



دانشکده: شیمی

پیش‌بینی فعالیت برخی ترکیبات دارویی ضد اختلال خواب با استفاده
از روش‌های خطی و غیرخطی QSAR

دانشجو: حسن رامه

استاد راهنما:

دکتر ناصرگودرزی

استاد مشاور:

دکتر منصور عرب چم جنگلی

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ماه ۹۱

به پاس تعبیر عظیم و انسانی‌شان از کلمه
ایثار و از خودگذشتگی

به پاس قلب‌های بزرگشان که فریادرس است و سرگردانی
و ترس در پناهِشان به شجاعت می‌گراید. به پاس عاطفه
سرشار و گرمای امیدبخش وجودشان که در این سردترین
روزگاران بهترین پشتیبان است و به پاس محبت‌های بی
دریغشان که هرگز فروکش نمی‌کند.

این مجموعه را به پدر، مادر و همسر عزیزم تقدیم می‌کنم

سپاس خدای را که سخنوران، در ستودن او بمانند
و
شمارندگان، شمردن نعمت‌های او ندانند و
کوشندگان، حق او را گزاردن نتوانند

اکنون که بیاری پروردگار این پایان‌نامه به
اتمام رسیده است برخود واجب می‌دانم، تلاش‌های
خستگی ناپذیر پدر و مادر دلسوزم، همسر فداکارم،
خانواده‌ی گرامی همسر و برادر و خواهران
مهربانم که همواره ره‌گشای مشکلاتم در تمام مراحل
زندگی بوده، ارج نهاده و مراتب قدر دانی و تشکر
قلبی خویش را از آنها ابراز دارم. همچنین لازم
می‌دانم از زحمات فراوان استاد فرهیخته و
توانمند جناب آقای دکتر ناصر گودرزی که
راهنمایی‌ها و نظرات ارزنده ایشان، نقش مهمی در
به ثمر رساندن این پروژه داشتند، صمیمانه تشکر
کنم. از استاد فرزانه و بزرگووارم جناب آقای
دکتر منصور عرب چمنجنگلی که به‌عنوان مشاور در
این پروژه نقش داشتند نیز نهایت تشکر و قدر
دانی را دارم. در نهایت از تمام دوستانم که با
تکیه بر دوستی و محبتشان توانستم این پروژه را
به پایان برسانم، به خصوص آقایان سید
مسعود حسینی، علی جندقی، مهدی آشورپوری، رئوف
غلامی و خانم کریمی تشکر می‌کنم.

چکیده

در قسمت اول این تحقیق، مطالعات ارتباط کمی ساختار- فعالیت (QSAR)، برای مدل‌سازی و پیش‌بینی فعالیت ضد اختلال خواب یکسری از ترکیبات دی‌هیدرو کینولین توسعه یافت. روش‌های رگرسیون مرحله‌ای (SR) و الگوریتم ژنتیک (GA) برای انتخاب توصیف‌کننده‌های مناسب استفاده شدند. توصیف‌کننده‌های انتخاب شده توسط این دو روش برای مدل‌سازی و پیش‌بینی فعالیت ضد اختلال خواب این ترکیبات وارد شبکه عصبی مصنوعی (ANN) شدند. به منظور ارزیابی مدل‌های بدست آمده از روش‌های مختلفی مانند به کارگیری سری تست، رد مرحله‌ای تک‌تک داده‌ها و Y-تصادفی استفاده گردید. نتایج به دست آمده بیانگر توانایی پیش‌بینی مناسب هر دو روش SR-ANN و GA-ANN می‌باشد. ضرایب تعیین سری تست با روش‌های SR-ANN و GA-ANN به ترتیب ۰/۹۰۴۲ و ۰/۸۴۰۲ بدست آمد.

در قسمت دوم این تحقیق، مدل‌سازی QSAR برای ۴۲ ترکیب از مشتقات پیریدوپیریدازین به عنوان بازدارنده‌های P38 MAP کیناز پیشنهاد شد. توصیف‌کننده‌های وارد شده بمنظور مدل‌سازی شبکه‌ی عصبی مصنوعی (ANN)، توسط دو روش رگرسیون مرحله‌ای و الگوریتم ژنتیک انتخاب شدند. ضرایب تعیین سری تست با روش‌های SR-ANN و GA-ANN به ترتیب ۰/۸۶۰۷ و ۰/۸۰۲۷ بدست آمد.

کلمات کلیدی: فعالیت ضد اختلال خواب، بازدارنده‌های P38 MAP کیناز، مطالعات ارتباط کمی ساختار- فعالیت، SR-ANN، GA-ANN.

نتایج حاصل از این پایان‌نامه در دو پوستر تحت عناوین:

“Application of linear and nonlinear QSAR methods for prediction of antitrypanosomal drugs activities of some 1, 2-dihydroquinolin-6-ols and their ester derivatives ”

“QSAR study of p38 MAP kinase inhibitors activity of some pyridopyridazin-6-one derivatives using SR-ANN and GA-ANN”

در پانزدهمین سمینار شیمی فیزیک ایران در دانشگاه تهران پذیرفته شد.

فصل اول: مقدمه

- ۱-۱-۱- اختلال خواب----- ۲
- ۱-۱-۱- مکانیسم آلودگی----- ۲
- ۲-۱- بازدارنده‌های p38 MAP کیناز----- ۳
- ۳-۱- ضرورت تحقیق----- ۳
- ۴-۱- پیشینه‌ی تحقیق----- ۴
- ۱-۴-۱- ضد اختلال خواب----- ۴
- ۲-۴-۱- بازدارنده‌های p38 MAP کیناز----- ۵

فصل دوم: کمومتریکس

- ۱-۲- کمومتریکس----- ۸
- ۲-۲- ارتباط کمی ساختار- فعالیت----- ۸
- ۱-۲-۲- فراهم کردن سری داده‌ها----- ۹
- ۲-۲-۲- بهینه‌سازی ساختار مولکول‌ها----- ۹
- ۳-۲-۲- محاسبه توصیف‌کننده‌ها----- ۱۰
- ۴-۲-۲- انتخاب بهترین توصیف‌کننده‌ها----- ۱۰
- ۵-۲-۲- رگرسیون خطی چندگانه----- ۱۰
- ۶-۲-۲- رگرسیون خطی ساده----- ۱۱
- ۷-۲-۲- رگرسیون مرحله‌ای----- ۱۱
- ۸-۲-۲- مقدمه ای بر الگوریتم ژنتیک----- ۱۲
- ۱-۸-۲-۲- تاریخچه----- ۱۲
- ۲-۸-۲-۲- تعریف الگوریتم ژنتیک----- ۱۳
- ۳-۸-۲-۲- مراحل کلی الگوریتم ژنتیک----- ۱۳
- ۴-۸-۲-۲- ساختار الگوریتم‌های ژنتیکی----- ۱۴
- ۱-۴-۸-۲-۲- کروموزوم----- ۱۴
- ۲-۴-۸-۲-۲- جمعیت----- ۱۵
- ۳-۴-۸-۲-۲- تابع برازندگی----- ۱۵
- ۵-۸-۲-۲- کدگذاری----- ۱۵
- ۱-۵-۸-۲-۲- روش‌های کدگذاری----- ۱۶
- ۶-۸-۲-۲- عملگرهای الگوریتم ژنتیک----- ۱۸
- ۱-۶-۸-۲-۲- عملگر انتخاب----- ۱۸
- ۲-۶-۸-۲-۲- عملگر پیوند----- ۲۱
- ۳-۶-۷-۲-۲- عملگر جهش----- ۲۲
- ۷-۸-۲-۲- نقاط قوت الگوریتم‌های ژنتیک----- ۲۳
- ۸-۸-۲-۲- محدودیت‌های الگوریتم ژنتیک----- ۲۴
- ۹-۸-۲-۲- الگوریتم ژنتیک و QSAR----- ۲۴
- ۹-۲-۲- ساختن مدل----- ۲۴

۲۵	۱۰-۲-۲-۱۰- مقدمه‌های بر شبکه عصبی
۲۵	۱-۱۰-۲-۲-۱- ساختمان و عملکرد نرون طبیعی
۲۶	۲-۱۰-۲-۲-۲- ساختمان و عملکرد نرون مصنوعی
۲۷	۳-۱۰-۲-۲-۳- بهبود تعمیم
۲۹	۴-۱۰-۲-۲-۴- توابع انتقال
۳۰	۵-۱۰-۲-۲-۵- ساختارهای شبکه
۳۲	۱۱-۲-۲-۱۱- ارزیابی مدل
۳۵	۳-۲- نرم‌افزارهای مورد استفاده
۳۵	۱-۳-۲- بسته نرم‌افزاری Hyperchem
۳۵	۲-۳-۲- بسته نرم‌افزاری Dragon
۳۵	۳-۳-۲- بسته نرم‌افزاری SPSS
۳۶	۴-۳-۲- نرم‌افزار MATLAB

فصل سوم: مطالعه ارتباط کمی ساختار- فعالیت برخی مشتقات دی هیدرو کینولین به عنوان داروی ضد اختلال خواب با استفاده از SR-ANN و GA-ANN

۳۸	۱-۳- مدل سازی فعالیت دارویی مشتقات دی هیدروکینولین به عنوان بازدارنده‌های بیماری اختلال خواب
۳۸	۱-۱-۳- انتخاب سری داده‌های استفاده شده در مدل سازی
۴۳	۲-۱-۳- رسم و بهینه‌سازی ساختار مولکول‌ها
۴۴	۳-۱-۳- محاسبه توصیف‌کننده‌ها
۴۴	۴-۱-۳- انتخاب توصیف‌کننده‌های مهم
۴۵	۱-۴-۱-۳- انتخاب توصیف‌کننده‌های مناسب توسط رگرسیون مرحله‌ای در SPSS
۴۶	۲-۴-۱-۳- انتخاب توصیف‌کننده‌های مناسب توسط الگوریتم ژنتیک در MATLAB
۵۰	۵-۱-۳- مدل سازی شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای - ۵۰
۵۰	۱-۵-۱-۳- بهینه‌سازی پارامترهای شبکه
۵۱	۲-۵-۱-۳- انتخاب تعداد متغیرهای ورودی شبکه، تابع آموزش، نوع تابع انتقال، تعدادگره‌های لایه‌های پنهان و تعداد دوره‌های آموزشی
۵۷	۳-۵-۱-۳- انتخاب مقدار بهینه μ
۵۸	۴-۵-۱-۳- ساختار شبکه عصبی مصنوعی بهینه شده
۵۹	۶-۱-۳- مدل سازی شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک
۵۹	۱-۶-۱-۳- بهینه‌سازی پارامترهای شبکه
۵۹	۲-۶-۱-۳- بهینه سازی همزمان تعداد متغیرهای ورودی شبکه، تابع آموزش، نوع تابع انتقال، تعدادگره‌های لایه های پنهان و تعداد دوره‌های آموزشی
۶۵	۳-۶-۱-۳- بهینه سازی پارامتر μ
۶۶	۴-۶-۱-۳- ساختار شبکه عصبی مصنوعی بهینه شده
۶۶	۵-۶-۱-۳- معماری شبکه عصبی بهینه شده
۶۷	۷-۱-۳- ارزیابی مدل های SR-ANN و GA-ANN
۶۷	۱-۷-۱-۳- ارزیابی مدل های SR-ANN و GA-ANN با استفاده از سری های ارزیابی و تست
۷۰	۲-۷-۱-۳- ارزیابی مدل های SR-ANN و GA-ANN توسط روش رد مرحله ای تک تک

- ۳-۱-۸- ارزیابی مدل‌های برتر SR-ANN و GA-ANN با استفاده از پارامترهای آماری ----- ۷۴
- ۳-۱-۹- ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون Y- تصادفی ----- ۷۴
- ۳-۱-۱۰- بررسی ارتباط توصیف‌کننده‌های منتخب با فعالیت ضد اختلال خواب ----- ۷۵
- ۳-۱-۱۱- بررسی میزان مشارکت توصیف‌کننده‌های منتخب در شبکه عصبی ----- ۸۰
- ۳-۱-۱۲- نتیجه‌گیری نهایی ----- ۸۱

فصل چهارم: مطالعه ارتباط کمی ساختار- فعالیت برخی مشتقات پیریدوپیریدازین به عنوان بازدارنده‌های

MAP P38 کیناز با استفاده از SR-ANN و GA-ANN

- ۴-۱-۱- مدل‌سازی فعالیت دارویی مشتقات پیریدوپیریدازین به عنوان بازدارنده‌های MAP P38 کیناز ----- ۸۳
- ۴-۱-۱- انتخاب سری داده‌ها ----- ۸۳
- ۴-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها ----- ۸۷
- ۴-۱-۳- محاسبه و انتخاب توصیف‌کننده‌های مهم ----- ۸۷
- ۴-۱-۳-۱- انتخاب توصیف‌کننده‌های مناسب توسط رگرسیون مرحله‌ای در SPSS ----- ۸۷
- ۴-۱-۳-۲- انتخاب توصیف‌کننده‌های مناسب توسط الگوریتم ژنتیک در MATLAB ----- ۹۰
- ۴-۱-۴- مدل‌سازی شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای ----- ۹۲
- ۴-۱-۴-۱- بهینه‌سازی پارامترهای شبکه ----- ۹۳
- ۴-۱-۴-۲- بهینه‌سازی همزمان تعداد متغیرهای ورودی شبکه، تابع آموزش، نوع تابع انتقال، تعداد گره‌های لایه‌های پنهان و تعداد دوره‌های آموزشی ----- ۹۳
- ۴-۱-۴-۳- بهینه‌سازی پارامتر μ ----- ۹۹
- ۴-۱-۴-۴- ساختار شبکه عصبی مصنوعی بهینه شده ----- ۱۰۰
- ۴-۱-۵- مدل‌سازی شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک - ۱۰۱ ----- ۱۰۱
- ۴-۱-۵-۱- بهینه‌سازی پارامترهای شبکه ----- ۱۰۱
- ۴-۱-۵-۲- بهینه‌سازی همزمان تعداد متغیرهای ورودی شبکه، تابع آموزش، نوع تابع انتقال، تعداد گره‌های لایه‌های پنهان و تعداد دوره‌های آموزشی ----- ۱۰۱
- ۴-۱-۵-۳- بهینه‌سازی پارامتر μ ----- ۱۰۷
- ۴-۱-۵-۴- ساختار شبکه عصبی مصنوعی بهینه شده ----- ۱۰۸
- ۴-۱-۵-۵- معماری شبکه عصبی بهینه شده ----- ۱۰۸
- ۴-۱-۶- ارزیابی مدل‌های SR-ANN و GA-ANN ----- ۱۰۹
- ۴-۱-۶-۱- ارزیابی مدل‌های SR-ANN و GA-ANN با استفاده از سری‌های ارزیابی و تست ----- ۱۰۹
- ۴-۱-۶-۲- ارزیابی مدل‌های SR-ANN و GA-ANN توسط روش رد مرحله‌ای تک‌تک ----- ۱۱۲
- ۴-۱-۷- ارزیابی مدل‌های برتر SR-ANN و GA-ANN با استفاده از پارامترهای آماری ----- ۱۱۶
- ۴-۱-۸- ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون Y- تصادفی ----- ۱۱۶
- ۴-۱-۹- بررسی ارتباط توصیف‌کننده‌های منتخب ----- ۱۱۷
- ۴-۱-۱۰- بررسی میزان مشارکت توصیف‌کننده‌های منتخب در شبکه عصبی ----- ۱۲۰
- ۴-۱-۱۱- نتیجه‌گیری نهایی ----- ۱۲۱
- آینده نگری ----- ۱۲۱
- فهرست منابع ----- ۱۲۲

فهرست اشکال

- شکل (۱-۲) - کدینگ مقدار ----- ۱۷
- شکل (۲-۲) - کدینگ درختی ----- ۱۸
- شکل (۳-۲) - چرخ رولت ----- ۲۰
- شکل (۴-۲) - شبیه سازی جهش به کمک نمودار ----- ۲۲
- شکل (۵-۲) - جهش باینری ----- ۲۳
- شکل (۶-۲) - ساختار یک نرون زیستی ----- ۲۶
- شکل (۷-۲) - ساختمان یک نرون محاسباتی ----- ۲۶
- شکل (۸-۲) - تغییرات خطای سری آموزش و سری ارزیابی ----- ۲۷
- شکل (۹-۲) - توابع انتقال (الف) خطی، (ب) سیگموئیدی ----- ۳۰
- شکل (۱-۳) - اسکلت اصلی ترکیبات مورد مطالعه ----- ۳۸
- شکل (۲-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید ----- ۵۲
- شکل (۳-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش بایزین و تابع انتقال لگاریتم سیگموئید ----- ۵۳
- شکل (۴-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید ----- ۵۴
- شکل (۵-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تابع تانژانت سیگموئید ----- ۵۵
- شکل (۶-۳) - نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ ----- ۵۷
- شکل (۷-۳) - ساختار شبکه عصبی بهینه شده ----- ۵۸
- شکل (۸-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید ----- ۶۰
- شکل (۹-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش بایزین و تابع انتقال لگاریتم سیگموئید ----- ۶۱
- شکل (۱۰-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید ----- ۶۲
- شکل (۱۱-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تابع تانژانت سیگموئید ----- ۶۳
- شکل (۱۲-۳) - نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ ----- ۶۵
- شکل (۱۳-۳) - ساختار شبکه عصبی به دست آمده پس از بهینه سازی ----- ۶۶
- شکل (۱۴-۳) - نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری ارزیابی الف: با مدل SR-ANN ب: بامدل GA-ANN ----- ۶۸
- شکل (۱۵-۳) - نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری تست الف: با مدل SR-ANN ب: بامدل GA-ANN ----- ۶۹

