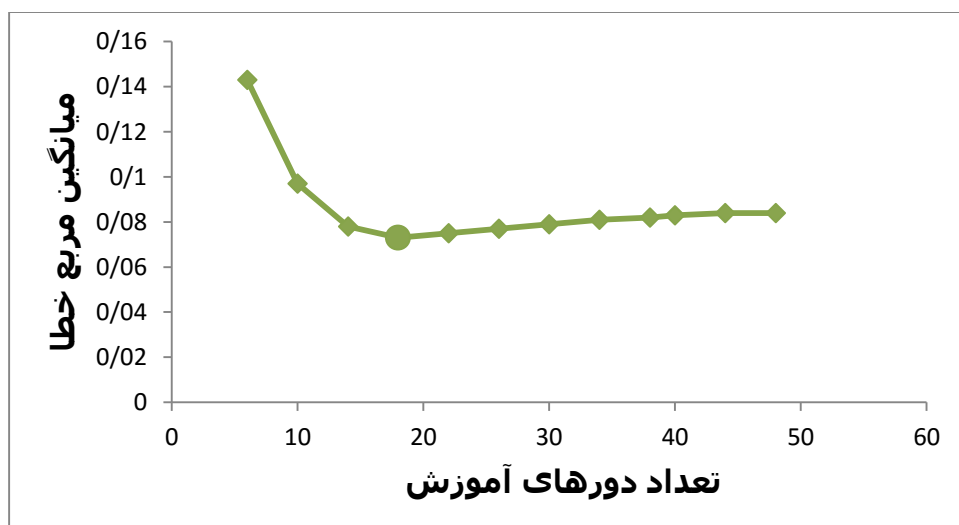
شکل (۴-۵) - نمودار سه بعدی خطا بر حسب تعداد توصیفگر و تعداد گره برای شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

#### ۴-۱-۶-۲- بهینه سازی تعداد دوره‌های آموزش

تعداد چرخه‌های آموزشی پارامتری است که جهت جلوگیری از آموزش بیش از حد شبکه بهینه سازی می شود. بدین منظور، ابتدا مقدار بهینه‌ی تعداد نرون ها (۶) در شبکه مورد نظر با تعداد متغیر و تابع آموزش بهینه قرار داده شد و با تغییر تعداد دوره‌های آموزشی با فواصل یکسان از ۶ تا ۴۸ ، مقادیر خطای شبکه‌ها بدست آمد. نتایج برای سری ارزیابی در جدول (۴-۱۰) و سپس منحنی میانگین مربع خطاها بر حسب تعداد دوره‌های آموزش در شکل (۴-۶) آورده شده است. مقدار بهینه دورها ۱۸ بدست آمد.

جدول (۴-۱۰)- مقادیر میانگین مربع خطای شبکه با تغییر تعداد دوره‌های آموزشی

تعداد دور	میانگین مربع خطا	تعداد دور	میانگین مربع خطا	تعداد دور	میانگین مربع خطا
۶	۰/۱۴۳	۲۲	۰/۰۷۵	۳۸	۰/۰۸۲
۱۰	۰/۰۹۷	۲۶	۰/۰۷۷	۴۰	۰/۰۸۳
۱۴	۰/۰۷۸	۳۰	۰/۰۷۹	۴۴	۰/۰۸۴
۱۸	۰/۰۷۴	۳۴	۰/۰۸۱	۴۸	۰/۰۸۴



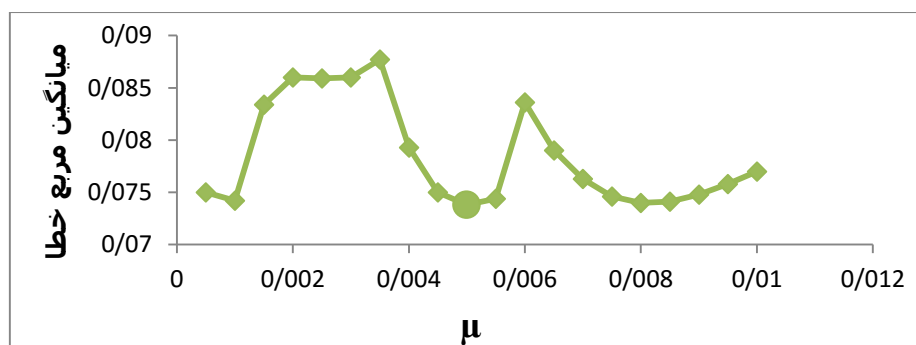
شکل (۴-۶)- نمودار میانگین مربع خطاها بر حسب تعداد دوره‌های آموزش متفاوت

#### ۴-۱-۶-۳- بهینه سازی پارامتر $\mu$

جهت بهینه‌سازی مقدار  $\mu$  (ضریب لونیبرگ)، ساختار شبکه با ۵ متغیر ورودی، ۶ گره در لایه مخفی و الگوریتم آموزشی تنظیم بایزین در نظر گرفته شد. سپس مقدار پارامتر  $\mu$  از ۰/۰۰۰۵ تا ۰/۰۲ با گام‌های ۰/۰۰۰۵ تغییر داده شد و میانگین مربع خطا برای هر مورد سری ارزیابی محاسبه گردید. پس از آن نمودار مقادیر میانگین مربع خطا بر حسب  $\mu$  ترسیم گردید. نقطه‌ی دارای کمترین خطا برای سری ارزیابی، به عنوان مقدار بهینه شده انتخاب شد که به علت زیاد بودن داده‌ها و با توجه به این که مقدار بهینه‌ی  $\mu$ ، ۰/۰۰۵ به دست آمده و در رنج ۰/۰۰۰۵ تا ۰/۰۱ قرار دارد، فقط داده‌های این رنج در جدول (۴-۱۱) و شکل (۴-۷) گزارش شده است.