- شکل (۳-۱۶) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها الف: با مدل SR-ANN ب: با مدل GA-ANN ----- ۷۳
- شکل (۳-۱۷) - نمودار باقیمانده‌ها بر حسب مقدار تجربی pIC_{50} برای کل داده‌ها الف: با مدل SR-ANN ب: با مدل GA-ANN ----- ۷۳
- شکل (۳-۱۸) - مشارکت توصیف‌کننده‌ها در شبکه عصبی بهینه ----- ۸۰
- شکل (۴-۱) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئید ----- ۹۴
- شکل (۴-۲) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش بایزین و تابع انتقال لگاریتم سیگموئید ----- ۹۵
- شکل (۴-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ- مارکوارت و تابع انتقال تانژانت سیگموئید ----- ۹۶
- شکل (۴-۴) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تابع تانژانت سیگموئید ----- ۹۷
- شکل (۴-۵) - نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ ----- ۹۹
- شکل (۴-۶) - ساختار شبکه عصبی بهینه شده ----- ۱۰۰
- شکل (۴-۷) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئید ----- ۱۰۲
- شکل (۴-۸) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش بایزین و تابع انتقال لگاریتم سیگموئید ----- ۱۰۳
- شکل (۴-۹) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ- مارکوارت و تابع انتقال تانژانت سیگموئید ----- ۱۰۴
- شکل (۴-۱۰) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تابع تانژانت سیگموئید ----- ۱۰۵
- شکل (۴-۱۱) - نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ ----- ۱۰۷
- شکل (۴-۱۲) - ساختار شبکه عصبی بهینه شده ----- ۱۰۸
- شکل (۴-۱۳) - نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری ارزیابی الف: با مدل SR-ANN ب: با مدل GA-ANN ----- ۱۱۰
- شکل (۴-۱۴) - نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای تست الف: با مدل SR-ANN ب: با مدل GA-ANN ----- ۱۱۱
- شکل (۴-۱۵) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها الف: با مدل SR-ANN ب: با مدل GA-ANN ----- ۱۱۵
- شکل (۴-۱۶) - نمودار باقیمانده‌ها بر حسب مقدار تجربی pIC_{50} برای کل داده‌ها الف: با مدل SR-ANN ب: با مدل GA-ANN ----- ۱۱۶
- شکل (۴-۱۷) - مشارکت توصیف‌کننده‌ها در شبکه عصبی بهینه ----- ۱۲۰

فهرست جداول

- جدول (۱-۳) - ساختار و مقادیر pIC_{50} ترکیبات مورد مطالعه ----- ۳۹
- جدول (۲-۳) - توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای ----- ۴۵
- جدول (۳-۳) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای ----- ۴۵
- جدول (۴-۳) - توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک ----- ۴۸
- جدول (۵-۳) - توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک ----- ۴۹
- جدول (۶-۳) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک ----- ۴۹
- جدول (۷-۳) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا ----- ۵۶
- جدول (۸-۳) - میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ ----- ۵۷
- جدول (۹-۳) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی ----- ۵۸
- جدول (۱۰-۳) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا ----- ۶۴
- جدول (۱۱-۳) - میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ ----- ۶۵
- جدول (۱۲-۳) - مشخصات شبکه بهینه ----- ۶۶
- جدول (۱۳-۳) - نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری ارزیابی ----- ۶۸
- جدول (۱۴-۳) - نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری تست ----- ۶۹
- جدول (۱۵-۳) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-ANN و GA-ANN با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها ----- ۷۱
- جدول (۱۶-۳) - پارامترهای آماری مدل‌های SR-ANN و GA-ANN ----- ۷۴
- جدول (۱۷-۳) - مقادیر R^2 برای سری تست و ارزیابی مدل SR-ANN با استفاده از آزمون Y-تصادفی ----- ۷۵
- جدول (۱۸-۳) - مقادیر R^2 برای سری تست و ارزیابی مدل GA-ANN با استفاده از آزمون Y-تصادفی ----- ۷۵
- جدول (۱۹-۳) - اثر متوسط توصیف‌کننده‌های بدست آمده توسط روش رگرسیون مرحله‌ای ----- ۷۶
- جدول (۲۰-۳) - مثال‌هایی از اثر توصیف‌کننده nROR بر فعالیت دارویی ----- ۷۸
- جدول (۱-۴) - ساختار و مقادیر pIC_{50} ترکیبات مورد مطالعه ----- ۸۴
- جدول (۲-۴) - توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای ----- ۸۸
- جدول (۳-۴) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای ----- ۸۹
- جدول (۴-۴) - توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک ----- ۹۱
- جدول (۵-۴) - توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک ----- ۹۲
- جدول (۶-۴) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک ----- ۹۲
- جدول (۷-۴) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا ----- ۹۸
- جدول (۸-۴) - میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ ----- ۹۹
- جدول (۹-۴) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی ----- ۱۰۰
- جدول (۱۰-۴) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا ----- ۱۰۶
- جدول (۱۱-۴) - میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ ----- ۱۰۷
- جدول (۱۲-۴) - مشخصات شبکه بهینه ----- ۱۰۸
- جدول (۱۳-۴) - نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری ارزیابی ----- ۱۱۰
- جدول (۱۴-۴) - نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری تست ----- ۱۱۱
- جدول (۱۵-۴) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-ANN و GA-ANN با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها ----- ۱۱۳

- جدول (۴-۱۶) - پارامترهای آماری برای مدل‌های SR-ANN و GA-ANN ----- ۱۱۶
- جدول (۴-۱۷) - مقادیر R^2 برای سری تست و ارزیابی مدل SR-ANN با استفاده از آزمون Y- تصادفی ----- ۱۱۷
- جدول (۴-۱۸) - مقادیر R^2 برای سری تست و ارزیابی مدل GA-ANN با استفاده از آزمون Y- تصادفی ----- ۱۱۷
- جدول (۴-۱۹) - اثر متوسط توصیف‌کننده‌های بدست آمده توسط روش رگرسیون مرحله‌ای ----- ۱۱۷

فصل اول

مقدمه

۱-۱- اختلال خواب^۱

در دوران پیش از تاریخ
ته شد. در سال ۱۹۰۲ دولت
انگلستان یک گروه پژوهشی را برای بررسی بیماری اختلال خواب به آفریقا فرستاد. یکی از اعضای این
گروه به نام آلدوکاستلا^۲ پس از کالبد شکافی قربانیان دریافت که وجود انواعی از انگل‌های ناشناخته

^۱ - Sleep disorders

^۲ - Aldo Kastela

عامل این بیماری می‌باشد. یک سال بعد پژوهشگری به نام دیوید بروک^۱ به این گروه پیوست. بروک پیش از این دریافته بود که مگس‌های تسه‌تسه پس از آلوده شدن توسط انگل تریپانوزوم^۲ مبتلا به بیماری تاگانا^۳ می‌شوند. این دو پژوهشگر دریافتند که بیماری اختلال خواب در مناطقی شیوع پیدا می‌کند که مگس‌های تسه‌تسه زندگی می‌کنند. در نهایت آن‌ها به این نتیجه رسیدند که کسانی به بیماری اختلال خواب دچار می‌گردند که نیش مگس تسه‌تسه آلوده به انگل تریپانوزوم وارد بدن آن‌ها شده باشد [۱]. تریپانوزوم‌ها در ابتدا باعث ایجاد یک بیماری همراه با تب و ماه‌ها یا سال‌ها بعد باعث اختلال عصبی و مرگ می‌گردند [۲].

۱-۱-۱- مکانیسم آلودگی

مکانیسم آلودگی بیماری اختلال خواب به این صورت است که در اثر نیش زدن مگس تسه‌تسه آلوده به تریپانوزوم، انگل تریپانوزوم وارد خون فرد می‌شود و با تکثیر آن در خون، اختلالهایی مانند تب، ورم، خستگی، چاق شدن و لاغر شدن، تورم غدد لنفاوی، کم‌خونی، بزرگی کبد و طحال، بزرگی ماهیچه قلب و حمله به بافت‌ها را ایجاد می‌کند و در نهایت موجب بیماری اختلال خواب می‌شود. دوره نهفتگی این انگل در بدن سه ماه و در بدن میزبان واسط حدود یک ماه طول می‌کشد [۳]. امروزه حدود ۶۶ میلیون نفر در سراسر دنیا در خطر ابتلا به بیماری اختلال خواب هستند و این آمار همچنان در حال گسترش است [۴]. بنابراین نیاز به درمان توسط مواد دارویی ارزان، موثر و سالم امری ضروری است.

۱-۲- بازدارنده‌های MAP^۴ p38 کیناز

^۱ - David Brooke

^۲ - Trypanosome

^۳ - Tagana

^۴ - Mitogen Activated Protein

p38 MAP کینازها برای اولین بار در سال ۱۹۹۴ مورد مطالعه قرار گرفتند. این کینازها نوعی پروتئین فعال شده با میتوزن می‌باشند که نقش مهمی در واکنش سلول‌ها با تنش‌های خارجی مانند سیتوکین‌ها دارند و در نهایت باعث مرگ سلول‌ها می‌شوند [۵].

درمان با بازدارنده‌های p38 MAP کیناز به این صورت است که، این بازدارنده‌ها مانع از فعالیت آنزیم‌های (کینازهای) p38 MAP در بدن، در نهایت باعث کاهش شدت بیماری می‌شوند. این بازدارنده‌ها از نظر توزیع بافت، تنظیم فعالیت کیناز و حساسیت متفاوت می‌باشند، همچنین عواملی مانند نقش p38 MAP کینازها در مرگ سلولی و ایجاد مکانیسم‌هایی که با آن بازدارنده‌های p38 MAP کیناز جلوی فعالیت آنزیم را بگیرند، فعالیت بیولوژیکی این بازدارنده‌ها را به طور چشم‌گیری افزایش داده است. این بازدارنده‌ها برای درمان بیماری‌های گوناگونی مانند استرس، آرتروز و بیماری‌های قلبی عروقی کارآمد می‌باشند. در حال حاضر، درمان بیماری توسط بازدارنده‌های p38 MAP کیناز به طور گسترده در مطالعات بالینی مورد مطالعه قرار گرفته است [۶].

۱-۳- ضرورت تحقیق

پیشرفت علوم در زمینه‌های مختلف روش‌های آزمایشگاهی کمک شایانی به بهبود نتایج حاصل از این روش‌ها کرده است. با وجود این، مشکلات و محدودیت‌هایی از جمله وجود عدم قطعیت در نتایج آزمایشگاهی، هزینه‌ی بالا، در دسترس نبودن امکانات و عدم توانایی در اندازه‌گیری همه‌ی ترکیبات باعث کشف و توسعه‌ی روش‌های نظری شده است [۷].

۱-۴- پیشینه‌ی تحقیق

۱-۴-۱- مطالعه بر روی ترکیبات ضد اختلال خواب

کرسٹو^۱ و همکارانش در سال ۲۰۰۵، میزان بازدارندگی ۲۶ ترکیب از مشتقات ۵- بنزوفوراکسان را در برابر بیماری اختلال خواب، با بکارگیری روش‌های QSAR^۲ دو بعدی و سه بعدی مقایسه کردند. در این مطالعه از رگرسیون خطی چندگانه^۳ استفاده شد و مقادیر مجذور ضریب همبستگی^۴ برای این دو روش را به ترتیب ۰/۹۳۹ و ۰/۹۹۷ بدست آوردند و نشان دادند که روش QSAR سه بعدی قدرت پیش‌بینی بهتری برای این دسته از ترکیبات دارد [۸].

اشمیت^۵ و همکارانش در سال ۲۰۰۹ ارتباط کمی ساختار- فعالیت ۴۰ ترکیب از مشتقات لاکتون‌ها را که به عنوان ترکیبات ضد اختلال خواب عمل می‌کنند، با استفاده از روش حداقل مربعات جزئی (PLS)^۶ بررسی کردند و مقادیر مجذور ضریب همبستگی و رد مرحله‌ای تک‌تک^۷ به ترتیب ۰/۸۹ و ۰/۸۵ بدست آمد [۹].

در سال ۲۰۱۰ نولوی^۸ و همکارانش مدل‌های QSAR سه بعدی را برای مجموعه‌ای شامل ۲۹ ترکیب از مشتقات ۱ و ۳ و ۴- تیادیازول ۲- آریل هالیدازون را به عنوان بازدارنده‌های اختلال خواب توسط روش تجزیه‌ی میدان مولکولی (MFA)^۹ مطالعه کردند به طوری که ضریب همبستگی برای سری آموزش و سری تست به ترتیب ۰/۹۴۵۵ و ۰/۸۰۸۷ بدست آمد [۱۰].

پاسکوالتو^{۱۰} و همکارانش در سال ۲۰۱۲ ارتباط کمی ساختار- فعالیت مشتقات فنوتیازین که بازدارنده آنزیم T.cruzi در انگل تریپانوزوم می‌باشند را با استفاده از روش‌های انتخاب پیش‌بینی کننده‌ی مرحله‌ای (OPS)^{۱۱} و حداقل مربعات جزئی (PLS) مقایسه کردند و از مقادیر مجذور ضریب

1 - Cerecetto

2 - Quantitative Structure-Activity Relationship

3 - Multiple linear regression

4 - Squared correlation coefficient

5 - Schmidt

6 - Partial Least Square

7 - Leave one out

8 - Noolvi

9 - Molecular Field Analysis

10 - Pasqualoto

11 - Ordered Predictor Selection

همبستگی و آزمون Y- تصادفی^۱ دریافتند که قابلیت پیش‌بینی مدل ارائه شده توسط PLS بهتر می‌باشد [۱۱].

در سال ۲۰۱۲ لزانو^۲ و همکارانش با مدل‌سازی QSAR مربوط به ۵۳ ترکیب از مشتقات تیوسمی کربازونها برای پیش‌بینی بازدارندگی اختلال خواب با بکارگیری روش حداقل مربعات جزئی (PLS)، مقدار مجذور ضریب همبستگی ۰/۸۵ را بدست آوردند [۱۲].

۱-۴-۲- مطالعه بر روی بازدارنده‌های MAP p38 کیناز

در سال ۲۰۰۷ تحقیقات QSAR سه‌بعدی توسط نایانا^۳ برای پیش‌بینی فعالیت بازدارندگی P38 MAP کیناز ۲۸ ترکیب از مشتقات بنزیمیدازولون صورت گرفت. در این تحقیق از روش تجزیه‌ی میدان مولکولی (MFA) استفاده شد که در نهایت مقادیر R^2 و Q^2 برای سری تست به ترتیب ۰/۹۸ و ۰/۶۵ بدست آمد [۱۳].

سارما^۴ و همکارانش در سال ۲۰۰۸ مدل‌های QSAR سه‌بعدی شامل ۴۱ ترکیب از مشتقات اوره را با اثر بازدارندگی P38 MAP کیناز توسط روش تجزیه‌ی میدان مولکولی (MFA) مطالعه کردند. آن‌ها با تقسیم تصادفی ترکیبات به دو سری آموزش (شامل ۳۲ ترکیب) و تست (شامل ۹ ترکیب) به بهینه‌سازی پارامترها پرداختند. مقادیر R^2 حاصل از رد مرحله‌ای تک‌تک (LOO)^۵ و سری تست حاصل از مدل بدست آمده به ترتیب ۰/۷۹۹ و ۰/۷۴ بدست آمد [۱۴].

آچایاه^۶ و همکارانش در سال ۲۰۰۸ در یک تحقیق QSAR سه‌بعدی برای پیش‌بینی فعالیت بازدارندگی P38 MAP کیناز شامل ۴۴ ترکیب فنوکسی پیریمیدینیل را مورد بررسی قرار دادند. در

^۱ - Y- randomization test

^۲ - Lozano

^۳ - Nayana

^۴ - Sarma

^۵ - Leave one out

^۶ - Achaiyah

این پژوهش از روش‌های COMFA^۱ و COMSIA^۲ استفاده شد و مقادیر R² برای این دو روش به ترتیب ۰/۹۰۸ و ۰/۸۹۴ بدست آمد [۱۵].

در سال ۲۰۱۰ کولکارنی^۳ و همکارانش مدل‌های QSAR شامل ۴۷ ترکیب از مشتقات ۵۴-ایمیدازو بی‌پیریدین و بنزیمیدازول را برای پیش‌بینی فعالیت بازدارندگی P38 MAP کیناز توسط روش حداقل مربعات جزئی (PLS) مطالعه کردند به طوری که مقادیر R² و Q² برای سری تست به ترتیب ۰/۸۱ و ۰/۷۵ بدست آمد [۱۶].

در تحقیق حاضر، ارتباط کمی ساختار- فعالیت برای پیش‌بینی فعالیت دارویی یکسری از ترکیبات ضد اختلال خواب و همچنین بازدارنده‌های p38 MAP کیناز به کار گرفته شد. بر اساس مطالعات صورت گرفته، برای ترکیبات بررسی شده در این پایان‌نامه قبلاً مدل‌سازی صورت نگرفته است و از روش‌های SR-ANN^۴ و GA-ANN^۵ برای مدل‌سازی استفاده شده است.

فصل دوم

¹ - Comparative Molecular Field Analysis

² - Comparative Molecular Similarity Indices Analysis

³ - Kulkarni

⁴ - Stepwise Regression- Artificial Neural Network

⁵ - Genetic Algorithm- Artificial Neural Network

کموتریکس

۱-۲- کموتریکس

کموتریکس^۱ شاخه‌ای از علم شیمی است که ابزارهای ریاضی و آمار را برای سامان بخشیدن به داده‌های شیمیایی و به‌دست آوردن اطلاعات بیشتر از آن‌ها به کار می‌گیرد. به دلیل پیشرفت سریع دستگاه‌های هوشمند آزمایشگاهی و نیز امکان استفاده از کامپیوترهای قدرتمند و نرم‌افزارهای ساده، کموتریکس به عنوان یک دانش مستقل، گسترش روزافزونی داشته است [۱۷].

^۱ - Chemometries

از نظر شیمیدانان فعالیت و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده مولکول انجام گیرد، به آن ارتباط کمی ساختار-ویژگی (QSPR)^۱ می‌گویند. اما وقتی خواصی از نوع بیولوژیکی (مانند فعالیت دارویی) توصیف می‌شود، اشاره به ارتباط کمی ساختار-فعالیت (QSAR) خواهد داشت [۱۸]. در مطالعات QSAR سعی بر این است تا رابطه‌ای هماهنگ میان فعالیت یک ترکیب با ویژگی‌های مولکولی آن پیدا شود، به گونه‌ای که بتوان از این روابط برای ارزیابی فعالیت ترکیبات جدید استفاده کرد [۱۹].

بسیاری از شیمیدان‌ها از آقای اسوانت ولد^۲ به عنوان اولین کسی که این روش‌ها را معرفی کرده است نام می‌برند و به او لقب پدر علم کمومتریکس را داده‌اند [۲۰].

۲-۲- ارتباط کمی ساختار-فعالیت

یکی از کاربردهای ویژه روش‌های کمومتریکس توسعه ارتباط کمی ساختار-فعالیت و ارزیابی داده‌های تجزیه‌ای است [۲۱]. مولکول‌های مشابه با تغییر کوچکی در ساختارشان می‌توانند فعالیت بیولوژیکی کاملاً متفاوتی داشته باشند و در واقع QSAR نیز سعی در پیدا کردن رابطه‌ی هماهنگ میان فعالیت بیولوژیکی و ویژگی‌های مولکولی، به منظور کاربرد این قواعد برای ارزیابی فعالیت ترکیبات جدید دارد. QSAR برای اولین بار در قرن نوزدهم مورد استفاده قرار گرفت [۲۱ و ۲۲]. مراحل انجام یک روش QSAR نیز به ترتیب شامل: فراهم کردن سری داده‌ها، بهینه‌سازی ساختار مولکول‌ها، محاسبه توصیف‌کننده‌ها، انتخاب بهترین توصیف‌کننده‌ها، ساختن مدل و ارزیابی مدل می‌باشد.

۲-۲-۱- فراهم کردن سری داده‌ها

^۱ - Quantitative Structure-Property Relationship

^۲ - Svante Wold

در توسعه‌ی مطالعات QSAR جمع‌آوری مجموعه‌ای از مولکول‌ها که مقادیر تجربی فعالیت مورد بررسی برای آن‌ها در شرایط یکسانی به دست آمده باشند در اولویت قرار دارد. عموماً جمع‌آوری مولکول‌های سری داده‌ها از طریق جستجو در مقالات و مجلات علمی صورت می‌گیرد و هر قدر سری داده‌ها بزرگتر و متنوع‌تر باشد، مدل حاصل از آن نیز معتبرتر خواهد بود.

۲-۲-۲- بهینه‌سازی ساختار مولکول‌ها

بهینه کردن ساختار هندسی مولکول‌ها، مرحله‌ی بعدی در مطالعات QSAR می‌باشد. بدین منظور از روش‌های شیمی محاسباتی استفاده می‌شود. ایجاد توصیف‌کننده‌های^۱ هندسی و هیبریدی بر مبنای ساختار و هندسه دقیق مولکولی استوار است و اگر ساختارها به شکل صورتبندی^۲ با حداقل انرژی نباشند، مقادیر غیر صحیحی برای این توصیف‌کننده‌ها ایجاد می‌شود [۲۳]. بنابراین باید ساختار هندسی ترکیب بهینه^۳ شود و مختصاتی از ساختار ترکیب به دست آید که در آن مختصات کارترین، انرژی مولکول حداقل باشد. بدین منظور از نرم‌افزار Hyperchem جهت رسم و بهینه‌سازی ساختار مولکول‌ها استفاده می‌شود [۲۴].

۲-۲-۳- محاسبه توصیف‌کننده‌ها

توصیف‌کننده‌ها مقادیر عددی هستند که ویژگی‌های مختلف ساختاری و الکترونی مولکول‌ها را به طور کمی نشان می‌دهند. به عبارت دیگر هر توصیف‌کننده بیانگر خصوصیت ویژه‌ای از مولکول است که ممکن است بر فعالیت مورد نظر مؤثر باشد. به عبارت دیگر می‌توان گفت که هر توصیف‌کننده، اطلاعات خاصی از مولکول را که بر کمیت مورد مدل‌سازی اثر می‌گذارد در اختیار قرار می‌دهد و از اهمیت بسزایی برخوردار است.

¹ - Descriptors

² - Conformation

³ - Optimize

۲-۲-۴- انتخاب بهترین توصیف‌کننده‌ها

قبل از مرحله انتخاب متغیر، فرآیند کاهش متغیر^۱ صورت می‌گیرد. زیاد بودن تعداد توصیف‌کننده‌ها و همچنین همبستگی بالای برخی از آن‌ها به یکدیگر، باعث طولانی و دشوار شدن فرآیند مدلسازی می‌شود. بنابراین با انتخاب زیرمجموعه‌ای از متغیرهای مستقل^۲ که اطلاعات اساسی مربوط به کل سری داده‌ها را دربردارند، متغیرهای با همبستگی بالا حذف می‌شوند. یکی از مهمترین مراحل QSAR انتخاب توصیف‌کننده‌های مناسب برای ایجاد مدل می‌باشد. از روش‌های انتخاب توصیف‌کننده‌ها می‌توان به رگرسیون مرحله‌ای، الگوریتم ژنتیک^۳ و روش جایگزینی^۴ اشاره کرد.

۲-۲-۵- رگرسیون خطی چندگانه

رگرسیون به مفهوم ایجاد یک رابطه بین یک متغیر از یک طرف و یک یا چند متغیر از طرف دیگر می‌باشد. در بسیاری از موارد نمی‌توان تغییرات یک متغیر را فقط به مقادیر یک متغیر دیگر مربوط کرد و یا به عبارت دیگر برای پیش‌بینی مقادیر یک متغیر، دانستن مقادیر دو یا چند متغیر دیگر لازم است که در این حالت از روش‌های رگرسیون چندگانه استفاده می‌شود [۲۵].

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

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_p X_p \quad (1-2)$$

که در آن Y متغیر وابسته (فعالیت دارویی)، X_1 و X_2 ، ...، X_p متغیرهای مستقل (توصیف‌کننده‌ها) و β_i ضرایب رگرسیون نامیده می‌شوند. باید توجه کرد که مقدار ضریب متغیر مستقل در معادله، بیانی از میزان اهمیت آن متغیر در پیش‌بینی مقادیر متغیر وابسته است.

¹ - Variable reduction

² - Independent variable(s)

³ - Genetic Algorithm

⁴ - Replacement method

۲-۲-۶- رگرسیون خطی ساده^۱

رگرسیون خطی ساده، یک محاسبه استاندارد رگرسیون خطی را برای تولید دسته‌ای از معادله های QSAR، که شامل یک معادله برای هر متغیر مستقل است، انجام می‌دهد. هر معادله شامل یک متغیر از هر دسته توصیف‌کننده است. این روش برای کشف ارتباطات بین ساختار و فعالیت مناسب است [۲۶]

۲-۲-۷- رگرسیون مرحله‌ای

یکی از روش‌های انتخاب توصیف‌کننده‌های مناسب، روش رگرسیون مرحله‌ای^۲ است که از طریق محاسبه ضریب همبستگی^۳ بین توصیف‌کننده و فعالیت بیولوژیکی انجام می‌شود. اگر توصیف‌کننده‌ای دارای ضریب همبستگی برابر یک باشد، بدین معنی است که می‌تواند فعالیت مورد نظر را به درستی توصیف کند و ضریب همبستگی صفر بدین معنی است که هیچ ارتباطی بین توصیف‌کننده و فعالیت دارویی وجود ندارد.

۲-۲-۸- مقدمه ای بر الگوریتم ژنتیک

الگوریتم ژنتیک یکی از زیر مجموعه‌های محاسبات تکامل یافته می‌باشد که رابطه مستقیمی با مبحث هوش مصنوعی^۴ دارد، در واقع الگوریتم ژنتیک یکی از زیر مجموعه‌های هوش مصنوعی است. محدوده‌ی کاری الگوریتم ژنتیک بسیار وسیع بوده و هر روزه با پیشرفت روزافزون علوم و تکنولوژی، استفاده از این روش در بهینه‌سازی و حل مسائل بسیار گسترش یافته است [۲۷].

¹ - Simple linear regression

² - Stepwise

³ - Correlation coefficient

⁴ - Artificial intelligence

الگوریتم ژنتیک را می‌توان یک روش جستجوی کلی نامید که از قوانین تکامل بیولوژیک طبیعی تقلید نموده و بر روی یکسری از جواب‌های مسأله به امید بدست آوردن جواب‌های بهتر، قانون بقای بهترین را اعمال می‌کند [۲۸].

۲-۲-۸-۱- تاریخچه

حساب تکاملی، برای اولین بار در سال ۱۹۶۰ توسط ریچنبرگ^۱ ارائه شد. بعدها نظریه او توسط محققان زیادی مورد بررسی قرار گرفت تا اینکه الگوریتم ژنتیک (GA) توسط جان هولند^۲ در سال ۱۹۷۵ در دانشگاه میشیگان ارائه شد [۲۹]. در سال ۱۹۹۲ جان کوزا^۳ از الگوریتم ژنتیک برای حل و بهینه‌سازی مسائل مهندسی پیشرفته استفاده کرد و توانست برای اولین بار روند الگوریتم ژنتیک را به زبان کامپیوتر تبدیل کند و برای آن یک زبان برنامه‌نویسی ابداع کرد. این روش برنامه‌نویسی را برنامه‌نویسی ژنتیک (GP)^۴ گویند و نرم‌افزاری که توسط وی ابداع گردید به نرم‌افزار LISP^۵ مشهور است که هم‌اکنون نیز این نرم‌افزار کاربرد زیادی در حل و بهینه‌سازی مسائل مهندسی پیدا کرده است [۳۰].

۲-۲-۸-۲- تعریف الگوریتم ژنتیک

حل هر مسأله با الگوریتم ژنتیک یک فرآیند پویا یا تکاملی را شامل می‌شود. الگوریتم با یک مجموعه‌ای به نام جمعیت آغاز می‌شود، سپس پاسخ‌ها از جمعیت گرفته شده و برای ایجاد جمعیت جدیدی بکار می‌روند. هدف این است که جمعیت جدید از جمعیت پیشین بهتر باشد، از اینرو معیار انتخاب^۶ پاسخ‌هایی که برای ایجاد پاسخ‌های جدید انتخاب می‌شوند شایستگی^۱ آن‌ها خواهد بود و به

¹ - Reichenberg

² - John Holland

³ - John Koza

⁴ - Genetic Programming

⁵ - Lots of Irritating Superfluous Parentheses

⁶ - Selection

عبارتی هرچه شایستگی یک پاسخ بیشتر باشد، شانس بیشتری برای تولید نسل دارد. این امر تا زمانی که شرایط خاصی (به عنوان مثال تعداد جمعیت‌ها یا بهبود بهترین جواب) حکمفرما گردد تکرار می‌شود.

الگوریتم ژنتیک در مسائل متنوعی نظیر بهینه‌سازی، شناسایی و کنترل سیستم، پردازش تصویر و مسایل ترکیبی، تعیین توپولوژی و آموزش شبکه‌های عصبی مصنوعی و سیستم‌های مبتنی بر تصمیم و قاعده به کار می‌رود [۳۱].

۲-۸-۳- مراحل کلی الگوریتم ژنتیک

- ۱- آغاز: تولید جمعیت اتفاقی n کروموزومی (پاسخ‌هایی حدسی برای مسأله).
- ۲- شایستگی: محاسبه‌ی شایستگی $f(X)$ متناظر با هر کروموزوم X در جمعیت.
- ۳- جمعیت جدید: ایجاد یک جمعیت جدید بوسیله تکرار مراحل زیر تا هنگامی که جمعیت جدید کامل گردد:
 - الف- انتخاب: انتخاب دو کروموزوم والدین از جمعیت با توجه به شایستگی آن‌ها (هرچه شایستگی بالاتر باشد شانس انتخاب بالاتر خواهد بود).
 - ب- پیوند: پیوند والدین با یک احتمال (نرخ) پیوند^۱ برای تشکیل فرزندان.
 - ج- جهش: تغییر فرزندان جدید با یک احتمال (نرخ) جهش^۲، از طریق تغییر مقدار ژن‌ها در کروموزوم.
 - د- پذیرش: قرار دادن فرزندان جدید در یک جمعیت جدید.
- ۴- جایگزینی: بکار بردن جمعیت تازه ایجاد شده برای سیکل بعدی اجرای الگوریتم.

¹ - Fitness

² - Crossover probability/rate

³ - Mutation probability/rate

۵- آزمایش: اگر شرایط انتهایی بدست آید برنامه متوقف شده و بهترین جواب موجود در جمعیت را معرفی می کند.

۶- بازگشت: بازگشت به مرحله ۲.

۲-۲-۸-۴- ساختار الگوریتم های ژنتیکی

به طور کلی، الگوریتم های ژنتیکی از اجزاء زیر تشکیل می شوند:

* کروموزوم

* جمعیت

* تابع برازندگی

۲-۲-۸-۴-۱- کروموزوم^۱

در الگوریتم های ژنتیکی، هر کروموزوم نشان دهنده یک نقطه در فضای جستجو و یک راه حل ممکن برای مسئله مورد نظر است. خود کروموزومها (راه حلها) از تعداد ثابتی ژن^۲ (متغیر) تشکیل شده اند و برای نمایش کروموزومها، معمولاً از کدگذاری های باینری^۳ استفاده می شود.

۲-۲-۸-۴-۲- جمعیت^۴

مفهوم جمعیت در الگوریتم ژنتیک بدین صورت است که چه تعداد کروموزوم در جمعیت وجود دارد. اگر تعداد کمی کروموزوم داشته باشیم، GA احتمال کمی برای اجرای پیوند دارد و تنها یک

¹ - Chromosome

² - Gene

³ - Binary codes

⁴ - Population

بخش کوچک از فضای جستجو مورد استفاده قرار می‌گیرد. از طرف دیگر اگر تعداد کروموزوم‌ها بسیار زیاد باشد، GA سرعت کمی خواهد داشت.

۲-۲-۸-۴-۳- تابع برازندگی^۱

برای مدل کردن، منطق انتخاب طبیعی در الگوریتم ژنتیکی از تابع برازندگی بهره می‌بریم که در واقع، این تابع مهمترین قسمت الگوریتم ژنتیک است. بسته به نوع مسأله، تابع برازندگی کروموزوم‌ها را به صورت کمی ارزیابی می‌کند، بطوری که انتخاب درست این تابع دارای بیشترین اثر در صحت عملکرد الگوریتم ژنتیک می‌باشد. برای مثال، این تابع می‌تواند محاسبه کننده خطا در پاسخ یک مسأله یا انرژی مربوط به ساختار یک پروتئین باشد. در این مثال، کروموزوم‌هایی برای تولید مثل شایسته‌تر انتخاب می‌شوند که خطای کمتری در جواب مسأله ایجاد کرده و یا پروتئینی با سطح انرژی ساختاری پایین‌تر را رمزبندی کنند.

۲-۲-۸-۵- کد گذاری^۲

الگوریتم ژنتیک به جای این که بر روی پارامترها یا متغیرهای مسأله کار کند، با شکل کد شده آن‌ها سروکار دارد. در آغاز حل یک مسأله با GA، کدگذاری کروموزوم‌ها اولین کاری است که باید انجام شود و کدگذاری نیز بستگی کامل به خود مسأله دارد. یک کروموزوم بایستی حاوی اطلاعاتی مطابق با جواب حاصل باشد. روشی که بیش از همه برای کدگذاری مورد استفاده قرار گرفته است روش کدینگ باینری می‌باشد. البته راه‌های بسیاری برای کدگذاری وجود دارد ولی روش کدگذاری اساساً به ماهیت مسأله بستگی دارد. در این بخش برخی روش‌های کدگذاری که تاکنون با موفقیت نسبی به کار رفته‌اند معرفی خواهد شد.

¹ - Fitness function

² - Encoding

۲-۲-۸-۵-۱- روش‌های کد گذاری

* کدینگ باینری

* کدینگ جایگشتی^۱

* کدینگ مقدار^۲

* کدینگ درختی^۳

کدینگ باینری

کدگذاری باینری ساده‌ترین کدگذاری و بهترین تبدیل برای عملگرهای ژنتیک است. در این کدگذاری هر کروموزوم به وسیله یک رشته دودویی نشان داده می‌شود و اعضای جمعیت به رشته‌هایی از صفر و یک تبدیل می‌شوند. همچنین این روش می‌تواند تعداد کروموزوم‌های ممکن زیادی را با کمک تعداد محدودی ژن ارائه دهد [۳۲].

کدینگ جایگشتی

در این روش، کروموزوم‌ها به صورت رشته‌ای از اعداد طبیعی نشان داده می‌شوند که هر کدام از این اعداد، مربوط به پارامتر ویژه‌ای در فضای حل مسأله است. ترتیب قرارگیری این اعداد مهم بوده و طول رشته دقیقاً با تعداد پارامترهای تعریف شده در مسأله برابر است [۳۲].

کدینگ مقدار

در این نوع روش کدگذاری، کروموزوم‌ها می‌توانند هر نوع داده مرتبط با مسأله را در رشته خود اختیار نمایند. این داده‌ها می‌توانند از نوع اعداد حقیقی، عبارات منطقی، دستورات جهت‌یابی و داده‌های کد شده به صورت رشته‌های حرفی باشند.

^۱ - Permutation coding

^۲ - Value coding

^۳ - Tree coding

در شکل زیر مثالی از کدگذاری مقدار نشان داده شده است. همان طور که مشاهده می گردد به

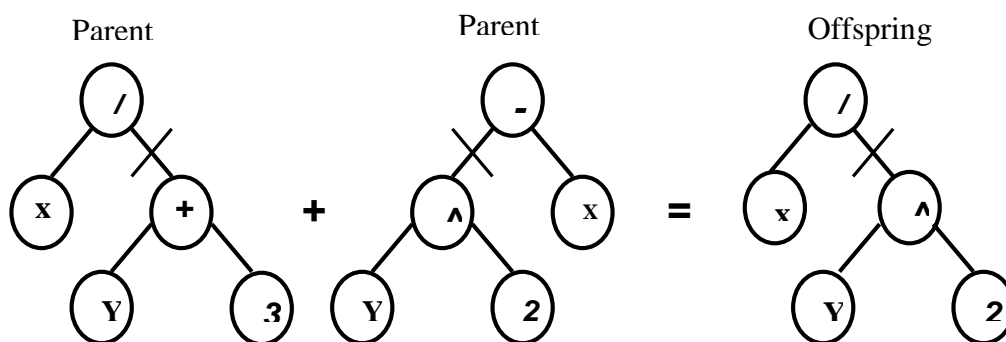
برخی ژن ها بطور تصادفی عددی اضافه شده یا از آن ها کم می گردد [۳۲].

$$(5.68 \mathbf{2.86} \mathbf{4.11} 5.55) = (5.68 \mathbf{2.73} \mathbf{4.22} 5.55)$$

شکل (۲-۱) - کدینگ مقدار

کدینگ درخت

این روش کدگذاری که توسط «جان کوزا» توسعه یافت، بیشتر برای برنامه نویسی ژنتیک توسط نرم افزار «LISP» مورد استفاده قرار می گیرد که برنامه ها را به عنوان شاخه های داده در ساختار درخت نشان می دهد. در این روش تغییرات تصادفی می توانند با عوض کردن عملگرها یا تغییر دادن ارزش یک گره داده شده در درخت، یا عوض کردن یک زیردرخت با دیگری به وجود آیند. به زبان ساده تر می توان این طور گفت که در این نوع کدگذاری یک درخت دودویی برای عبارت (کروموزوم) تشکیل می دهیم. مثالی از این نوع کدگذاری در شکل (۲-۲) نشان داده شده است [۳۳ و ۳۲].