جدول (۴-۱۱)- مقادیر میانگین مربع خطا در مقادیر متفاوت  $\mu$

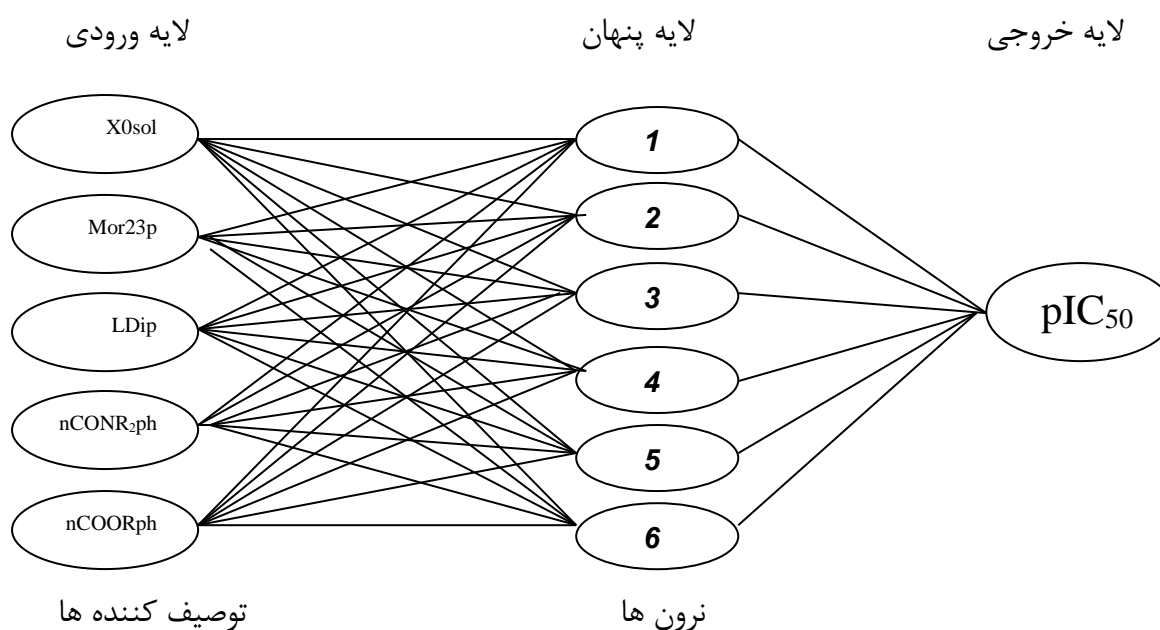
مقادیر $\mu$	میانگین مربع خطا	مقادیر $\mu$	میانگین مربع خطا
۰/۰۰۰۵	۰/۰۷۵۰	۰/۰۰۵۵	۰/۰۷۴۴
۰/۰۰۱	۰/۰۷۴۲	۰/۰۰۶	۰/۰۸۳۶
۰/۰۰۱۵	۰/۰۸۳۴	۰/۰۰۶۵	۰/۰۷۹۰
۰/۰۰۲	۰/۰۸۶۰	۰/۰۰۷	۰/۰۷۶۳
۰/۰۰۲۵	۰/۰۸۵۹	۰/۰۰۷۵	۰/۰۷۴۶
۰/۰۰۳	۰/۰۸۶۰	۰/۰۰۸	۰/۰۷۴۰
۰/۰۰۳۵	۰/۰۸۷۷	۰/۰۰۸۵	۰/۰۷۴۱
۰/۰۰۴	۰/۰۷۹۳	۰/۰۰۹	۰/۰۷۴۸
۰/۰۰۴۵	۰/۰۷۵۰	۰/۰۰۹۵	۰/۰۷۵۸
۰/۰۰۵	۰/۰۷۳۸	۰/۰۱	۰/۰۷۷۰



شکل (۴-۷)- نمودار میانگین مربع خطاها بر حسب مقادیر پارامتر  $\mu$

#### ۴-۶-۱-۴- مشخصات و هندسه‌ی شبکه‌ی عصبی بهینه

با توجه به نتایج مرحله بهینه‌سازی، شبکه‌ی عصبی که بتواند فعالیت ضدانگلی مشتقات بنزیمیدازول را پیش‌بینی کند، شبکه‌ای پیش‌خور و سه لایه با الگوی آموزشی پس‌انتشار، الگوریتم بایزین و تابع انتقال لگاریتم زیگموئید می‌باشد. ساختار هندسی این شبکه‌ی بهینه‌ی انتخاب شده در شکل (۴-۸) رسم شده است.



شکل (۴-۸) - هندسه شبکه‌ی عصبی بهینه

پارامترهای بهینه‌شده‌ی شبکه‌ی عصبی برای ترکیبات مورد بررسی در جدول (۴-۱۲) ارائه شده است.

جدول (۴-۱۲) - پارامترهای بهینه‌شده‌ی شبکه‌ی عصبی

تابع آموزش	Train br
تابع انتقال لایه پنهان	Logsig
تعداد نرون لایه پنهان	۶
تعداد متغیرهای ورودی	۵
تعداد دوره‌های آموزش	۱۸
پارامتر $\mu$	۰/۰۰۵

## ۴-۲- ارزیابی مدل‌های بدست آمده و مقایسه مدل‌ها

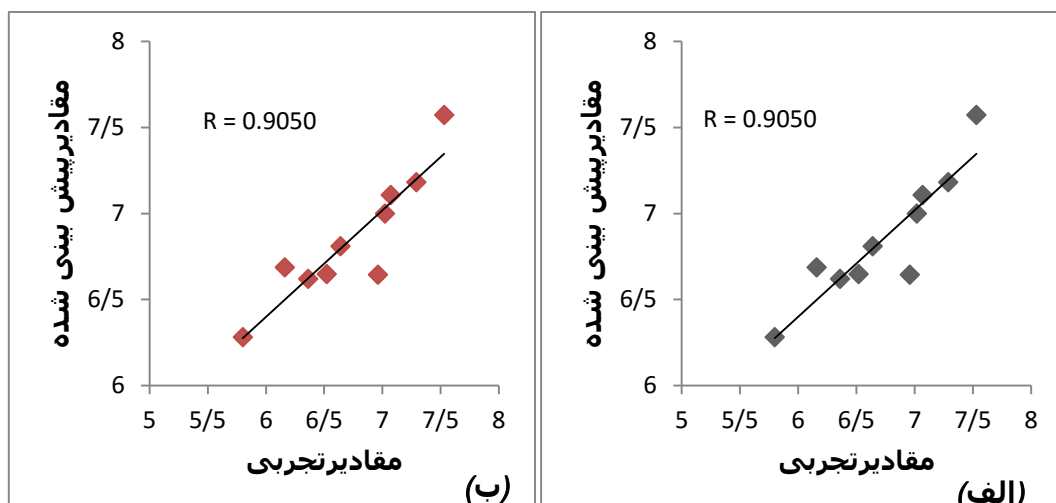
### ۴-۲-۱- ارزیابی مدل‌ها با استفاده از سری ارزیابی

جهت ارزیابی مدل با استفاده از سری ارزیابی، در مدل غیرخطی ابتدا به وسیله‌ی سری آموزش، شبکه آموزش داده شد و بعد از آن شبکه‌ی آموزش دیده برای پیش بینی داده‌های  $pIC_{50}$  برای سری ارزیابی (شامل ۱۰ مولکول که به طور تصادفی انتخاب شده‌اند) مورد استفاده قرار گرفت و درصد خطای نسبی برای مقادیر پیش بینی شده محاسبه گردید. در مدل خطی نیز جهت ارزیابی از ضرائب مدل برتر استفاده شد. نتایج حاصل در جدول (۴-۱۳) و نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی نیز در شکل (۴-۹) نشان داده شده است.

جدول (۴-۱۳)- نتایج حاصل از ارزیابی مدل‌های برتر MLR و ANN با استفاده از سری ارزیابی

شماره ترکیب	$pIC_{50}$				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۱	۵/۸۰	۶/۲۷	۶/۲۸	۸/۱۰	۸/۲۸
۵	۶/۵۲	۶/۶۹	۶/۶۵	۲/۶۱	۱/۹۹
۱۱	۶/۹۶	۶/۶۸	۶/۶۴	-۴/۰۲	-۴/۶۰
۱۶	۶/۶۴	۶/۸۷	۶/۸۱	۳/۴۶	۲/۵۶
۲۱	۷/۰۲	۷/۰۹	۷/۰۰	۰/۱۰	-۰/۲۸
۲۶	۷/۵۳	۷/۷۴	۷/۵۷	۲/۷۹	۰/۵۳
۳۱	۶/۳۶	۶/۶۳	۶/۶۲	۴/۲۵	۴/۰۹
۳۷	۶/۱۶	۶/۷۳	۶/۶۹	۹/۲۵	۸/۶۰
۴۲	۷/۰۷	۷/۲۱	۷/۱۲	۱/۹۸	۰/۷۱
۴۶	۷/۲۹	۷/۳۰	۷/۱۸	۰/۱۴	-۱/۶۴





شکل (۴-۹) - نمودار مقادیر پیش بینی  $pIC_{50}$  بر حسب مقادیر تجربی برای ارزیابی الف) با مدل MLR ب) با مدل ANN

#### ۴-۲-۲- ارزیابی مدل ها با استفاده از سری تست

مطالعه‌ی قدرت پیش‌بینی مدل‌های منتخب خطی و غیرخطی با استفاده از سری تست، در مدل خطی با استفاده از ضرائب مدل برتر و در مدل غیرخطی با استفاده از شبکه عصبی بهینه شده بوسیله ۱۰ داده‌ی سری تست که در آموزش استفاده نشده‌اند، انجام شد. نتایج حاصله در جدول (۴-۱۴) و همچنین شکل (۴-۱۰) نشان داده شده است که حاکی از برتری مدل خطی بر مدل غیرخطی می‌باشند.

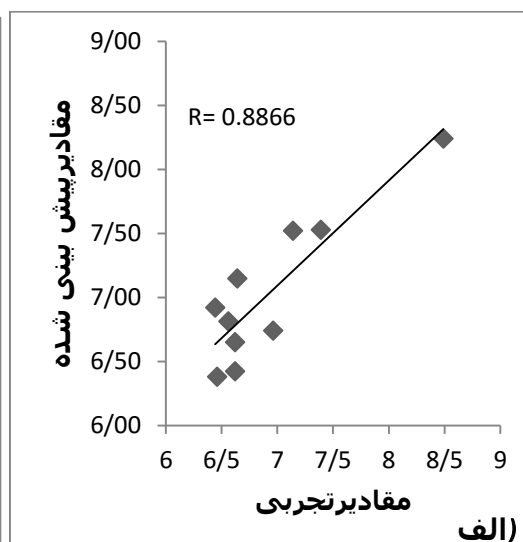
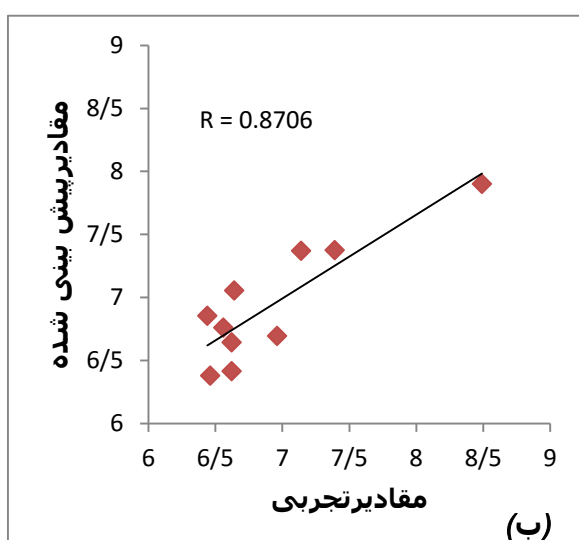
#### ۴-۲-۳- ارزیابی مدل ها به روش رد مرحله ای تک تک

از دیگر روشهایی که برای ارزیابی بیشتر مدل‌های برتر ارائه شده توسط MLR و ANN به کار رفته است، تکنیک رد مرحله ای تک تک می باشد که نتایج حاصل از این روش در جدول (۴-۱۵) و شکل (۴-۱۱) ارائه شده است. این نتایج تعمیم پذیری بالای مدل طراحی شده توسط هر دو روش خطی و غیرخطی را بیان می‌کند. همچنین در شکل (۴-۱۲) مقادیر باقیمانده‌ها بر حسب مقادیر

pIC<sub>50</sub> تجربی ترکیبات مورد بررسی ترسیم شده است که تمرکز نقاط به سمت محور صفر، نشان دهنده‌ی عدم وجود خطای سیستماتیک در مدل است.

جدول (۴-۱۴) - نتایج حاصل از ارزیابی مدل های برتر MLR و ANN با استفاده از سری تست

شماره ترکیب	pIC <sub>50</sub>				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۴	۶/۵۶	۶/۸۱	۶/۷۶	۳/۸۱	۳/۰۵
۸	۶/۴۶	۶/۳۸	۶/۳۸	-۱/۲۴	-۱/۲۴
۱۳	۶/۶۲	۶/۶۵	۶/۶۴	۰/۴۵	۰/۳۰
۱۸	۶/۴۴	۶/۹۲	۶/۸۶	۷/۴۵	۶/۵۲
۲۴	۶/۶۴	۷/۱۵	۷/۰۶	۷/۶۸	۶/۳۳
۳۰	۶/۹۶	۶/۷۴	۶/۷۰	-۳/۱۶	-۳/۷۴
۳۵	۶/۶۲	۶/۴۲	۶/۴۱	-۳/۰۲	-۳/۱۷
۴۰	۷/۳۹	۷/۵۳	۷/۳۸	۱/۸۹	-۰/۱۴
۴۴	۷/۱۴	۷/۵۲	۷/۳۷	۵/۳۲	۳/۲۲
۴۹	۸/۴۹	۸/۲۴	۷/۹۰	-۲/۹۴	-۶/۹۵



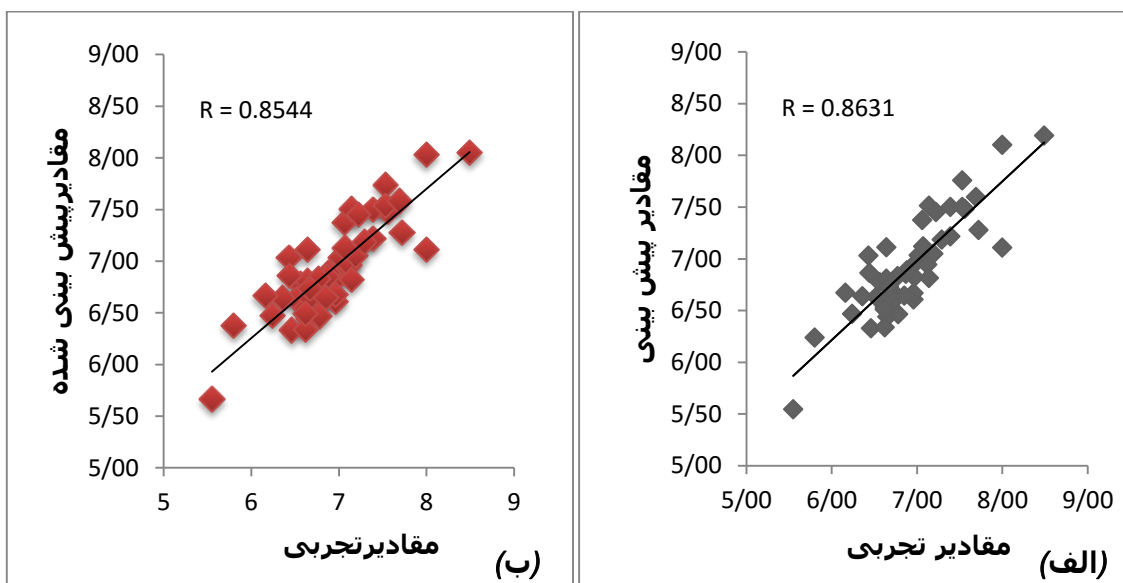
شکل (۴-۱۰) - نمودار مقادیر پیش بینی pIC<sub>50</sub> بر حسب مقادیر تجربی برای تست الف) با مدل MLR (ب) با مدل ANN