شکل (۲-۲) - کدینگ درختی [۳۳]

۲-۲-۸-۶- عملگرهای الگوریتم ژنتیک

در الگوریتم‌های ژنتیکی، در طی مرحله تولید مثل^۱ از عملگرهای ژنتیکی استفاده می‌شود. با تأثیر این عملگرها بر روی یک جمعیت، نسل^۲ بعدی آن جمعیت تولید می‌شود. عملگرهای انتخاب^۳، پیوند^۴ و جهش^۵ معمولاً بیشترین کاربرد را در الگوریتم‌های ژنتیکی دارند [۳۴-۳۶].

۲-۲-۸-۶-۱- عملگر انتخاب

قبل از طرح کلی GA، کروموزوم‌ها برای اینکه نقش والدین را در فرآیند پیوند بازی کنند از میان جمعیت انتخاب می‌شوند. اما مسأله چگونگی انتخاب این کروموزوم‌ها می‌باشد. در مرحله انتخاب، یک جفت از کروموزوم‌ها برگزیده می‌شوند تا با هم ترکیب شوند، عملگر انتخاب، رابط بین دو نسل است و بعضی از اعضای نسل کنونی را به نسل آینده منتقل می‌کند، بعد از انتخاب، عملگرهای ژنتیک روی دو عضو برگزیده اعمال می‌شوند، معیار در انتخاب اعضاء، ارزش تطابق آن‌ها می‌باشد اما روند انتخاب حالتی تصادفی دارد.

روش‌های متداول انتخاب

- انتخاب چرخ رولت^۶

- انتخاب نخبه گرایی^۷

- انتخاب رقابتی^۸

¹ - Reproduction

² - Generation

³ - Selection

⁴ - Crossover

⁵ - Mutation

⁶ - Roulette wheel selection

⁷ - Elitism selection

⁸ - Rivalry selection

انتخاب چرخ رولت

انتخاب چرخ رولت که اولین بار توسط «هولند»^۱ پیشنهاد شد یکی از مناسب‌ترین انتخاب‌های تصادفی بوده که ایده آن، احتمال انتخاب می‌باشد. احتمال انتخاب متناظر با هر کروموزوم، براساس برازندگی آن محاسبه شده که اگر f_k مقدار برازندگی کروموزوم k ام باشد، احتمال بقای متناظر با آن کروموزوم (P_k) عبارت است از:

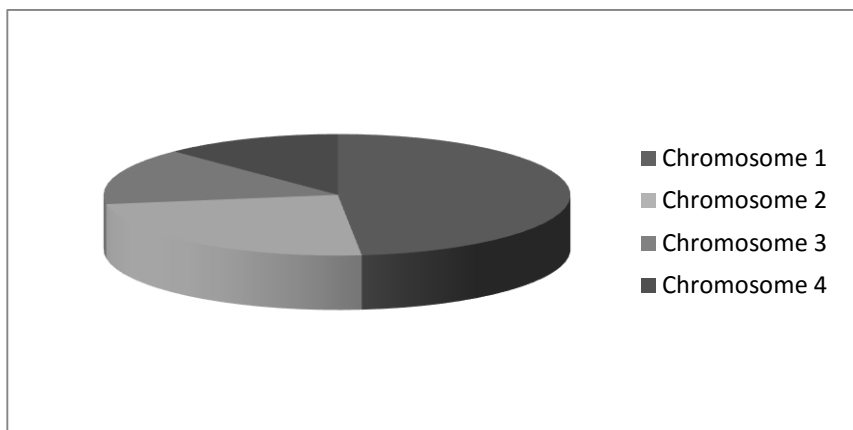
$$P_k = \frac{f_k}{\sum_{i=1}^n f_i} \quad (2-2)$$

کروموزوم‌ها براساس P_k مرتب می‌شوند و q_k همان مقادیر تجمعی P_k می‌باشد که به صورت زیر به دست می‌آید:

$$q_k = \sum_{i=1}^k P_i \quad (3-2)$$

چرخ رولت به این صورت عمل می‌کند که برای انتخاب هر کروموزوم یک عدد تصادفی بین یک و صفر تولید کرده و عدد مذکور در هر بازه‌ای که قرار گرفت، کروموزوم متناظر با آن انتخاب می‌شود. البته روش پیاده‌سازی چرخ رولت به این شکل است که ما یک دایره را در نظر گرفته و آن را به تعداد کروموزوم‌ها طوری تقسیم می‌کنیم که هر بخش متناظر با مقدار برازندگی کروموزوم مربوط باشد سپس چرخ را چرخانده و هر کجا که چرخ متوقف شد به شاخص چرخ نگاه کرده، کروموزوم مربوط به آن بخش انتخاب می‌گردد.

¹ - Holland



شکل (۲-۳) - چرخ رولت [۳۶]

انتخاب نخبه‌گرایی

ایده‌ی نخبه‌گرایی، ویژگی تازه‌ای به پروسهٔ انتخاب اضافه می‌کند، در نخبه‌گرایی، بهترین عضو هر جمعیت زنده می‌ماند و در جمعیت بعد حضور دارد، به عبارت دیگر عضوی که بالاترین تطابق را دارد به طور خودکار به جمعیت جدید منتقل می‌شود. این روش ابتدا در سال ۱۹۷۵ توسط کندی جونز^۱ معرفی شد. اعمال نخبه‌سالاری در الگوریتم ژنتیک، معمولاً باعث بهبود کارایی آن می‌شود.

انتخاب رقابتی

این روش تعدادی از اعضای جمعیت را به تصادف انتخاب می‌کند و سپس اگر شرطی خاص برقرار باشد، بهترین یا تعدادی از بهترین‌های آن‌ها را به عنوان والد بر می‌گزیند، اگر شرط برقرار نشود، بدترین عضو یا تعدادی از بدترین‌ها، در تشکیل جمعیت آینده به عنوان والد در نظر گرفته می‌شوند.

۲-۲-۸-۶-۲ - عملگر پیوند

در جریان عمل پیوند به صورت اتفاقی بخش‌هایی از کروموزوم‌ها با یکدیگر تعویض می‌شوند. این موضوع باعث می‌شود که فرزندان ترکیبی از خصوصیات والدین خود را به همراه داشته باشند و

^۱ - Kennedy Jones

دقیقاً مشابه یکی از والدین نباشند. هدف تولید فرزند جدید می باشد به این امید که خصوصیات خوب دو موجود در فرزندشان جمع شده و یک موجود بهتری را تولید کند.

روش های متداول پیوند

❖ پیوند تک نقطه ای^۱

اگر عملیات پیوند را در یک نقطه انجام دهیم، به آن پیوند تک نقطه ای می گویند. پیوند بدین صورت انجام می گیرد که حاصل ترکیب کروموزوم های پدر و مادر می باشد. روش تولید مثل نیز بدین صورت است که ابتدا نقطه ای که قرار است تولید مثل از آنجا آغاز شود، بصورت تصادفی انتخاب می گردد، سپس اعداد بعد از آن به ترتیب از بیت های کروموزوم های پدر و مادر قرار می گیرد.

❖ پیوند دو نقطه ای^۲

در این روش دو مکان را به صورت تصادفی انتخاب کرده و مقادیر را بین این دو نقطه جابجا می کنیم.

❖ پیوند چند نقطه ای^۳

می توانیم این عملیات را در چند نقطه انجام دهیم، که به آن پیوند چند نقطه ای می گویند.

❖ پیوند جامع^۴

اگر تمام نقاط کروموزوم را بعنوان نقاط پیوند انتخاب کنیم به آن پیوند جامع می گوئیم.

¹ - Single point crossover

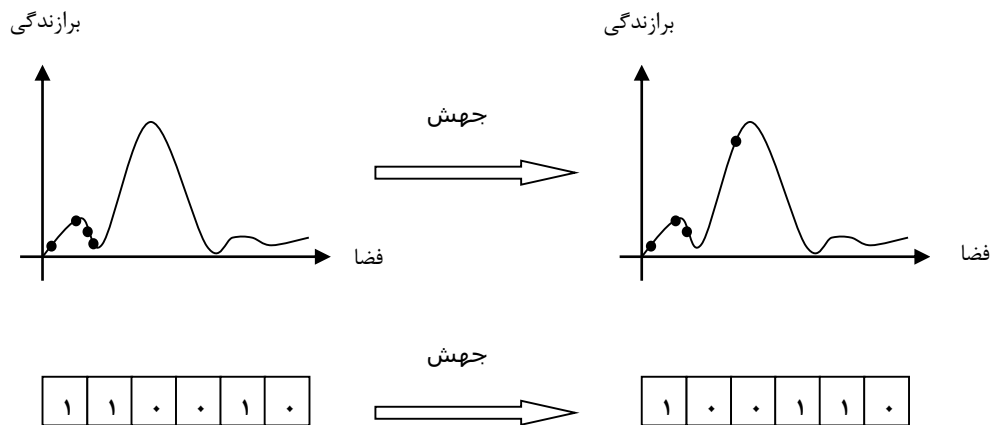
² - Two point crossover

³ - Multi point crossover

⁴ - Uniform crossover

۲-۲-۸-۶-۳- عملگر جهش

در الگوریتم‌های ژنتیکی، جهش با هدف جلوگیری از سقوط تمامی جواب‌ها در جمعیت به یک بهینه‌ی محلی انجام می‌گیرد، همچنین جهش باعث جستجو در فضاهای دست نخورده مسأله می‌شود. می‌توان استنباط کرد که مهمترین وظیفه جهش، اجتناب از همگرایی به بهینه محلی است. در اشکال زیر نحوه جهش و کارکرد آن نمایش داده شده است:

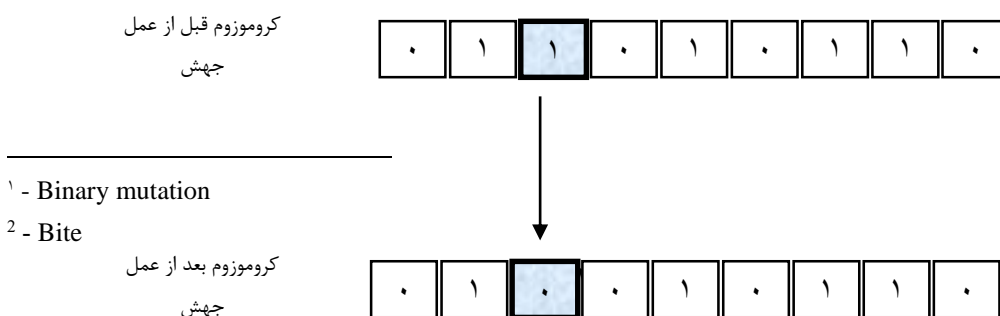


شکل (۲-۴) - شبیه سازی جهش به کمک نمودار [۳۶]

انواع جهش

جهش باینری^۱

در الگوریتم ژنتیک با جهش باینری، این عملگر اغلب با تولید تصادفی یکی از اعداد ۰ و ۱ و جایگزینی آن به جای بیت^۲ مورد نظر صورت می‌گیرد، اما در برخی کاربردهای ژنتیک عمل جهش دودویی در یک بیت با متمم ساختن آن بیت انجام می‌شود. به این صورت که اگر بیت مورد نظر صفر بوده به بیت یک و بر عکس تبدیل خواهد شد. آزمایش‌ها نشان داده که روش دوم مناسب‌تر است.



شکل (۲-۵) - جهش باینری

جهش حقیقی^۱

در جهش حقیقی، عملگر جهش باعث تولید تصادفی یک مقدار جدید در یک موقعیت خاص در کروموزوم می‌شود، در نتیجه این تغییرات تصادفی در جمعیت کروموزوم‌ها، نواحی بیشتری از فضای کاوش بررسی شده و از همگرایی بی‌موقع (ناگهانی محلی) الگوریتم جلوگیری می‌شود.

۲-۲-۸-۷ - نقاط قوت الگوریتم‌های ژنتیک

اولین و مهمترین نقطه قوت الگوریتم‌های ژنتیک این است که ذاتاً موازی‌اند. اکثر الگوریتم‌های دیگر موازی نیستند و فقط می‌توانند فضای مسأله مورد نظر را در یک جهت در یک لحظه جستجو کنند و اگر راه‌حل پیدا شده یک پاسخ بهینه محلی باشد و یا زیر مجموعه‌ای از پاسخ اصلی باشد باید تمام کارهایی که تا به حال انجام شده را کنار گذاشت و دوباره از اول شروع کرد. از آنجایی که الگوریتم ژنتیک چندین نقطه شروع دارد، در یک لحظه می‌تواند فضای مسأله را از چند جهت مختلف جستجو کند. نقطه قوت دیگر الگوریتم‌های ژنتیک که در ابتدا یک کمبود به نظر می‌رسد این است که، هیچ چیزی در مورد مسائلی که حل می‌کنند، نمی‌دانند.

۲-۲-۸-۸ - محدودیت‌های الگوریتم ژنتیک

یکی از محدودیت‌های الگوریتم ژنتیک، نوشتن عملگر ارزیاب است که منجر به پیدا کردن بهترین راه‌حل برای مسأله می‌شود. اگر این عملگر به خوبی و قوی انتخاب نشود، ممکن است باعث

^۱ - Real mutation

شود که راه‌حلی برای مسأله پیدا نکنیم یا مسأله‌ای دیگر را به اشتباه حل کنیم. به علاوه برای انتخاب تابع مناسب برای ارزیاب، پارامترهای دیگری مثل اندازه جمعیت، نرخ ترکیب، قدرت و نوع انتخاب هم باید مورد توجه قرار گیرند. مشکل دیگر این است که اگر یک ژنوم که فاصله‌اش با سایر ژنوم‌های نسلش زیاد باشد خیلی زود دیده می‌شود (ایجاد می‌شود)، ممکن است محدودیت ایجاد کند و راه‌حل را به سوی جواب بهینه محلی سوق دهد. این اتفاق معمولاً در جمعیت‌های کم اتفاق می‌افتد [۳۷].

۲-۸-۹- الگوریتم ژنتیک و QSAR

اولین کاربرد الگوریتم ژنتیک در کمومتریکس بوسیله لاردی^۱ در سال ۱۹۹۲ گزارش شد که الگوریتم ژنتیک را برای انتخاب متغیر و رگرسیون خطی چند گانه را برای بهینه‌سازی شکل و اندازه خراباها بکار برد [۳۸]. الگوریتم ژنتیک می‌تواند برای انتخاب متغیر و مدل‌سازی با روش‌های دیگر از قبیل رگرسیون خطی چند گانه، حداقل مربعات جزئی و شبکه عصبی به کار برده شود [۳۹].

۲-۲-۹- ساختن مدل

مدل، در واقع یک رابطه ریاضی است که بیان‌کننده‌ی رابطه بین متغیر مستقل و متغیر وابسته است و به کمک آن می‌توان با داشتن مقادیر متغیر مستقل، متغیر وابسته را تخمین زد. مرحله اصلی در مطالعات QSAR ساختن مدل می‌باشد. در این تحقیق از روش شبکه عصبی برای ساخت مدل استفاده شد و متغیرهای به دست آمده از دو روش رگرسیون مرحله‌ای و الگوریتم ژنتیک، به طور جداگانه در ساخت مدل استفاده شدند و در نهایت مقایسه‌ای بین دو مدل صورت گرفت.

۲-۲-۱۰- مقدمه‌ای بر شبکه عصبی

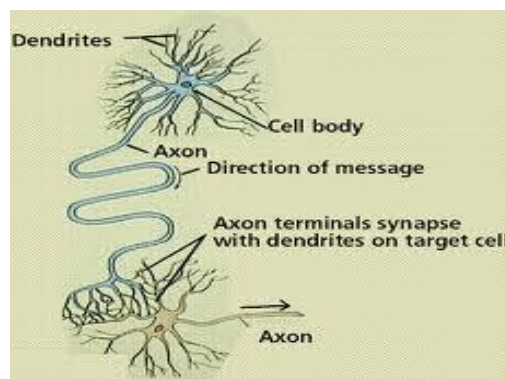
شبکه عصبی یک برنامه نرم‌افزاری است که می‌تواند همانند مغز انسان عمل نماید. در واقع یک شبکه عصبی مصنوعی ایده‌ای است برای پردازش اطلاعاتی که از سیستم عصبی زیستی الهام گرفته و

^۱ - Lardy

مانند مغز انسان عمل می‌کند. شبکه عصبی مصنوعی یکی از بهترین روش‌های مدل‌سازی است، زیرا توانایی زیادی در شبیه‌سازی سیستم‌های غیرخطی^۱ دارد [۴۰]. طی چند دهه اخیر تلاش‌های بسیار جدی جهت طراحی مدارهای الکترونیکی که قادر باشند شبکه‌های عصبی زیستی را همانندسازی کنند صورت گرفته است. نحوه پردازش علائم و ذخیره اطلاعات توسط سیستم عصبی انسان برای مدت‌های طولانی ناشناخته بوده است. برای مدل‌سازی یک شبکه عصبی مصنوعی می‌توان از یک مدل ریاضی که خصوصیات یک سیستم بیولوژیکی را توصیف می‌کند، استفاده کرد. سپس با استفاده از یک رایانه می‌توان این مدل را به سرعت و به نحو مناسبی شبیه‌سازی نمود [۴۱].

۲-۲-۱-۱۰-۱- ساختمان و عملکرد نرون طبیعی

طبق نظر محققان واحد اصلی و سازنده سیستم عصبی، اجزایی به نام نرون^۲ می‌باشند. ساختمان اصلی یک نرون شامل جسم سلولی^۳، آکسون^۴ و دندریت‌ها^۵ می‌باشد که دو قسمت آخر، عناصر ارتباطی نرون را تشکیل می‌دهد. نرون‌ها واحدهای اصلی و تشکیل‌دهنده سیستم عصبی بوده و عملکردهای بیوشیمیایی و پردازشگری سیگنال را دارا می‌باشند [۴۲]. شکل (۲-۶) ساختار یک نرون زیستی را نشان می‌دهد.