جدول (۴-۱۵) - نتایج حاصل از ارزیابی مدل ANN و MLR با روش رد مرحله ای تک تک

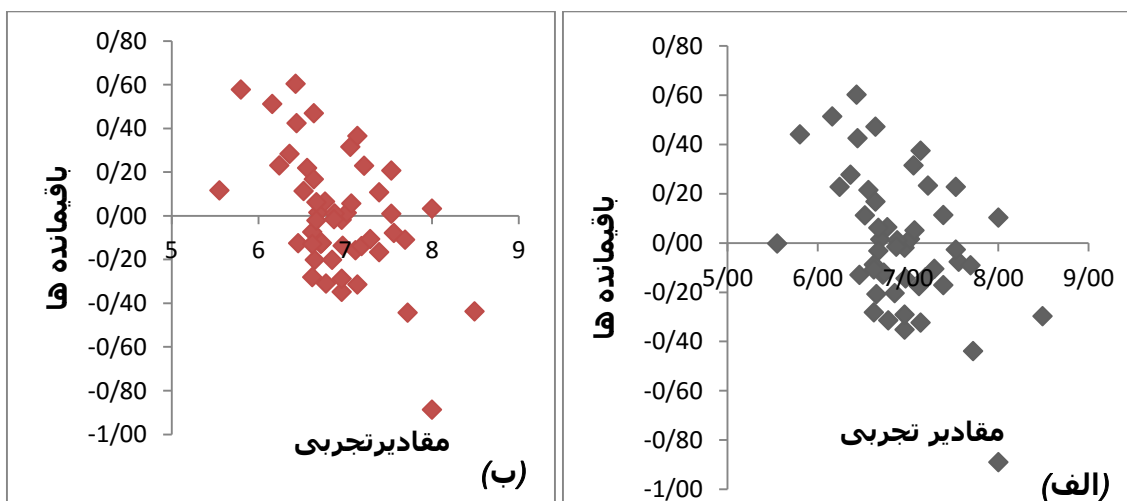
شماره ترکیب	pIC <sub>50</sub>				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۱	۵/۸۰	۶/۲۴	۶/۳۸	۷/۵۹	۱۰/۰۰
۲	۶/۶۵	۶/۴۴	۶/۴۵	-۳/۱۶	-۳/۰۱
۳	۶/۷۲	۶/۵۹	۶/۵۹	-۱/۹۳	-۱/۹۳
۴	۶/۵۶	۶/۷۸	۶/۷۸	۳/۳۵	۳/۳۵
۵	۶/۵۲	۶/۶۳	۶/۶۳	۱/۶۹	۱/۶۹
۶	۶/۲۴	۶/۴۷	۶/۴۷	۳/۶۸	۳/۶۸
۷	۶/۴۳	۷/۰۳	۷/۰۳	۹/۳۳	۹/۳۳
۸	۶/۴۶	۶/۳۳	۶/۳۴	-۲/۰۱	-۱/۸۶
۹	۵/۵۵	۵/۵۵	۵/۶۷	۰/۰۰	۲/۱۶
۱۰	۶/۶۹	۶/۷۰	۶/۷۱	۰/۱۵	۰/۳۰
۱۱	۶/۹۶	۶/۶۱	۶/۶۱	-۵/۰۳	-۵/۰۳
۱۲	۶/۹۷	۶/۸۳	۶/۸۳	-۲/۰۱	-۲/۰۱
۱۳	۶/۶۲	۶/۵۳	۶/۵۵	-۱/۳۶	-۱/۰۶
۱۴	۷/۷۲	۷/۲۸	۷/۲۸	-۵/۷۰	-۵/۷۰
۱۵	۶/۹۶	۶/۹۴	۶/۹۴	-۰/۲۹	-۰/۲۹
۱۶	۶/۶۴	۶/۸۱	۶/۸۱	۲/۵۶	۲/۵۶
۱۷	۶/۷۳	۶/۶۱	۶/۶۱	-۱/۷۸	-۱/۷۸
۱۸	۶/۴۴	۶/۸۷	۶/۸۶	۶/۶۸	۶/۵۲
۱۹	۶/۶۷	۶/۶۴	۶/۶۵	-۰/۴۵	-۰/۳۰
۲۰	۷/۵۶	۷/۴۸	۷/۴۸	-۱/۰۶	-۱/۰۶
۲۱	۷/۰۲	۷/۰۴	۷/۰۳	۰/۲۸	۰/۱۴
۲۲	۶/۸۸	۶/۸۹	۶/۸۹	۰/۱۵	۰/۱۵
۲۳	۶/۸۷	۶/۸۶	۶/۸۶	-۰/۱۵	-۰/۱۵

شماره ترکیب	pIC <sub>50</sub>				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۲۴	۶/۶۴	۷/۱۱	۷/۱۱	۷/۰۸	۷/۰۸
۲۵	۷/۱۲	۶/۹۵	۶/۹۶	-۲/۳۹	-۲/۲۵
۲۶	۷/۵۳	۷/۷۶	۷/۷۴	۳/۰۵	۲/۷۹
۲۷	۷/۳۹	۷/۲۲	۷/۲۲	-۲/۳۰	-۲/۳۰
۲۸	۷/۱۹	۷/۰۵	۷/۰۵	-۱/۹۵	-۱/۹۵
۲۹	۶/۷۸	۶/۴۷	۶/۴۷	-۴/۵۷	-۴/۵۷
۳۰	۶/۹۶	۶/۶۷	۶/۶۷	-۴/۱۷	-۴/۱۷
۳۱	۶/۳۶	۶/۶۴	۶/۶۴	۴/۴۰	۴/۴۰
۳۲	۷/۰۶	۷/۳۸	۷/۳۸	۴/۵۳	۴/۵۳
۳۳	۶/۷۷	۶/۸۳	۶/۸۴	۰/۸۹	۱/۰۳
۳۴	۶/۶۷	۶/۷۳	۶/۷۳	۰/۹۰	۰/۹۰
۳۵	۶/۶۲	۶/۳۴	۶/۳۴	-۴/۲۳	-۴/۲۳
۳۶	۶/۶۲	۶/۵۱	۶/۴۹	-۱/۶۶	-۱/۹۶
۳۷	۶/۱۶	۶/۶۷	۶/۶۷	۸/۲۸	۸/۲۸
۳۸	۶/۸۵	۶/۶۵	۶/۶۵	-۲/۹۲	-۲/۹۲
۳۹	۷/۱۴	۶/۸۲	۶/۸۳	-۴/۴۸	-۴/۳۴
۴۰	۷/۳۹	۷/۵۰	۷/۵۰	۱/۴۹	۱/۴۹
۴۱	۷/۵۳	۷/۵۰	۷/۵۴	-۰/۴۰	۰/۱۳
۴۲	۷/۰۷	۷/۱۲	۷/۱۳	۰/۷۱	۰/۸۵
۴۳	۸/۰۰	۷/۱۱	۷/۱۱	-۱۱/۱۳	-۱۱/۱۳
۴۴	۷/۱۴	۷/۵۱	۷/۵۱	۵/۱۸	۵/۱۸
۴۵	۷/۶۹	۷/۶۰	۷/۵۸	-۱/۱۷	-۱/۴۳
۴۶	۷/۲۹	۷/۱۹	۷/۱۸	-۱/۳۷	-۱/۵۱

شماره	pIC <sub>50</sub>				
	مقدار واقعی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۴۷	۷/۲۲	۷/۴۵	۷/۴۵	۳/۱۹	۳/۱۹
۴۸	۸/۰۰	۸/۱۰	۸/۰۳	۱/۲۵	۰/۳۸
۴۹	۸/۴۹	۸/۱۹	۸/۰۵	-۳/۵۳	-۵/۱۸



شکل (۴-۱۱) - نمودار مقادیر پیش بینی pIC<sub>50</sub> بر حسب مقادیر تجربی کل داده ها الف) با مدل MLR ب) با مدل ANN



شکل (۴-۱۲) - نمودار باقیمانده ها بر حسب مقادیر واقعی برای کل داده ها الف) با مدل MLR ب) با مدل ANN

#### ۴-۲-۴- آزمون Y- تصادفی

این آزمون جهت اثبات قدرت مدل ایجاد شده و شانس نبودن نتایج بدست آمده انجام می‌گیرد. در این مرحله، یک مجموعه اعداد تصادفی در دامنه‌ی متغیر هدف، تعریف کرده و بهترین شبکه با آن آزموده شد. لذا داده‌های pIC<sub>50</sub> بطور تصادفی ایجاد شدند و مدل QSAR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از مقادیر بازدارنده‌ها توسعه یافت. نتایج حاصل از چندین مجموعه‌ی تصادفی ایجاد شده در جدول (۴-۱۶) نشان داده شده است که با توجه به نتایج، هیچ وابستگی برجسته‌ای بین خواص ساختاری و داده‌های بازداری برای ترکیبات وجود ندارد (ضریب همبستگی پایین)، که نشان دهنده‌ی عدم همبستگی تصادفی در مدل شبکه عصبی مصنوعی ایجاد شده می‌باشد.

جدول (۴-۱۶)- مقادیر R برای سری ارزیابی و سری تست بعد از ۱۰ بار تست Y-تصادفی

تکرار	سری ارزیابی	سری تست
۱	۰/۰۲۹۶	۰/۲۲۱۱
۲	-۰/۰۷۴۲	۰/۲۰۱۹
۳	۰/۱۴۴۷	۰/۲۸۶۶
۴	۰/۰۵۵۹	۰/۳۷۹۶
۵	۰/۲۴۶۴	-۰/۱۶۲۳
۶	۰/۶۹۳۰	۰/۰۹۷۹
۷	-۰/۰۲۳۱	۰/۲۱۲۴
۸	۰/۲۸۳۳	۰/۱۲۹۹
۹	-۰/۱۷۵۸	-۰/۵۲۰۱
۱۰	۰/۰۶۳۲	-۰/۰۷۳۲

## ۴-۲-۵- ارزیابی مدل‌های ارائه شده با استفاده از پارامترهای آماری

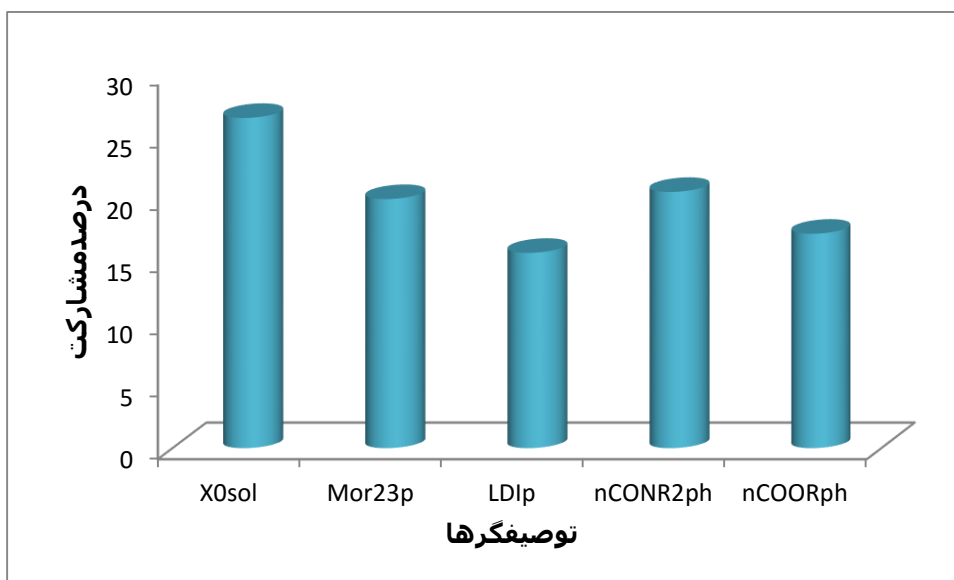
برای ارزیابی کارایی شبکه عصبی مصنوعی و رگرسیون خطی چندگانه‌ی ارائه شده برای پیش بینی بازدارندگی ترکیبات بنزیمیدازول، چهار پارامتر آماری MSE، R، REP و PRESS محاسبه شد (در بخش ۲-۲-۸ توضیح داده شده است) و نتایج حاصل نیز در جدول (۴-۱۷) ذکر شده است. جدول (۴-۱۷) - پارامترهای آماری محاسبه شده

پارامتر	سری ارزیابی (n=10)		سری تست (n=10)		کل داده‌ها (n=49)	
	ANN	MLR	ANN	MLR	ANN	MLR
MSE	۰/۰۷۴۰	۰/۰۸۴۰	۰/۰۸۹۰	۰/۰۸۷۰	۰/۰۷۹۲	۰/۰۷۴۸
PRESS	۰/۷۴۰۰	۰/۸۴۰۰	۰/۸۹۰۰	۰/۸۷۰۰	۳/۸۸۰۸	۳/۶۶۵۲
REP(%)	۴/۰۳۹۰	۴/۳۰۳۰	۴/۳۰۳۰	۴/۲۵۵۰	۴/۰۶۸۰	۳/۹۴۹۰
R	۰/۹۰۵۰	۰/۹۰۵۰	۰/۸۷۰۶	۰/۸۸۶۶	۰/۸۵۴۴	۰/۸۶۳۱

با توجه به نتایج حاصل در این بخش، ملاحظه می‌شود که شبکه عصبی و رگرسیون خطی چندگانه، هر دو از قدرت پیش بینی نسبتاً بالایی برخوردارند اما با توجه به نتایج سری تست، روش رگرسیون خطی چندگانه در پیش بینی فعالیت این ترکیبات دارویی نتایج بهتری ارائه می‌دهد.

## ۴-۲-۶- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی

بر اساس روش ذکر شده در بخش (۳-۲-۶)، میزان مشارکت توصیفگرهای منتخب در پیش‌بینی فعالیت دارویی مورد نظر برآورد شد و نتایج در شکل (۴-۱۳) ارائه شده است.



شکل (۴-۱۳) - مشارکت توصیف‌گرها در شبکه عصبی بهینه

بر اساس نمودار فوق توصیف‌گرهای X0sol و nCONR2ph دارای بیشترین اثر مشارکت می باشند. توصیف‌گر X0sol در برگیرنده اطلاعات دوبعدی مولکول است بنابراین می توان گفت که فعالیت بازداری این ترکیبات وابستگی قابل ملاحظه‌ای به توپولوژی ترکیب بازدارنده دارد. همچنین توصیف‌گر nCONR2ph از دسته توصیف‌گرهای یک بعدی است که نقش عامل آمیدی را در میزان بازدارندگی ترکیب دارویی نشان می‌دهد.