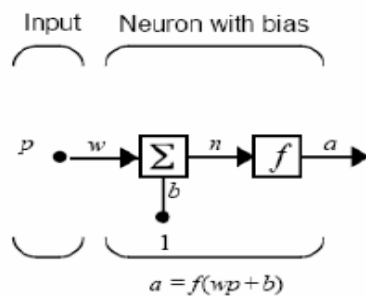


شکل (۲-۶) - ساختار یک نرون زیستی [۴۳]

1 - Non liner
2 - Neuron
3 - Cell body
4 - Axon
5 - Dendrite

۲-۱۰-۲-۲ - ساختمان و عملکرد نرون مصنوعی

نرون عصبی را می‌توان به صورت شبکه‌ای از عناصر محاسباتی به هم پیوسته به نام گره در نظر گرفت. هر گره، تبدیلی بر روی سیگنال‌های ورودی‌اش صورت داده و سیگنال خروجی تولید می‌کند که به نوبه‌ی خود به گره‌های دیگر فرستاده خواهد شد.



شکل (۲-۷) - ساختمان یک نرون محاسباتی [۴۴]

در شکل بالا، از علامت p برای نشان دادن یک سیگنال ورودی استفاده شده است. در واقع در این مدل، یک سیگنال ورودی پس از تقویت یا تضعیف شدن به اندازه پارامتر w ، به صورت یک سیگنال الکتریکی به اندازه pw وارد نرون می‌شود. به جهت ساده‌سازی مدل ریاضی، فرض می‌شود که در هسته سلول عصبی، سیگنال ورودی با سیگنال دیگری به اندازه b جمع می‌گردد.

w و b دو پارامتر تنظیم شونده در نرون‌ها می‌باشند و ایده‌ی اصلی شبکه‌ی عصبی این است که با تغییر مقادیر w و b ، شبکه یک رفتار یا تصمیم را اتخاذ کند [۴۴].

۲-۱۰-۳ - بهبود تعمیم^۱

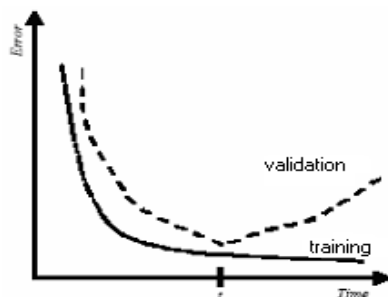
یکی از مشکلات در آموزش شبکه‌های عصبی، برآزش بیش از حد^۲ شبکه است. اگر شبکه بیش از حد آموزش داده شود، شروع به خاطر سپردن داده‌ها می‌کند و از پیدا کردن ارتباط موجود در داده‌ها دور می‌شود. به عبارتی در بیش برآزش، خطا بر روی مجموعه آموزش به مقدار قابل قبول می‌رسد

^۱- Improving generalization

^۲- Over fitting

ولی با ارائه داده های جدید به عنوان ورودی، خطای شبکه بسیار بالا می‌باشد. در اصطلاح گفته می‌شود شبکه تعمیم برای حالت‌های جدید را یاد نگرفته است. یک روش برای ارتقای عمومیت شبکه، بکارگیری شبکه‌هایی 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. در جعبه ابزار مطلب برای جلوگیری از بیش برآزش و ارتقای عمومیت روش‌های مختلفی از جمله توقف زودرس^۱، تنظیم^۲ وجود دارد [۴۵].

در توقف زودرس برای آموزش شبکه، داده‌ها به‌طور تصادفی به سه زیر مجموعه آموزش، ارزیابی، تست تقسیم می‌شوند، درحین آموزش شبکه، زمانی که شبکه سعی در تطابق بیش از حد دارد، طبق شکل (۸-۲) خطای مجموعه ارزیابی شروع به افزایش می‌کند که در این هنگام باید آموزش متوقف شود.



شکل (۸-۲) - تغییرات خطای سری آموزش و سری ارزیابی

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.

$$F = \beta E_D + \alpha E_W \quad (۴-۲)$$

^۱ - Early stopping
^۲ - Regularization

استفاده از این تابع کارایی اصلاح شده سبب می‌شود که شبکه، وزن‌ها و بایاس‌های کوچکتری داشته باشد و این پاسخ شبکه را هموارتر می‌کند. در این رابطه α و β پارامترهای تابع کارایی، E_D و E_W به ترتیب میانگین مربعات خطاها و میانگین مربعات وزن‌ها بوده و طبق روابط (۲-۵) و (۲-۶) تعریف می‌شوند.

$$E_D = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - t_i)^2 \quad (۲-۵)$$

$$E_W = \frac{1}{N} \sum_{j=1}^N (w_j)^2 \quad (۲-۶)$$

در الگوریتم آموزشی تنظیم بایزین، اصلاح وزن‌ها و بایاس‌ها همانند الگوریتم لونبرگ 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 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، α و β

۲- اجرای الگوریتم لونبرگ مارکوارت جهت کاهش تابع کارایی (رابطه ۲-۴)

۳- محاسبه تعداد مؤثر پارامترهای تابع کارایی (γ)

$$\gamma = N - 2\alpha \text{tr}(H)^{-1} \quad (۲-۷)$$

که در این رابطه N تعداد کل پارامترها و $\text{tr}(H)^{-1}$ اثر ماتریس معکوس هسیان است. که منظور از اثر ماتریس همان مجموع عناصر قطری ماتریس می‌باشد.

۴- تنظیم مجدد پارامترهای تابع کارایی با استفاده از روابط (۲-۸) و (۲-۹)

$$\alpha = \frac{\gamma}{2E_W} \quad (۲-۸)$$

$$\beta = \frac{n - \gamma}{2E_D} \quad (9-2)$$

که در رابطه (۹-۲) n تعداد خطاها می باشد.

۵- تکرار مراحل ۲ تا ۴ برای رسیدن به همگرایی

شرایط پایان آموزش و توقف این الگوریتم همانند الگوریتم لونبرگ مارکوارت است [۴۶].

۲-۲-۱۰-۴- توابع انتقال^۱

سه تابع انتقال رایج در شبکه‌های عصبی که در این پژوهش برای بهینه‌سازی شبکه به کار

گرفته شده‌است عبارت‌اند از:

تابع انتقال لگاریتم سیگموئید^۲ (logsig)

از این تابع انتقال در شبکه‌های پس‌انتشار استفاده می‌شود. این تابع مقادیر ورودی را در

محدوده‌ی منفی بی‌نهایت تا مثبت بی‌نهایت دریافت کرده و خروجی بین ۰ و ۱ تولید می‌نماید.

تابع انتقال تانژانت هایپربولیک^۳ (tansig)

این تابع انتقال مقادیر ورودی را در محدوده‌ی منفی بی‌نهایت تا مثبت بی‌نهایت دریافت کرده و

خروجی بین ۱+ و ۱- تولید می‌نماید.

تابع انتقال خطی^۴ (purelin)

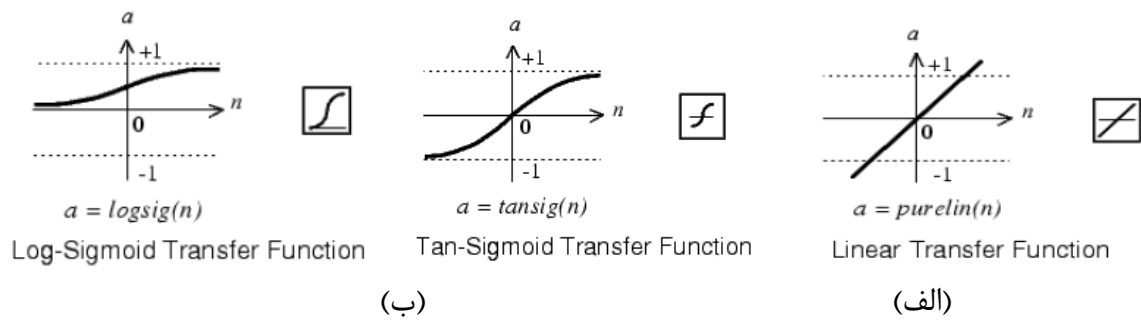
¹ - Transfer functions

² - Logarithm sigmoid transfer function

³ - Hyperbolic tangent

⁴ - Linear transfer function

نرون‌هایی که از این تابع انتقال استفاده می‌کنند، معمولاً برای تقریب خطی در فیلترهای خطی به کار می‌روند. این تابع همان مقدار ورودی را به عنوان خروجی بر می‌گرداند. شکل (۹-۲) مقایسه‌ی بین توابع انتقال را نشان می‌دهد.



شکل (۹-۲) - توابع انتقال (الف) خطی، (ب) سیگموئیدی [۴۷]

۲-۲-۱۰-۵ - ساختارهای شبکه

- ❖ شبکه‌های تک لایه^۱
- ❖ شبکه‌های چند لایه^۲
- ❖ شبکه‌های پرسپترون چند لایه^۳
- ❖ شبکه‌های پیشخور^۴
- ❖ شبکه‌های برگشتی^۵

شبکه‌های تک لایه

¹ - One Layer Networks
² - Multi Layer Networks
³ - Multi Layer Perceptron Networks
⁴ - Feed Forward Networks
⁵ - Recurrent Networks

شبکه‌های تک لایه، شبکه‌هایی هستند که مسیر پاسخ در آن‌ها همواره رو به جلو پردازش می‌شود و به نرون‌های لایه قبل باز نمی‌گردد. در ساده‌ترین شکل شبکه عصبی، یک لایه ورودی از گره‌های منبع داریم که در نهایت از طریق وزن‌ها به نرون‌های لایه خروجی وصل می‌شوند [۴۸].

شبکه‌های چند لایه

این دسته از شبکه‌های عصبی، شامل یک یا چند لایه میانی (پنهان) است که نرون‌های این لایه‌ها، نرون‌های مخفی^۱ نیز نامیده می‌شوند. با اضافه کردن چند لایه میانی، شبکه توانایی تجزیه و تحلیل‌های با درجه پیچیدگی بالاتر را خواهد داشت. توابع فعال‌سازی در لایه خروجی عمدتاً بر اساس نیاز کاربر انتخاب می‌شوند. هر نرون در هر لایه از شبکه به واسطه وزن به هر کدام از نرون‌های لایه بعدی خود متصل شده است. چنین ساختاری را شبکه کامل متصل می‌گویند. حال چنان‌چه بخشی از این ارتباطات وجود نداشته باشند، آن‌را نیمه‌متصل^۲ می‌گویند [۴۸].

شبکه‌های پرسپترون چند لایه [۴۹]

مدل ریاضی یک سلول عصبی که از روی الگوی ریاضی آن ساخته شده است پرسپترون^۳ نامیده می‌شود.

پرسپترون را می‌توان بصورت یک سطح تصمیم ابرصفحه^۴ در فضای n بعدی نمونه‌ها در نظر گرفت. پرسپترون برای نمونه‌های یک طرف صفحه مقدار ۱ و برای مقادیر طرف دیگر مقدار ۱- بوجود

می‌آورد. خروجی پرسپترون نیز توسط رابطه زیر مشخص می‌شود:

$$O(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{if } w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n > 0 \\ -1 & \text{Otherwise} \end{cases} \quad (10-2)$$

شبکه‌های پیشخور [۵۰ و ۵۱]

¹ - Hidden Neurons
² - Semi-conjunct
³ - Perceptron
⁴ - Hyperplane

شبکه‌های پیشخور، شبکه‌هایی هستند که مسیر پاسخ در آن‌ها همواره رو به جلو پردازش می‌شود و به نرون‌های لایه قبل باز نمی‌گردد. این نوع شبکه‌ها به سیگنال اجازه می‌دهد که از ورودی تا خروجی در مسیر یک‌طرفه عبور کند. بنابراین بازخوردی^۱ وجود ندارد؛ بدین معنی که خروجی هر لایه تأثیری بر همان لایه و همچنین لایه‌های قبلی ندارد.

شبکه‌های برگشتی [۵۱ و ۵۰]

در ساختار این نوع از شبکه‌های عصبی حداقل یک حلقه پسخور وجود دارد که در آن خروجی هر نرون، سایر نرون‌های شبکه را تغذیه می‌کند، بدین صورت که یک سیگنال از یک نرون به همان نرون یا نرون‌های همان لایه یا لایه‌های قبل برمی‌گردد.

لازم بذکر است برای اطلاعات بیشتر در رابطه با انواع شبکه‌های عصبی می‌توان به مراجع شماره‌ی (۵۲، ۵۳، ۵۴ و ۵۵) مراجعه نمود.

۲-۱۱-۲- ارزیابی مدل

منظور از ارزیابی مدل، ارائه‌ی شاخص‌های کمی است. به وسیله این شاخص‌ها صحت نتایج ارائه شده توسط مدل، مورد سنجش قرار می‌گیرند. در این تحقیق متناسب با طبیعت پیوسته‌ی جواب‌ها از شاخص‌های ارزیابی مختلفی استفاده شد که نحوه‌ی محاسبه‌ی هر یک از پارامترها به صورت زیر می‌باشد:

ضریب همبستگی^۲

ساده‌ترین راه برای بررسی میزان همبستگی دو یا چند متغیر، محاسبه آماره‌ی ضریب همبستگی آن‌هاست. ضریب همبستگی دو متغیر x, y با رابطه (۲-۱۱) تعریف می‌شود.

^۱ - Feedback

^۲ - Correlation Coefficient

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (11-2)$$

ضریب تعیین^۱

ضریب تعیین به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و نشان‌دهنده‌ی نسبت تغییرات متغیر وابسته توسط متغیر مستقل است. رابطه ریاضی مربوط به ضریب تعیین به صورت زیر است:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (12-2)$$

که SSR ^۲ طبق رابطه (۱۳-۲) بیانگر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است.

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (13-2)$$

عبارت SST ^۳ طبق رابطه (۱۴-۲) نشانگر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (14-2)$$

SSE ^۴ نیز مبین مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (15-2)$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16-2)$$

^۱ - Determination Coefficient

^۲ - Sum Square Regression

^۳ - Sum Square Total

^۴ - Sum Square Error

مجموع مربع باقیمانده‌ها^۱ (PRESS)

PRESS برابر مجموع مربعات اختلاف بین مقدار کمیت مشاهده شده (y_i) و مقدار تخمین زده شده (\hat{y}_i) است.

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (۱۷-۲)$$

خطای استاندارد پیش‌بینی^۲ (SEP)

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (۱۸-۲)$$

خطای مطلق میانگین^۳ (MAE)

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (۱۹-۲)$$

خطای نسبی پیش‌بینی^۴ (REP)

$$REP(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (۲۰-۲)$$

میانگین مربع خطاها^۵ (MSE)

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

میانگین خطای نسبی^۶ (MRE)

$$MRE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \times 100 \quad (۲۲-۲)$$

^۱ - Predictive Residual Sum of Squares

^۲ - Standard Error of Prediction

^۳ - Mean Absolute Error

^۴ - Relative Error of Prediction

^۵ - Mean Square Error

^۶ - Mean Relative Error

۲-۳- نرم افزارهای مورد استفاده

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

۲-۳-۱- بسته نرم‌افزاری Hyperchem

بسته نرم‌افزاری Hyperchem یک نرم‌افزار مدل‌سازی و شبیه‌سازی است که برای رسم شکل مولکول‌ها و بهینه‌سازی ساختار با استفاده از روش‌های کوانتومی و مکانیکی، استفاده می‌شود. به کمک این نرم‌افزار، ساختار ترکیبات به طور تقریبی رسم شد، در ادامه با اضافه کردن اتم‌های هیدروژن و انتخاب گزینه‌ی بهینه‌سازی^۱ هندسی از قسمت محاسبه، شکل هندسی مولکول بهینه شد.

۲-۳-۲- بسته نرم‌افزاری Dragon

نرم‌افزار Dragon توسط گروه کمومتریکس دانشگاه میلان [۵۶] طراحی و عرضه شده است و امکان محاسبه بیش از ۱۴۸۱ توصیف‌کننده مختلف، که به ۱۸ دسته‌ی اصلی تقسیم می‌شوند را فراهم می‌کند. جهت محاسبه توصیف‌کننده به کمک این نرم‌افزار لازم است ساختار هندسی بهینه مولکول مورد استفاده قرار گیرد.

۲-۳-۳- بسته نرم‌افزاری SPSS^۲

نرم‌افزار SPSS [۵۷]، به منظور تجزیه و تحلیل داده‌ها تدوین شده است و یکی از قدرتمندترین ابزارها برای تحلیل آماری داده‌ها به شمار می‌آید. نخستین نسخه آن در سال ۱۹۷۰ توسط جمعی از فارغ‌التحصیلان دانشگاه استنفورد آمریکا ارائه شد. یادگیری آسان، سادگی استفاده و قدرت بالا در

^۱ - Optimization

^۲ - Statistical Package for the Social Science

انجام محاسبات پیچیده، SPSS را به یکی از رایج‌ترین بسته‌های نرم‌افزاری مبدل نموده‌است. در این تحقیق از این نرم‌افزار برای بدست آوردن بهترین مدل استفاده شده است.

۲-۳-۴- نرم‌افزار MATLAB^۱

نرم‌افزار MATLAB [۵۸] یکی از جامع‌ترین و کارآمدترین نرم‌افزارهای علمی و محاسباتی است که طی چند سال گذشته تهیه و به بازار عرضه شده است و در سال‌های اخیر با تدوین نسخه‌های جدیدتر و کامل‌تر روز به روز بر غنای آن افزوده می‌شود.

MATLAB به معنای آزمایشگاه ماتریس است. در این نرم‌افزار، ورودی‌ها اساساً به صورت ماتریس در نظر گرفته می‌شوند و هیچ نیازی به مشخص کردن ابعاد ماتریس نمی‌باشد. اطلاعات شیمیایی به عنوان ورودی و به صورت یک ماتریس به نرم‌افزار داده می‌شود. در محیط این برنامه با استفاده از آرایه‌ها و فرامین موجود امکان مدل‌سازی غیرخطی فعالیت بیولوژیکی با ساختار ترکیبات فراهم می‌گردد. در حال حاضر MATLAB به ابزاری پر قدرت برای ترسیم داده‌ها، برنامه‌نویسی و انجام محاسبات مهندسی و پژوهشی تبدیل شده است.