#### ۴-۲-۷- بررسی ارتباط توصیف‌گرهای انتخاب شده با فعالیت بازدارندگی

در این بخش با توجه به توصیف‌کننده‌های وارد شده در مدل‌های ANN و MLR یک بررسی اجمالی روی اثرات مختلف این متغیرها بر پارامتر بازداری ترکیبات مورد مطالعه صورت می‌گیرد. بهترین مدل منتخب شامل ۵ توصیف‌کننده است که هر کدام بیانگر خصوصیات متفاوتی از مولکولهای مورد بررسی می‌باشد.



## ۴-۲-۷-۱- توصیفگرهای Topological [۲۶]

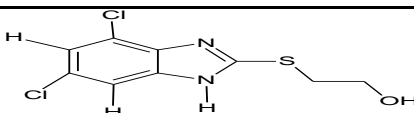
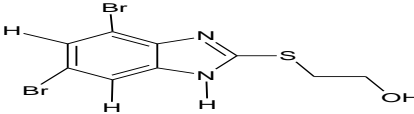
این دسته توصیفگرها که از گراف مولکولی تهی از هیدروژن محاسبه می‌شوند، از ساده‌ترین توصیفگرهای مولکولی بوده و به ساختار فضایی مولکول ارتباطی ندارند. توصیفگرهای توپولوژیکی نشان‌دهنده‌ی نوع اتم، نوع و تعداد پیوند و چگونگی ارتباط با یکدیگر بوده و به سادگی از روی ساختمان دوبعدی مولکول قابل محاسبه‌اند. از میان این توصیفگرها،  $X_{0sol}$  که اندیس ارتباطی حلال پوشی نام دارد، توسط مدل انتخاب شده است. به طور کلی این توصیفگرهای مولکولی آنتروپی حلال پوشی و برهمکنش‌های پراکندگی در محلول را شرح می‌دهند که بصورت زیر تعریف می‌شود:

$$X_{q}^S = \frac{1}{2^{m+1}} \cdot \sum_{k=1}^K \frac{(\prod_{a=1}^n L_a)_k}{(\prod_{a=1}^n \delta_a)_k^{1/2}} \quad \text{رابطه ی (۱-۴)}$$

که  $L_a$  عدد کوانتومی اصلی (۲ برای اتم C، N و O و ۳ برای Si، S و CL.....) مربوط به  $a$  امین اتم در  $K$  امین زیر گراف و  $\delta_a$  درجه ی رأس،  $K$  تعداد کل زیرگراف های  $m$  ام و  $n$  نیز تعداد یال ها در زیر گراف می باشد.

با توجه به اثر مثبت این توصیفگر، با افزایش مقدار این توصیفگر، قدرت بازدارندگی ترکیب افزایش یافته است. همچنین طبق تعریف این توصیفگر (رابطه (۱-۴)) با بزرگتر شدن مولکول ها، بدلیل بزرگتر شدن صورت کسر، مقدار این توصیفگر و متناسب با آن،  $pIC_{50}$  افزایش می یابد که این اثر به وسیله ی چند مثال در جدول (۴-۱۸) نشان داده شده است. ظاهر شدن این توصیفگر نشان دهنده‌ی اهمیت آنتروپی حلال پوشی در فعالیت دارویی این ترکیبات می‌باشد.

جدول (۴-۱۸)- چند مثال از اثر توصیفگر  $X_{0sol}$

شماره ترکیب	ترکیب	$pIC_{50}$
۳۷		۶/۱۶
۳۸		۶/۱۵

<sup>1</sup>- Solvation connectivity index

#### ۲-۷-۲-۴- توصیفگرهای 3D MORSE [۴۹ و ۲۶]

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

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

I شدت الکترون پراکنده شده،  $A_i$  و  $A_j$  خاصیت اتمیاتی  $i$  و  $j$ ،  $s$  زاویه پراکندگی،  $r_{ij}$  فاصله بین اتم-های  $i$  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.

از این نمونه توصیفگرهای سه بعدی،  $\text{Mor23p}$  در مدل برتر ظاهر شده است که با توجه به ضریب منفی این توصیفگر در مدل برتر، با افزایش مقدار این توصیفگر مقدار  $\text{pIC}_{50}$  کاهش یافته و به دنبال آن بازدارندگی ترکیب کمتر می‌شود.

#### ۳-۷-۲-۴- توصیفگرهای Charge

با توجه به تئوری کلاسیک، برهم‌کنشهای بین مواد از نظر ماهیت، الکتروستاتیک یا کووالانسی است. تغییرات بار الکتریکی در مولکول به عنوان نیروی پیش‌برنده‌ی برهم‌کنشهای الکتروستاتیک است. از جمله این توصیفگرها که در مدل برگزیده شده  $\text{LDip}$ ، شاخص دوقطبی محلی، می‌باشد. علامت منفی آن در مدل، اهمیت برهم‌کنشهای الکتروستاتیک بین بازدارنده و مولکول هدف را نشان می‌دهد و به عنوان یک شاخص برای اندازه‌گیری میزان این برهم‌کنشهای الکتروستاتیک به کار می‌رود و با افزایش مقدار این توصیفگر، مقدار  $\text{pIC}_{50}$  کاهش می‌یابد [۶۰].

<sup>1</sup>-3D- Molecular Representation of Structure based on Electron diffraction

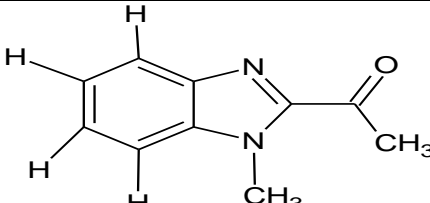
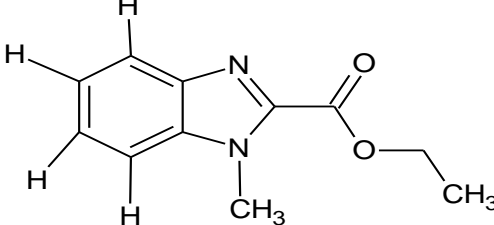
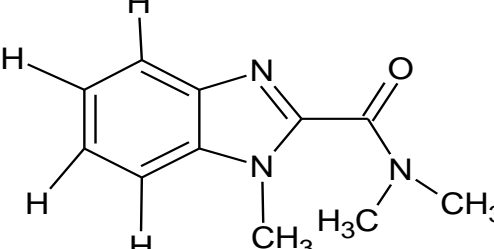
<sup>2</sup>-3D – MORSE-signal 23 weighted by polarizabilit

<sup>3</sup>- Local Dipole index

#### ۴-۷-۲-۴- توصیفگرهای functional groups

می شوند، که اولی تعداد آمیدهای آروماتیک نوع سوم و دومی تعداد استرهای آروماتیک را در ترکیب نشان می دهد. با توجه به ضریب منفی اولی و ضریب مثبت دومی در مدل برتر، با افزایش مقدار هرکدام از این دو توصیفگر، مقدار  $pIC_{50}$  برای اولی کاهش و برای دومی افزایش می یابد. به عبارت دیگر با افزایش تعداد عامل استری و کاهش تعداد عامل آمیدی در ترکیب، قدرت بازداري ترکیب دارویی افزایش می یابد. در جدول (۴-۱۹) با چند مثال تأثیر توصیفگر  $nCOOR_{ph}$  و  $nCONR_2$  بر فعالیت دارویی نشان داده شده است.