^۱ - Matrix Laboratory

فصل سوم

مطالعه ارتباط کمی ساختار -
فعالیت برخی مشتقات دی-
هیدروکینولین به عنوان
داروهای ضد اختلال خواب با
استفاده از
SR-ANN و GA-ANN

ولین به عنوان بازدارنده

۱-۳- مدل سازی فعالیت

های بیماری اختلال خواب

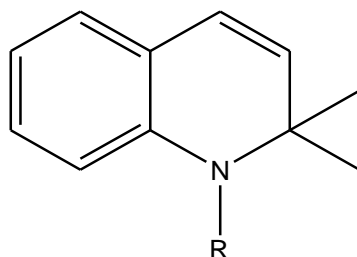
سازی SR-ANN و GA-ANN

به منظور مطالعه ارتباط

برای پیش بینی فعالیت ضد اختلال خواب یک طبقه‌ی جدید از ترکیبات ۱ و ۲- دی‌هیدروکینولین و همچنین مشتقات استر آنها مورد استفاده قرار گرفت. توانایی مدل‌های به دست آمده از هر دو روش در پیش بینی قدرت بازدارندگی این مشتقات، مورد ارزیابی قرار گرفت. به طور کلی این بخش شامل انتخاب سری داده‌ها، رسم و بهینه سازی ساختار مولکول‌ها، محاسبه توصیف‌کننده‌های مولکولی، انتخاب توصیف‌کننده‌های مولکولی مناسب، مدل سازی و ارزیابی مدل برتر است.

۳-۱-۱- انتخاب سری داده‌های استفاده شده در مدل‌سازی

فعالیت دارویی (IC_{50}) برخی مشتقات دی‌هیدرو کینولین به عنوان سری داده‌ها مورد استفاده قرار گرفت [۵۹]. IC_{50} غلظتی از دارو است که ۵۰٪ اثر بازدارندگی داشته باشد. سری داده‌های مورد مطالعه شامل ۴۸ ترکیب می‌باشد. برای مدل‌سازی، مقادیر فعالیت به صورت $\text{Log}(IC_{50} \times 10^{-6})$ (برحسب میکرومولار) وارد شدند که دارای گستره ۴/۰۷ تا ۸/۱۵ می‌باشند. اسکلت اصلی ترکیبات مورد بررسی در شکل (۱-۳) ارائه شده است. همچنین جزئیات استخلاف‌های ترکیبات سری داده‌ها و مقادیر عددی pIC_{50} در جدول (۱-۳) آورده شده است.

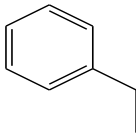
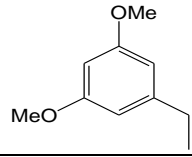
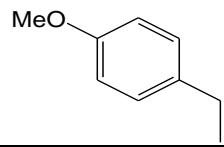
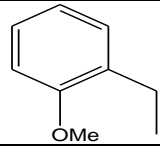
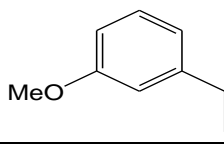
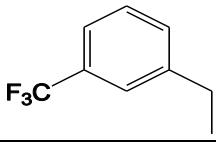
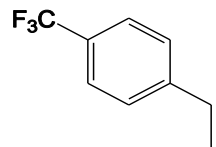
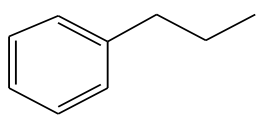


شکل (۱-۳)- اسکلت اصلی ترکیبات مورد مطالعه

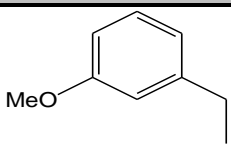
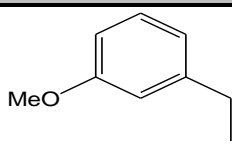
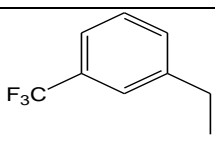
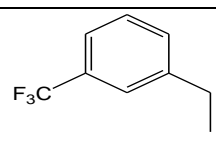
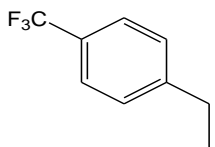
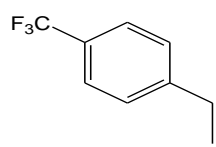
جدول (۱-۳)- ساختار و مقادیر pIC_{50} ترکیبات مورد مطالعه

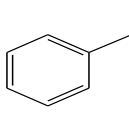
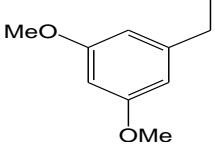
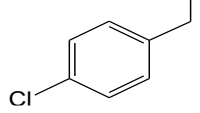
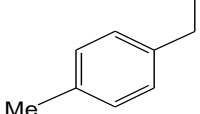
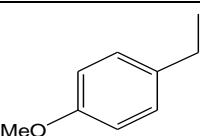
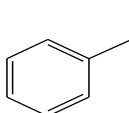
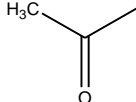
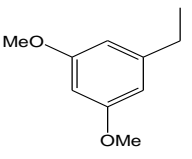
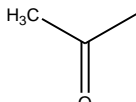
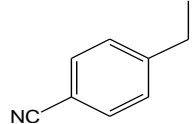
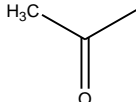
شماره ترکیب	سری داده	R_1	R_2	$IC_{50}(\mu M)$	pIC_{50}
۱	آموزش	H		۰/۰۵۴	۷/۲۷
۲	ارزیابی	H	H	۰/۷۸۰	۶/۱۱

¹ - Half maximal inhibitory concentration

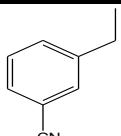
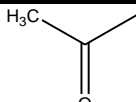
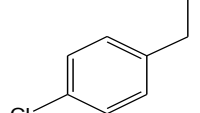
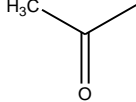
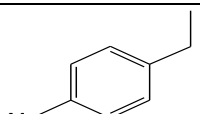
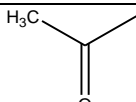
۳	آموزش	H		۴/۱۰۰	۵/۳۹
۴	تست	H		۱۳/۰۰۰	۴/۸۹
۵	آموزش	H		۵/۲۰۰	۵/۲۸
۶	ارزیابی	H		۵/۲۰۰	۵/۲۸
۷	آموزش	H		۱۱/۰۰۰	۴/۹۶
۸	ارزیابی	H		۸/۸۰۰	۵/۰۵
۹	آموزش	H		۴/۵۰۰	۵/۳۵
۱۰	آموزش	H		۳/۷۰۰	۵/۴۳

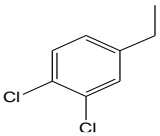
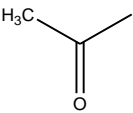
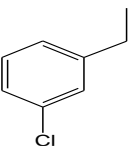
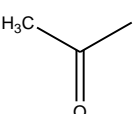
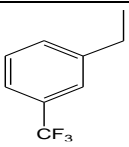
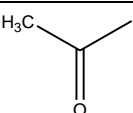
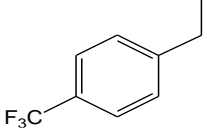
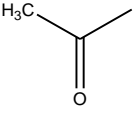
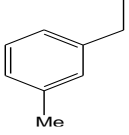
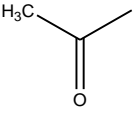
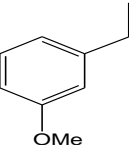
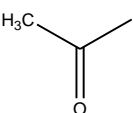
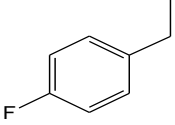
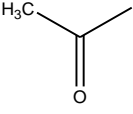
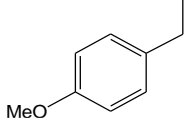
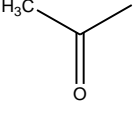
ادامه جدول (۱-۳)

شماره ترکیب	سری داده	R ₁	R ₂	IC ₅₀ (μM)	pIC ₅₀
۱۱	تست			۶۳/۰۰۰	۴/۲۰
۱۲	آموزش			۳/۶۰۰	۵/۴۴
۱۳	آموزش			۸۶/۰۰۰	۴/۰۷

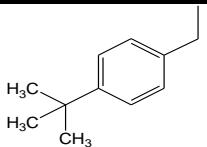
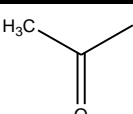
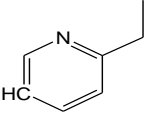
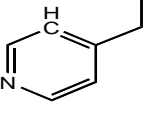
۱۴	آموزش		H	۰/۰۰۷	۸/۱۵
۱۵	ارزیابی		H	۰/۰۵۸	۷/۲۴
۱۶	تست		H	۰/۰۶۱	۷/۲۱
۱۷	آموزش		H	۰/۱۴۰	۶/۸۵
۱۸	ارزیابی		H	۰/۱۷۰	۶/۷۷
۱۹	آموزش			۰/۰۱۴	۷/۸۵
۲۰	آموزش			۰/۱۰۰	۷/۰۰
۲۱	تست			۰/۲۶۰	۶/۵۹

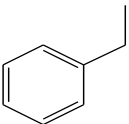
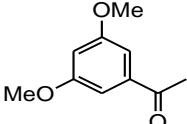
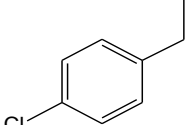
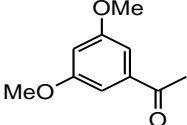
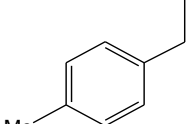
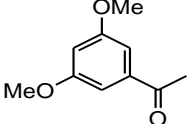
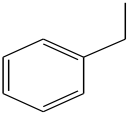
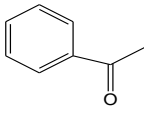
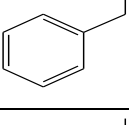
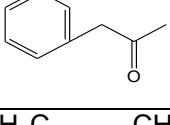
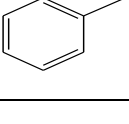
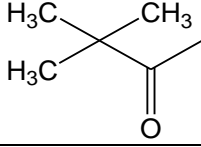
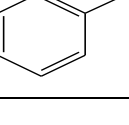
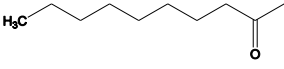
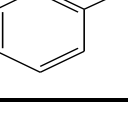
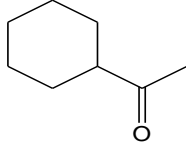
ادامه جدول (۱-۳)

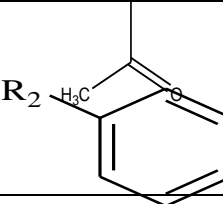
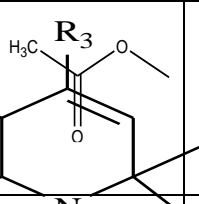
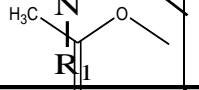
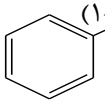
شماره ترکیب	سری داده	R ₁	R ₂	IC ₅₀ (μM)	pIC ₅₀
۲۲	آموزش			۰/۱۸۰	۶/۷۵
۲۳	ارزیابی			۰/۱۸۰	۶/۷۵
۲۴	تست			۰/۳۹۰	۶/۴۱

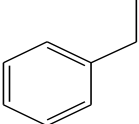
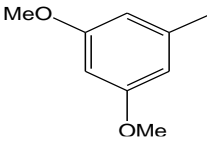
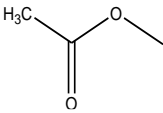
۲۵	آموزش			۰/۰۵۳	۷/۲۸
۲۶	آموزش			۰/۰۳۹	۷/۴۱
۲۷	آموزش			۰/۰۴۱	۷/۳۹
۲۸	تست			۰/۳۶۰	۶/۴۴
۲۹	ارزیابی			۰/۳۴۰	۶/۴۷
۳۰	تست			۰/۳۲۰	۶/۵۰
۳۱	آموزش			۰/۰۴۱	۷/۳۹
۳۲	آموزش			۰/۰۴۸	۷/۳۲

ادامه جدول (۱-۳)

شماره ترکیب	سری داده	R ₁	R ₂	IC ₅₀ (μM)	pIC ₅₀
۳۳	آموزش			۰/۶۷۰	۶/۱۷
۳۴	تست		H	۱/۱۰۰	۵/۹۶
۳۵	آموزش		H	۰/۳۴۰	۶/۴۷

۳۶	آموزش			۰/۰۲۳	۷/۶۴
۳۷	ارزیابی			۰/۰۴۸	۷/۳۲
۳۸	تست			۰/۰۳۳	۷/۴۸
۳۹	آموزش			۰/۰۱۳	۷/۸۹
۴۰	آموزش			۰/۰۱۵	۷/۸۲
۴۱	آموزش			۰/۰۲۸	۷/۵۵
۴۲	تست			۰/۰۱۷	۷/۷۷
۴۳	آموزش			۰/۰۲۸	۷/۵۵

۴۴	ارزیابی			CH ₃	۵/۳۰۰	۵/۲۸
۴۵	آموزش	CH ₃		CH ₃	۱/۴۰۰	۵/۸۵
شماره ترکیب	سری داده	R ₁	R ₂	R ₃	IC ₅₀ (μM)	pIC ₅₀
۴۶	آموزش		ادامه جدول (۱-۳) H	CH ₃	۵۷/۰۰۰	۴/۲۴

۴۷	ارزیابی		OH	H	۰/۰۱۳	۶/۸۹
۴۸	آموزش			H	۰/۲۰۰	۶/۷۰

۳-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها

در این مرحله از مطالعه، ساختار مولکولی هر ترکیب ابتدا در نرم‌افزار Hyperchem 8.0 ترسیم شد. سپس با احتساب اتم‌های هیدروژن، ساختار سه بعدی ترکیبات با استفاده از روش نیمه تجربی کوانتومی AM1 بهینه شد و بهینه‌سازی تا زمانی ادامه یافت که جذر میانگین مربعات گرادیان^۱ انرژی به ۰/۰۰۱ کیلو کالری بر مول برسد.

۳-۱-۳- محاسبه توصیف‌کننده‌ها

ساختارهای سه بعدی رسم شده برای ۴۸ ترکیب موجود به عنوان داده‌های ورودی به نرم‌افزار Dragon 2.1 داده شد که این نرم‌افزار تعداد ۱۴۸۱ توصیف‌کننده را در ۱۸ گروه مختلف برای داده‌ها محاسبه نمود.

۳-۱-۴- انتخاب توصیف‌کننده‌های مهم

مهمترین مرحله در مطالعات QSAR انتخاب توصیف‌کننده‌های مناسبی است که بتوانند فعالیت دارویی را به ساختار مولکول ارتباط دهند. با عنایت به این موضوع و با توجه به اینکه تعداد زیاد توصیف‌کننده باعث پیچیدگی محاسبات می‌شود و همچنین تعدادی از توصیف‌کننده‌ها حاوی اطلاعات یکسان و در نتیجه اثر یکسان بر روی فعالیت مهارکنندگی هستند، لذا باید روش کاهش

^۱ - Root Mean Square Gradient

تعداد متغیرها به کار گرفته شود. به همین منظور، ابتدا توصیف‌کننده‌هایی که دارای مقادیر یکسانی برای کل مولکول‌ها بودند از فهرست توصیف‌کننده‌ها حذف شدند. سپس با بررسی همبستگی بین متغیرها، از میان هر دو توصیف‌کننده‌ای که ضریب همبستگی بزرگتر از ۰/۹ داشتند، آنکه همبستگی کمتری با متغیر وابسته داشت، حذف شد. بدین ترتیب در پایان این مرحله تعداد ۱۴۱ توصیف‌کننده باقی ماند. در این تحقیق برای انتخاب توصیف‌کننده‌ها از روش خطی رگرسیون مرحله‌ای (SR) و روش الگوریتم ژنتیک (GA) استفاده شده است.

۳-۴-۱- انتخاب توصیف‌کننده‌های مناسب توسط رگرسیون مرحله‌ای در SPSS

در روش رگرسیون مرحله‌ای، متغیرها یکی پس از دیگری وارد مدل شده و ابتدا متغیری وارد مدل می‌شود که بالاترین میزان همبستگی را با متغیر وابسته دارد. با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آن‌ها سطح معناداری خود را از دست داده باشند، قبل از ورود متغیر جدید از مدل خارج می‌شوند. بدین ترتیب از میان ۱۴۱ توصیف‌کننده ارائه شده تعداد ۶ توصیف‌کننده انتخاب شدند. نام این توصیف‌کننده‌ها همراه با معنی و طبقه آن‌ها در جدول (۳-۲) ارائه شده است.

جدول (۳-۲) - توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای

No	Symbol	Class	Meaning
1	qneg	Charge	Maximum negative charge
2	MATS6m	2D autocorrelation	Moran auto correlation - lag6
3	MATS1e	2D autocorrelation	Moran auto correlation - lag1
4	nROR	Functional group	Number of ethers
5	Mor10u	3D MORSE	3D -MORSE signal10
6	JGI1	Galvez topol. charge indices	Mean topological charge

همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۳-۳) ارائه شده که این ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول (۳-۳) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای

	qneg	MATS6m	MATS1e	nROR	Mor10u	JGI1
qneg	۱/۰۰۰					
MATS6m	-۰/۲۲۷	۱/۰۰۰				
MATS1e	۰/۳۰۳	-۰/۳۱۸	۱/۰۰۰			
nROR	-۰/۲۲۵	-۰/۱۳۵	-۰/۰۲۷	۱/۰۰۰		
Mor10u	-۰/۵۰۳	۰/۵۰۶	-۰/۴۰۸	-۰/۰۰۷	۱/۰۰۰	
JGI1	۰/۰۰۸	-۰/۵۳۹	-۰/۴۶۵	-۰/۰۴۸	۰/۲۰۸	۱/۰۰۰

۳-۱-۴-۲- انتخاب توصیف‌کننده‌های مناسب توسط الگوریتم ژنتیک در MATLAB

الگوریتم ژنتیک یک روش بهینه‌سازی است که در علوم مختلف توجه خاصی را به خود جلب کرده است. الگوریتم ژنتیک هیچ چیزی درباره مسأله‌ای که حل می‌کند، نمی‌داند و برای اجرای جستجوی مؤثر فقط به مقادیر تابع هدف نیاز دارد.