جدول (۴-۱۹)- چند مثال از اثر توصیفگر  $nCONR_2$  و  $nCOOR_{ph}$  بر فعالیت دارویی

شماره ترکیب	ترکیب	$pIC_{50}$
۳۴		۶/۶۷
۳۲		۷/۰۶
۳۱		۶/۳۶

1- Number of amids(aromatic)

2- Number of esters(aromatic)

3- Functional groups

## ۴-۳- نتیجه‌گیری نهایی

همانطور که گفته شد هدف اصلی از کار حاضر، توسعه روشهای خطی و غیرخطی جهت مطالعه ارتباط کمی بین ساختار و فعالیت ضدانگلی مشتقات بنزیمیدازول می‌باشد. نتایج نشان می‌دهند که مدل‌های خطی و غیرخطی توسعه یافته می‌توانند به عنوان روش‌های موفق برای مدل‌سازی و پیشگویی فعالیت بازدارندگی ترکیبات مورد مطالعه باشند و این مدل‌ها می‌توانند به محققان در طراحی داروهایی با فعالیت بازدارندگی قویتر کمک کنند.

## آینده‌نگری

استفاده از مدل‌های QSAR می‌تواند منجر به صرفه جویی زیادی در وقت و هزینه در راستای طراحی و بهینه‌سازی ساختار داروها شود.

از آن جایی که در تحقیق حاضر توصیفگرهای به دست آمده از روش رگرسیون خطی چندگانه به عنوان ورودی شبکه عصبی مصنوعی قرار گرفته است و چون در روشهای رگرسیون خطی چندگانه تنها برهم‌کنشهای خطی بین پارامترها لحاظ می‌شود، لذا می‌توان از روشهایی همچون ژنتیک الگوریتم، الگوریتم طرح‌ریزی متوالی<sup>۱</sup> برای انتخاب توصیفگرهای مهم استفاده کرد و می‌توان روشهایی مانند SVM<sup>۲</sup> و LS-SVM<sup>۳</sup> را به جای شبکه عصبی به کار برد.

---

1- Successive Projection Algorithm

2- Supported Vector Machine

3- Least Square Supported Vector Machine

## مراجع و منابع

- [۱]- Feldmeier H., Krantz I., (1993), "A synoptic inventory of needs for research on women and tropical parasitic diseases", Acta. Trop, 55, pp 117-138.
- [۲]- [http:// www.parasitic diseases.com/](http://www.parasiticdiseases.com/)
- [۳]- [http:// www.parasitic diseases.com/](http://www.parasiticdiseases.com/)
- [۴]- D.Baker G., Malineni S., Taylor H.M., (1998), "Experimental infection of inbred mouse strains with *spironucleus muris*", vet. parasitol., 77., pp 305-310.
- [۵]- Stear M.J., Murray M., (1994), " Genetic resistance to parasitic disease particularly of resistance in ruminants to gastrointestinal nematodes", Vet. Parasitol, 54., pp 161-176.
- [۶]- Valdez-Padilla D., Rodriguez-Morales S., Hernandez-Campos A., Hernandez-Luis F., Yépez-Mulia L., Tapia-Contreras A., Castillo R., (2009), "Synthesis and antiprotozoal activity of novel 1- methylbenzimidazole derivatives " , Bioorg. Med. Chem., 17, pp 1724-1730.
- [۷]- [http:// www.giardia intestinalis.com/history](http://www.giardiaintestinalis.com/history)
- [۸]- Katz M., Despommier D. D., Gwadz R., (1989), " Parasitic diseases", second ed. New York: springer-verlag.
- [۹]- Garsia L. S., Shimizu R. Y., (1997), "Evaluation of nine immunoassay kits (enzyme immunoassay and direct fluorescence) for detection of *giardia lamblia* in human fecal specimens", J. Clin. Microbiol, 35, pp 1526-1529.
- [۱۰]- Gardner T. B., Hill D. R., (2001), "Treatment of giardiasis", Clin. Microbiol. Rev., 14, pp 114-128.
- [۱۱]- E. Weksberg T., C. Lynch G., Krause K., Pettitt B.M., (2007), "Molecular Dynamics simulations of *Trichomonas vaginalis* ferredoxin show a loop-cap transition", Biophys. J, 92, pp 3337-3345.
- [۱۲]- Chose O., Noel C., Gerbod D., Brenner C., Viscogliosi E., Roseto A., (2002), " A Form of Cell Death with Some Features Resembling Apoptosis in the Mitochondrial Unicellular Organism *Trichomonas vaginalis*", Exp. Cell. Res, 276, PP 32-39.

[۱۳]- دکتر بهادر سرکاری، دکتر حدیثا تدین، شهر بانو عسکریان، الهام فرنیاء، مهر انگیز عسکریان (۱۳۸۸)،

"اثر عصاره سیر و آنغوز هبرر شد و تکثیر انگل تریکوموناس وازینالیس " ، مجله علمی دانشگاه علوم پزشکی گران، شماره ۳، ص ۱۷-

.۱۳

[۱۴]- Meneses-Marcel A., Marrero-Ponce Y., Machado-Tugores Y., Montero-Torres A., Montero Pereira D., Antonio Escario J., Jose Nogal-Ruiz J., Ochoa C., J. Aran V., R. Martinez-Fernandez A., N. Garcia Sanchez R., (2005), "A linear discrimination analysis based virtual screening of trichomonacidal lead- like compounds", Bioorg. Med. Chem. Lett, 15, pp 3838-3843.

[۱۵]- Meneses-Marcel A., Marrero-Ponce Y., Machado-Tugores Y., Montero-Torres A., Montero Pereira D., Antonio Escario J., Jose Nogal-Ruiz J., Ochoa C., J. Aran V., R. Martinez-Fernandez A., N. Garcia Sanchez R., (2006), "predicting antitrichomonal activity: A computational screening using atom-based bilinear indices and experimental proofs", Bioorg. Med. Chem, 14, pp 6502-6524.

[۱۶]- Ozden S., Atabey D., Yildiz S., Goker H., (2008), "Synthesis anti-staphylococcal activity and QSARs of some novel 2- anilinobenzazoles", Eur. J. Med. Chem, 43, pp 1390-1402.

[۱۷]- Tan Sh. M., Jiao J., Zhu X. L., Zhou Y. P., Song D. D., Gong H., Yu R. Q., (2010), "QSAR studies of a diverse series of antimicrobial agents against *Candida albicans* by classification and regression trees", Chemometr. Intell. Lab. Syst, 103, pp 184-190.

[۱۸]- منصور عرب چم جنگلی (۱۳۸۶)، "پیش بینی فعالیت دارویی ضدایدز (سیتوتوکسیتی) مشتقات ۵- فنیل-۱- فنیل آمینو H<sub>1</sub>-ایمیدازول به وسیله شبکه عصبی مصنوعی"، دانشگاه صنعتی شاهرود، گزارش طرح پژوهشی

[۱۹]- Hansch C., Fujita T., (1964), " methods for the correlation of biological activity and chemical structure" , J. Am. Chem. Soc., 86, 1616-1620.

[۲۰]- شعیری ف، (۱۳۸۵)، پایان نامه کارشناسی ارشد، "پیشگویی فعالیت پیرازولوکوئینولین ها با استفاده از شبکه عصبی موجک و کاربرد روش RAFA در اندازه گیری همزمان آمپی سیلین و آموکسی سیلین"، دانشگاه علوم پایه دامغان.

[۲۱]- Guha R., (2005), "Methods to improve the reliability, validity and interpretability of QSAR models", PhD thesis, Pennsylvania state university.

[۲۲]- Yasri A., Hartsough D., (2001), "Toward an optimal procedure for variable selection and QSAR model building", J. Chem. Inf. Comput. Sci. ,41, pp 1218-1227.

[۲۳]- Clark T., (1985) " A Handbook of Computational Chemistry", John Wiley & Sons, New York, pp 304.

[۲۴]- HyperChem7.0 Toronto, Canada: HyperCube Inc, [http:// www.hyper.com](http://www.hyper.com).

[۲۵]- Atkins P.W., Friedman R.S., (1996), "Molecular Quantum Mechanics", 3<sup>rd</sup> Ed, oxford University press, New York, pp 307.

[۲۶]- Todeschini R., Consonni V., (2000), "Handbook of molecular descriptors", John Wiley, New York.

[۲۷]- Pitts W. H, McCulloch W. S, (1947), "*How we know universals: The perception of auditory and visual forms*", Bull. Math. Biophys, 9, pp 127- 147.

[۲۸]- Wasserman, Philip. D., (1989), "*Biostatistics A foundation for Analysis in the Health Science*", John Wiley, New York, pp 290.

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

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

[۳۱]- Rumelhart D.E. and McClelland J.L., (1986), "*Parallel Distributed Processing: Exploration in the Microstructure of Cognition*", vol1, Cambridge, MA: MIT Press, Chapter 1.

[۳۲]- منهای م، (۱۳۸۷) " مبانی شبکه های عصبی (هوش محاسباتی) " ، جلد اول، چاپ پنجم، مرکز نشر دانشگاه صنعتی امیر کبیر، تهران .

[۳۳]- منهای م، (۱۳۸۶) " مبانی شبکه های عصبی " ، مرکز نشر پر و فسور حسابی.

[۳۴]- کیا م. (۱۳۸۷)، " شبکه های عصبی در *MATLAB* "، چاپ اول، انتشارات کیان رایانه، ص ۲۱-۵۹.

[۳۵]- هاگان. ت. ه.، دیموث. ه.، بیل م. (۱۳۸۲) " طراحی شبکه های مصنوعی "، ترجمه ی کیا.م. انتشارات کیان رایانه سبز.

[۳۶]- کیا.م. (۱۳۸۷) " شبکه های عصبی در *MATLAB* "، چاپ دوم، انتشارات کیان رایانه سبز

[۳۷]- <http://foram.takdownload.ir/threads/25029-؟-شبکه-های-مصنوعی-چیست-؟>

[۳۸]- Kermani B.G., Schiffman S.S., Troy Nagle H., (2005), "*Performance of the Levenberg-Marquardt neural network training method in electronic nose applications*", sensors and Actuators B, 110, pp 13-22.

[۳۹]- Wang X.G, Tang Z., Tamura H., Ishii M., (2004), "*A modified error function for the back propagation algorithm*", Neurocomputing, 57, pp 477- 484.

[۴۰]- Wang X.G., Tang Z., Tamura H., Ishii M., Sun W., (2004), "*An improved backpropagation algorithm to avoid the local minima problem*", Neurocomputing, 56, pp 455- 460.

[۴۱]- Hagan M. T., Menhaj M. B., 1994, "*Training feed forward networks with the Marquardt algorithm*", IEE Trans. Neural network., vol5, pp 989-993.

[۴۲]- 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.

[۴۳]- Caballero J., Fernandez L, Abreu J.I, Fernandez M., (2006), "Amino acid sequence autocorrelation vectors and ensembles of Bayesian-regularized genetic neural networks for prediction of conformational stability of human Lysozyme mutant", J. Chem. Inf. Model., 46, pp 1255-1268.

[۴۴]- چترچی س.، هادی ع.، پرایس ب.، (۱۳۸۴)، "تحلیل رگرسیون با مثال"، نیرومند ح.ع.، چاپ اول، انتشارات دانشگاه فردوسی مشهد

[۴۵]- Hirschen K., Schafer M., (2006), " Bayesian regularization neural networks for optimizing fluid flow processes", Comput. Methods Appl. Mech. Eng, 195, pp 481-500.

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

[۴۷]- Todeschini R., Consonni V., Mauri A., Pavan M., V.13-20124, Milano, Italy, Dragon Software version 3.0.

[۴۸]- SPSS 13.0 for Windows Evaluation version, Release 13.0, SPSS Inc., 2005.

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

[۵۰]- Andrzejewska M., Yépez-Mulia L., Cedillo R., Tapia-Contreras A., Vilpo J., Kazimierczuk Z., (2002), "Synthesis, antiprotozoal and anticancer activity of substituted 2- trifluoromethyl- and 2- pentafluoroethylbenzimidazoles" , Eur. J. Med. Chem., 37, pp 973-978.

[۵۱]- Valdez-Padilla D., Rodriguez-Morales S., Hernandez-Campos A., Hernandez-Luis F., Yépez-Mulia L., Tapia-Contreras A., Castillo R., (2009), "Synthesis and antiprotozoal activity of novel 1- methylbenzimidazole derivatives " , Bioorg. Med. Chem., 17, pp 1724-1730.

[۵۲]- Hernandez-Luis F., Hernandez-Campos A., Castillo R., Navarrete-Vazquez G., Soria-Arteche O., Hernandez-Hernandez M., Yépez-Mulia L., (2010), " Synthesis and biological activity of 2-(trifluoromethyl)-1H-benzimidazole derivatives against some protozoa and *Trichinella spiralis* ", Eur. J. Med. Chem., 45, pp 3135-3141.

[۵۳]- Andrzejewska M., Yépez-Mulia L., Tapia A., Cedillo-Rivera R., E. Laudy A., J. Starosciak B., Kazimierczuk Z., (2004), " Synthesis, and antiprotozoal and antibacterial activity of S-substituted 4,6 dibromo -and 4,6 dichloro-2-mercapto benzimidazoles " , Eur. J. pharm.sci., 21, pp 323–329.

[۵۴]- Zahouily M., Rakik J., Lazar M., Banlaoui M.A., Rayadh A., Komiha N., (2007), "Exploring QSAR of non-nucleoside reverse transcriptase inhibitors by artificial neural networks: HEPT derivatives", ARKIVOC, 15, pp 245-256.

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



[Δϕ]- 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.

[Δγ]- Hemmer M.C., Steinhauer V., Gasteiger J., (1999), "*The prediction of the 3D structure of organic molecules from their infrared spectra*", J. Vib. Spectrosc. ,19, pp 151-164.

[Δλ]- Consonni V., Todeschini R., Pavan M. (2002), "*Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. I. Theory of the Novel 3DMolecular Descriptors*", J. Chem. Inf. Comput. Sci., 42, pp 682-692.

[Δϑ]- Navarrete-Vazquez G., Yopez L., Hernandez-Campos A., Tapia A., Hernandez-Luis F., Cedillo R., Gonzalez J., Martinez-Fernandez A., Martinez-Grueirod M., Castillo R., (2003), "*Synthesis and Antiparasitic Activity of Albendazole and Mebendazole Analogues*", Bioorg. Med. Chem., 11, pp 4615-4622.

[ϕ·]- Karelson M., Lobanov V.S., Katritzky A.R. (1996), "*Quantum-chemical descriptors in QSAR/QSPR studies*", Chem. Rev., 96, pp 1027–1044.