یکی از روش‌های انتخاب متغیر در مطالعات QSAR استفاده از روش GA-PLS می‌باشد. کمترین مربعات جزئی (PLS) یک روش آماری برای انتخاب بهترین مدل می‌باشد. بهترین مدل انتخاب شده برای مجموعه‌ای از داده‌ها، مدلی است که در آن مجموع مربعات باقیمانده^۱ مقداری کمینه باشد. منظور از باقیمانده‌ها، اختلاف بین داده‌های تئوری و تجربی می‌باشد که از مدل به دست می‌آید. از آنجایی که تعداد متغیر زیاد عملکرد تحلیلی PLS را مختل می‌کند باید یک روش انتخاب متغیر را برای یافتن مجموعه‌ی مناسب از توصیف‌کننده‌ها به کار گرفت که در این پروژه روش GA برای انتخاب متغیر بکار گرفته شد [۶۰ و ۶۱]. برنامه‌ی GA-PLS استفاده شده در این پایان‌نامه تحت نرم‌افزار MATLAB نوشته شده است.

^۱ - Sum of squared residuals

در فرآیند اجرای الگوریتم ژنتیک، ابتدا به صورت اتفاقی یک جمعیت اولیه از توصیف‌کننده‌ها تولید شده و برآزندگی یا شایستگی تک‌تک اعضای هر نسل یعنی همان توصیف‌کننده‌ها، محاسبه و با توجه به شایستگی‌ها، نسل‌های بعدی با اعمال سه پارامتر جمعیت، پیوند و جهش تولید یا باز ترکیب می‌شوند.

جمعیت

این برنامه کروموزوم ساخته شده به وسیله الگوریتم ژنتیکی را گرفته و بر اساس آن، توصیف‌کننده‌های منتخب را بطور خودکار از مجموعه‌ی داده‌ها انتخاب می‌کند. همچنین به صورت تصادفی ترکیبات را به دو سری آموزش و آزمون تقسیم می‌نماید. اکثر محققان برای اندازه جمعیت، عددی بین ۳۰ تا ۱۰۰ پیشنهاد کرده‌اند. این مجموعه از جمعیت‌های انتخاب شده اولین نسل را ایجاد نموده و از نسل اولیه نیز نسل‌های بعدی ایجاد می‌شود. سرانجام در نسل نهایی، یک جمعیت انتخاب شده که موجب حل بهینه‌ی مسأله است، به دست می‌آید [۶۲]. در این مطالعه، اندازه جمعیت عدد ۳۰ فرض شده است [۶۱].

عملگر پیوند

عملگر پیوند می‌تواند دو نقطه‌ای باشد که در این صورت، دو نقطه به طور تصادفی انتخاب شده و پیوند بین آن دو اتفاق می‌افتد. این عملیات در چند نقطه انجام شده که به آن باز ترکیبی چند نقطه ای می‌گویند. معمولاً احتمال پیوند برای هر زوج کروموزوم $0/1$ تا $0/9$ در نظر گرفته می‌شود [۶۳]. در این تحقیق این مقدار $0/5$ فرض شده است [۶۱].

عملگر جهش

جهش، عملگر دیگری است که جواب‌های ممکن را ایجاد می‌کند. در جهش ممکن است ژنی از مجموعه ژن‌های جمعیت حذف شود یا ژنی که تا به حال در جمعیت وجود نداشته است به آن اضافه

شود. جهش یک ژن به معنای تغییر آن ژن است و وابسته به نوع کدگذاری، روش‌های متفاوت جهش استفاده می‌شود [۶۴]. احتمال وقوع این عملگر بین ۰/۰۰۵ تا ۰/۰۱ انتخاب می‌شود (۰/۵ تا ۱ درصد).

تعداد تکرار	توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک
1	FDI- MATS6m- nROR- Mor10u- Mor17m- MSD- ATS5v- qneg- Mor11m- MATS1e
2	FDI- qneg- nROR- MATS6m- H8u- RDF030m- Mor17m- Mor10u- Mor27v- MATS1e- GATS6e

در این مطالعه مقدار ۰/۰۱ در نظر گرفته شده است [۶۱].

برنامه‌ی الگوریتم ژنتیک استفاده شده در این پروژه یک برنامه‌ی انتخاب متغیر می‌باشد، بنابراین برای اجرای آن نیازی به مرحله‌ی رمزگشایی نبوده و تمام ژن‌ها تک بیتی در نظر گرفته شدند، همچنین هر ژن معرف یک توصیف‌کننده می‌باشد. به منظور اجرای برنامه الگوریتم ژنتیک تعداد ۱۴۱ توصیف‌کننده‌ای که بالاترین میزان همبستگی را با متغیر وابسته داشتند به عنوان ورودی به این برنامه داده شدند. در جدول (۳-۴) توصیف‌کننده‌های انتخاب شده در هر بار اجرای الگوریتم ژنتیک گزارش شده است. بدلیل اینکه توصیف‌کننده‌های تصادفی حذف گردند و نتایج خوبی حاصل گردد این روش ۵ بار تکرار گردید.

3	FDI- qneg- Mor17m- MATS6m- Mor11m- MSD- RDF030m- Mor10u- nROR- MATS1e- Mor11u
4	FDI- MATS6m- qneg- Mor10u- Mor26u- Mor11u- MATS1e- Mor17m- nROR
5	FDI- MATS6m- qneg- nROR- H8u- Mor13v- Mor127v- MATS1e- Mor10u- Mor26u- Mor17m

جدول (۳-۴) - توصیف‌کننده‌های انتخاب شده توسط الگوریتم ژنتیک

سپس توصیف‌کننده‌های انتخاب شده در این ۵ بار تکرار را مجدداً وارد برنامه GA نموده و توصیف‌کننده‌های انتخاب شده در این مرحله به‌عنوان توصیف‌کننده‌های انتخاب شده نهایی در نظر گرفته شد که به ترتیب زیر می‌باشند:

FDI- MATS6m- qneg- nROR- Mor10u- Mor17m- MATS1e

با توجه به موارد ذکر شده، توصیف‌کننده‌های انتخاب شده نهایی توسط این روش همراه با گروه و معنی آن‌ها در جدول (۳-۵) آورده شده است.

جدول (۳-۵) - توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک

No	Symbol	Class	Meaning
1	FDI	Geometrical	Folding degree index
2	MATS6m	2D autocorrelation	Moran auto correlation – lag6
3	qneg	Charge	Maximum negative charge
4	nROR	Functional group	Number of ethers
5	Mor10u	3D MORSE	3D-MORSE- signal10

6	Mor17m	3D MORSE	3D –MORSE-signal17
7	MATS1e	2D auto correlation	Moran auto correlation – lag1

همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۳-۶) ارائه شده که این

ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول (۳-۶) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک

	FDI	MATS6m	qneg	nROR	Mor10u	Mor17m	MATS1e
FDI	۱/۰۰۰						
MATS6m	۰/۱۶۴	۱/۰۰۰					
qneg	۰/۴۶۶	-۰/۲۲۷	۱/۰۰۰				
nROR	-۰/۵۵۹	-۰/۱۳۵	-۰/۲۲۵	۱/۰۰۰			
Mor10u	-۰/۱۲۶	۰/۵۰۶	-۰/۵۰۳	-۰/۰۰۷	۱/۰۰۰		
Mor17m	۰/۱۰۱	۰/۱۶۴	۰/۴۶۶	-۰/۵۵۹	-۰/۱۲۶	۱/۰۰۰	
MATS1e	-۰/۱۶۲	-۰/۳۱۸	۰/۳۰۳	-۰/۰۲۷	-۰/۴۰۸	-۰/۱۶۲	۱/۰۰۰

۳-۱-۵- مدل‌سازی شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های

انتخاب شده توسط رگرسیون مرحله‌ای

شبکه عصبی انتخاب شده در این مطالعه یک شبکه پیشخور با الگوریتم آموزشی پس‌انتشار می

باشد که الگوریتم آن با استفاده از نرم‌افزار مطلب نوشته شد. شبکه طراحی شده از سه لایه تشکیل

شده است. هر شبکه دارای حداقل یک لایه ورودی، یک لایه خروجی و تعدادی لایه پنهان می‌باشد.

متأسفانه هیچ راهنمای مناسبی برای انتخاب تعداد لایه‌های پنهان وجود ندارد بنابراین ساختار شبکه

های عصبی، اغلب به روش سعی و خطا ایجاد می‌شود. مقادیر توصیف‌کننده‌های وارد شده در مدل

خطی به عنوان ورودی به شبکه عصبی داده شد. مقدار تجربی فعالیت ضد اختلال خواب نیز به عنوان

هدف به دنبال ورودی‌ها وارد شبکه شدند تا پاسخ شبکه با آنها سنجیده شود. همچنین با توجه به اینکه اگر دفعات تکرار کم باشد خطای زیادی داریم و اگر تعداد تکرار زیاد باشد، آموزش بیش از حد رخ می‌دهد بنابراین باید تعداد چرخه‌هایی که طی آن زوج ورودی- هدف در شبکه وارد می‌شوند، بهینه شود.

برای به دست آوردن بهترین مدل و کمترین خطا، پارامترهای شبکه (تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه، تعداد گره‌ها در لایه مخفی و تعداد دورهای آموزش) بهینه سازی شدند.

۳-۱-۵-۱- بهینه‌سازی پارامترهای شبکه

برای بهینه‌سازی پارامترهای شبکه و شناسایی بهترین پارامترهای آن، شبکه‌هایی با تعداد ورودی‌های از ۲ تا ۶ توصیف‌کننده ایجاد شدند. هر شبکه با توابع آموزش لوببرگ- مارکوارت و تنظیم بایزین با تعداد گره‌های متفاوت از ۲ تا ۱۰ گره در لایه پنهان آموزش داده شد. این درحالی بود که برای بدست آوردن بهترین تابع انتقال در لایه پنهان مدل‌های شبکه‌ی عصبی مصنوعی طراحی شده، از توابع لگاریتم سیگموئید و تانژانت سیگموئید به عنوان تابع انتقال استفاده گردید و تعداد دورهای آموزش نیز از ۲ تا ۱۵ تغییر داده شد. همچنین معیار نیز به حداقل رساندن میانگین مربع خطا (MSE) در نظر گرفته شد.

۳-۱-۵-۲- انتخاب تعداد متغیرهای ورودی شبکه، تابع آموزش، نوع تابع انتقال، تعداد

گره‌های لایه‌های پنهان و تعداد دورهای آموزش

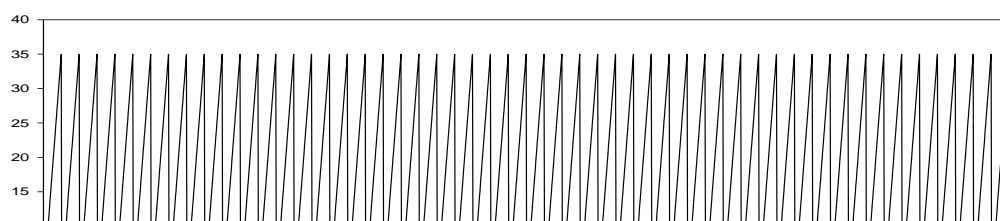
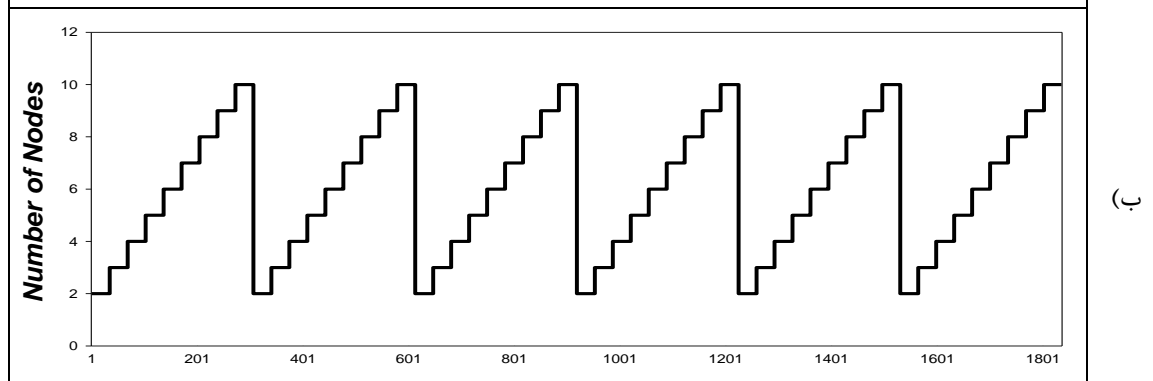
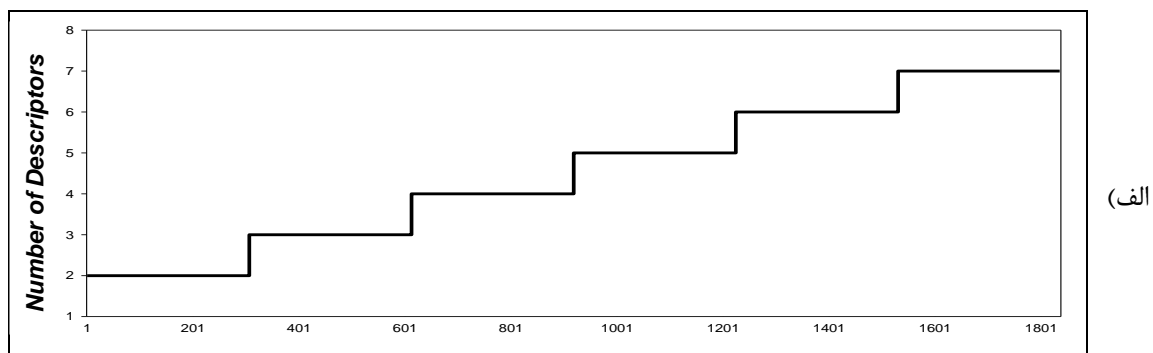
در این مرحله، بهینه‌سازی و انتخاب بهترین تعداد متغیرهای ورودی شبکه، نوع تابع آموزشی، نوع تابع انتقال، تعداد نرون‌های لایه پنهان و تعداد دورهای آموزش به طور همزمان انجام گرفت. این کار موجب می‌شود بتوانیم نقش تمام پارامترها را به طور همزمان و یکجا بررسی نماییم.

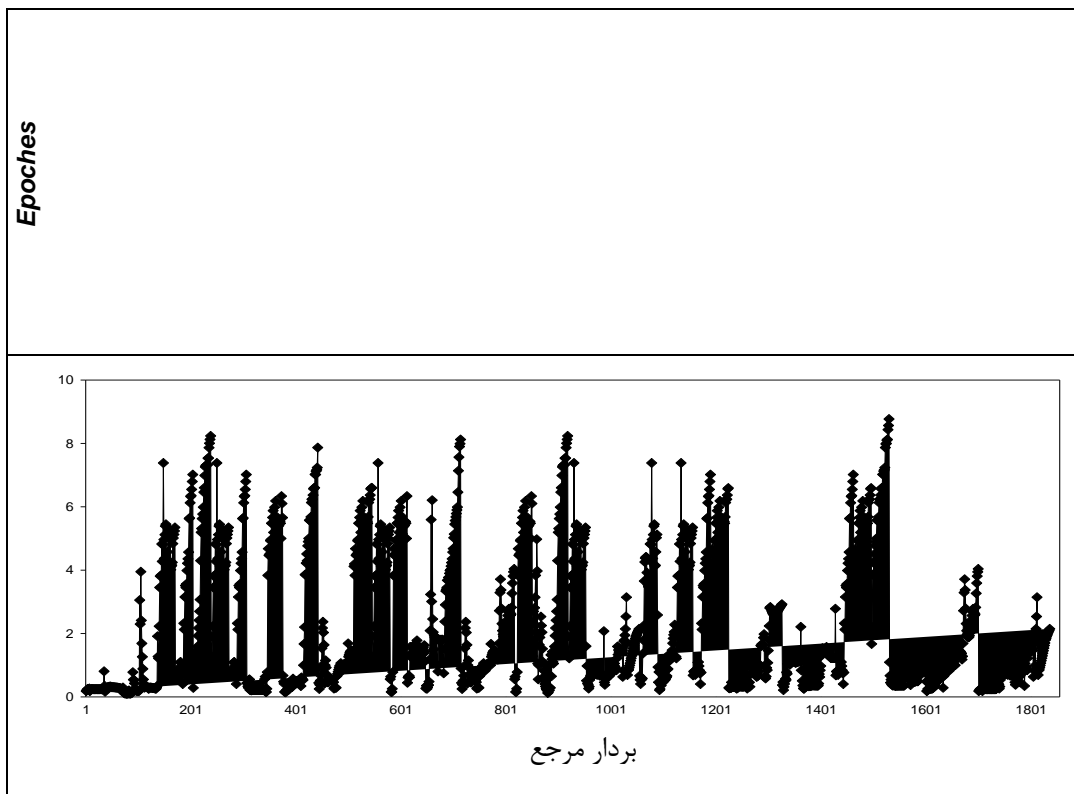
جهت بهینه‌سازی همزمان تمام پارامترها، در مجموع ۲۵۲۰ شبکه مورد ارزیابی قرار گرفته است که در زیر به تعداد تمام موارد مورد بررسی اشاره شده است.

$$۲۵۲۰ = \text{دوره‌های آموزش (۱۴)} \times \text{گره (۹)} \times \text{تابع انتقال (۲)} \times \text{تابع آموزش (۲)} \times \text{توصیف‌کننده (۵)}$$

بنابراین خودکار شدن شبکه برای بدست آوردن نتایج مربوط به هر شبکه منحصر بفرد می‌تواند به میزان زیادی زمان مورد نیاز جهت بهینه‌سازی شبکه را کاهش دهد.

بخشی از روند تغییرات پارامترهای شبکه در حین بهینه‌سازی همزمان پارامترها به همراه مقادیر MSE به دست آمده به صورت نموداری بر حسب یک بردار مرجع فرضی در شکل‌های (۳-۲)، (۳-۳)، (۴-۳) و (۵-۳) آمده است.

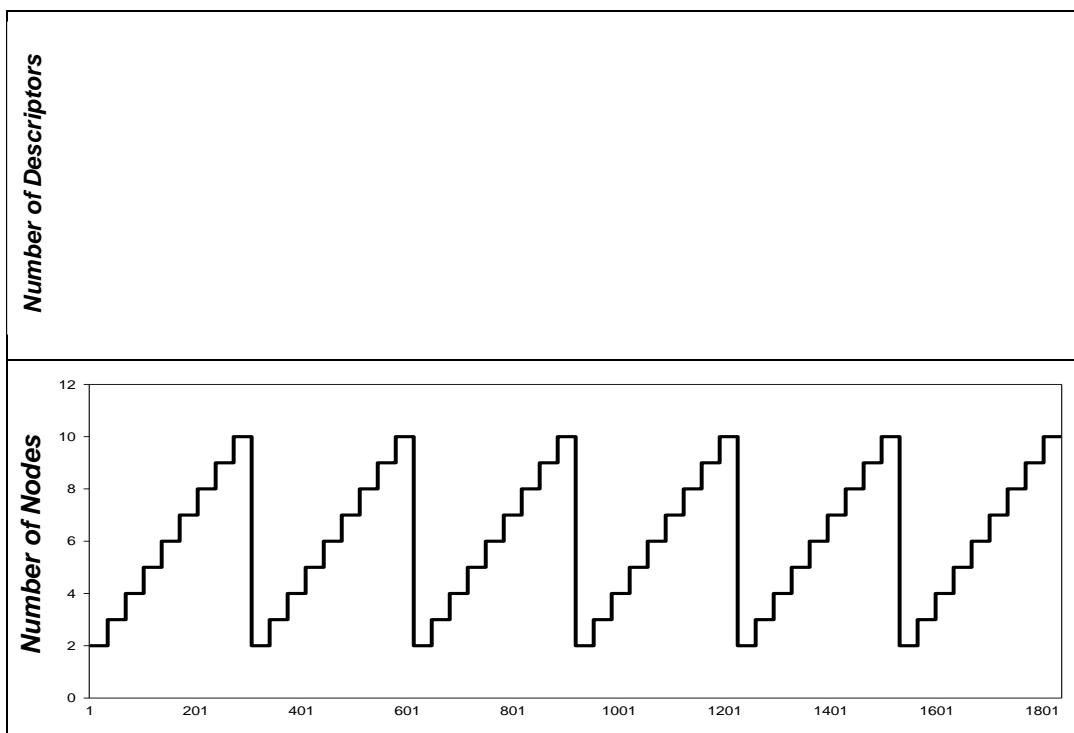




ج

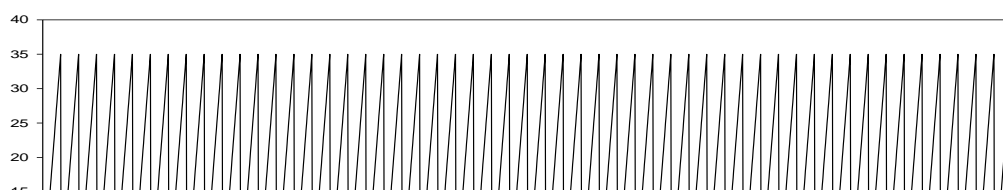
د

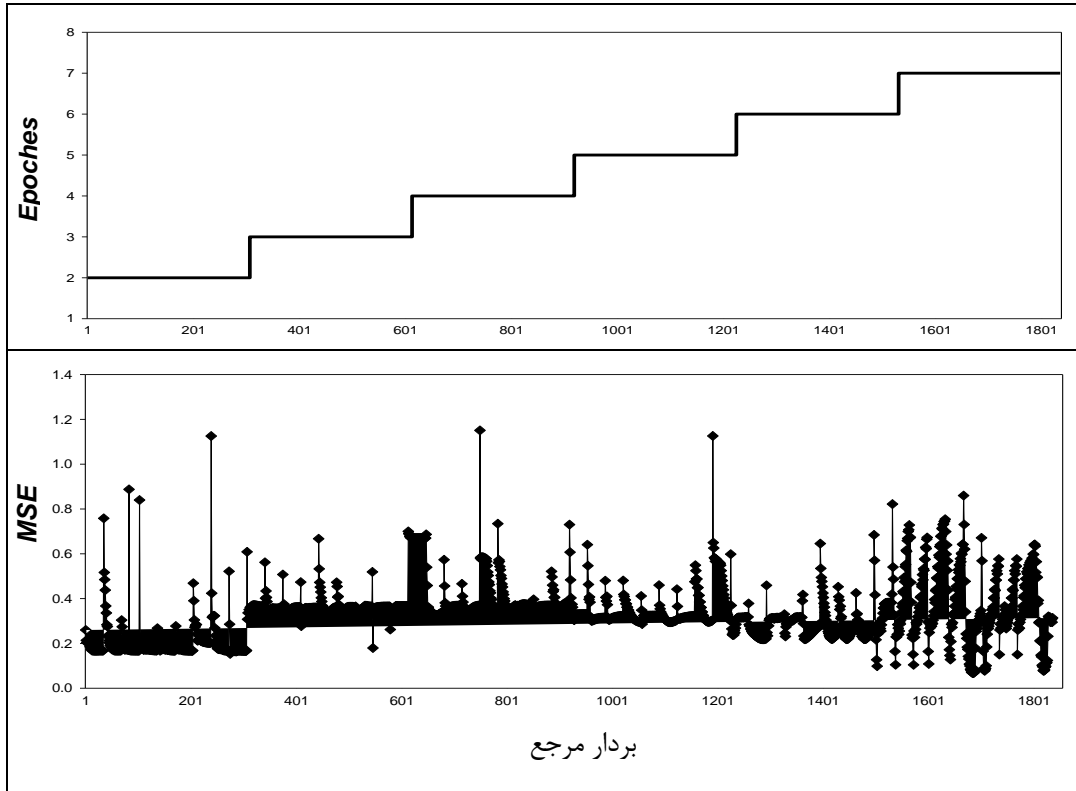
شکل (۳-۲) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر برای تابع MSE آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید



الف

ب

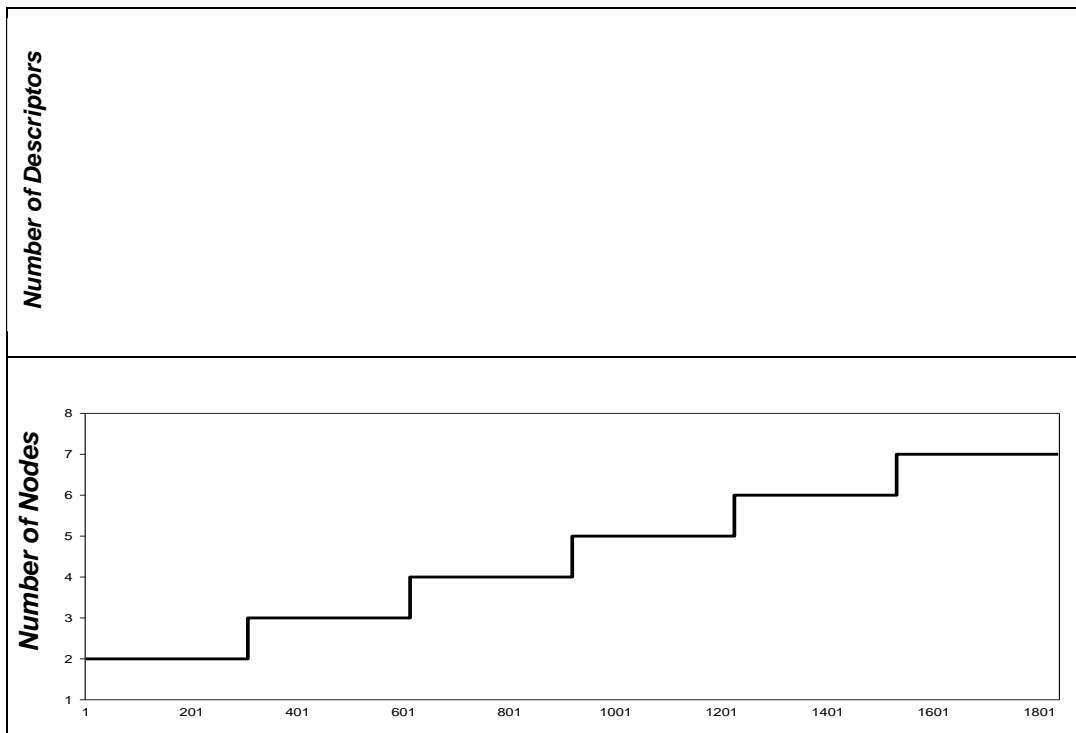




(ج)

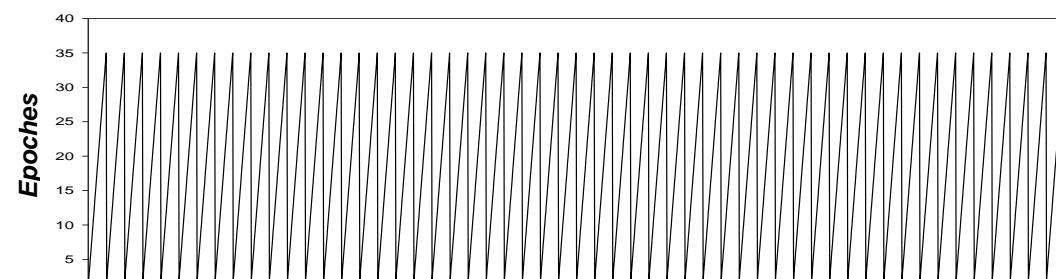
(د)

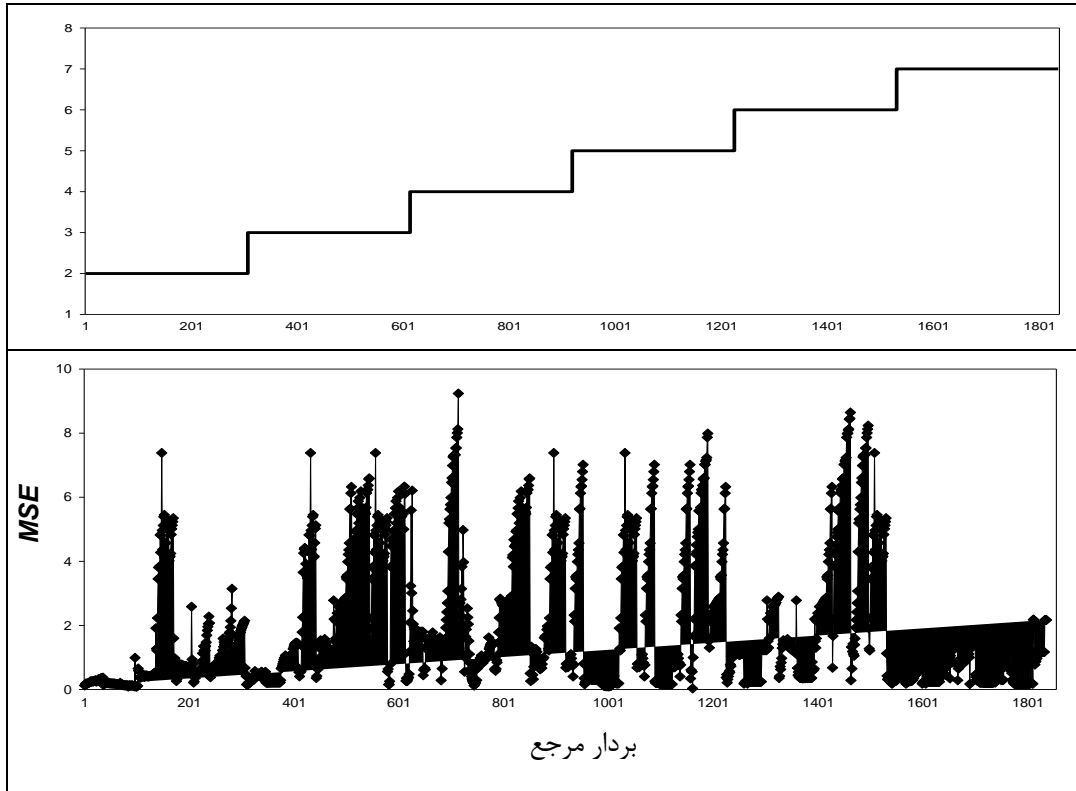
شکل (۳-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئید



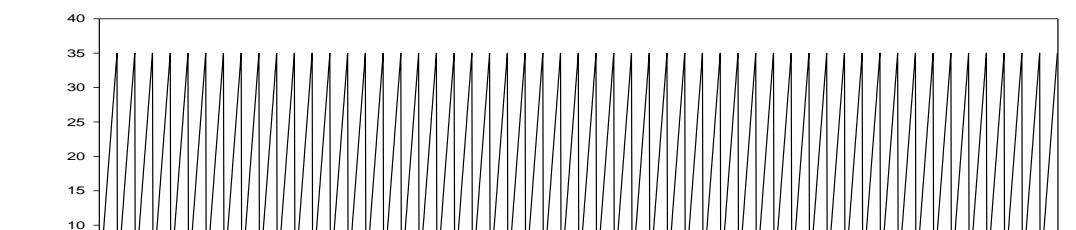
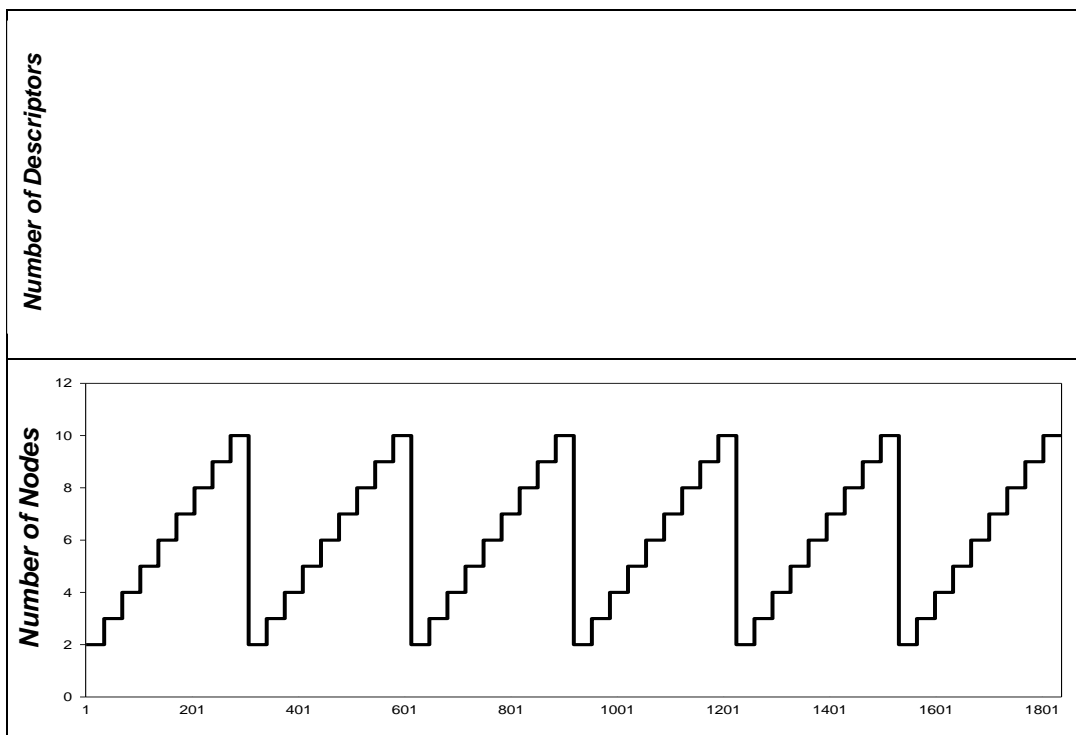
(الف)

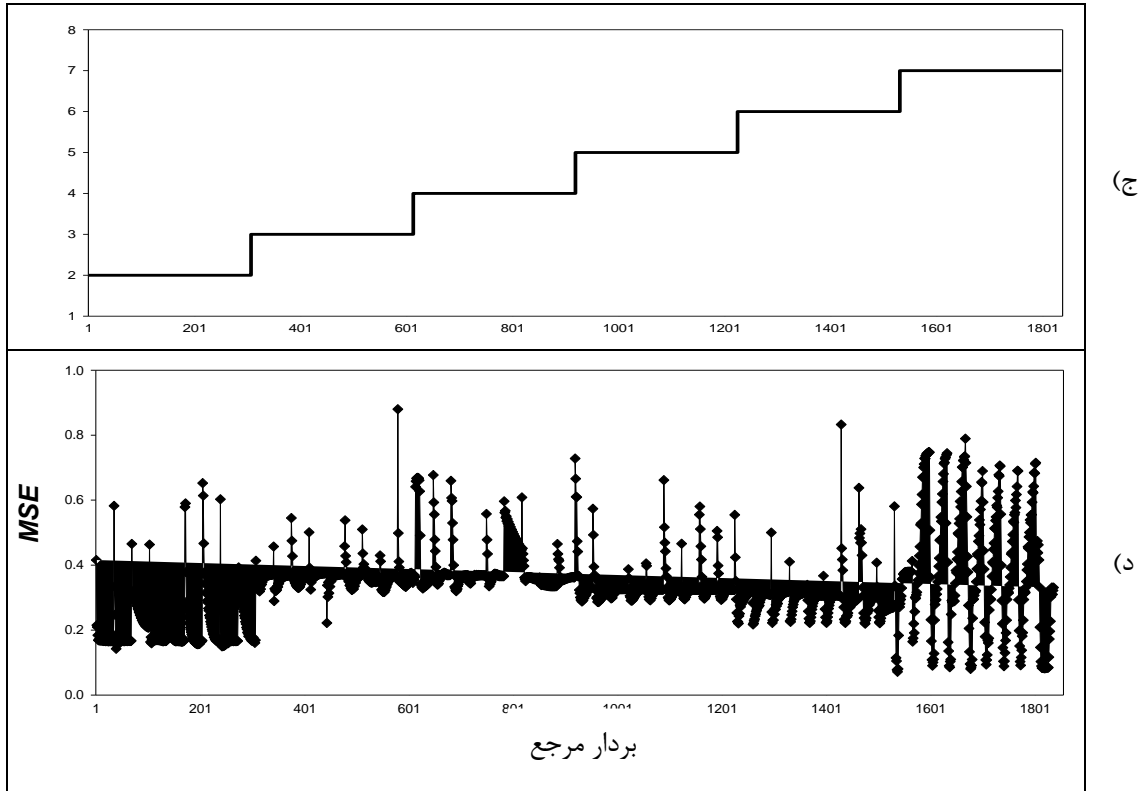
(ب)





شکل (۳-۴) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید





شکل (۳-۵) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقدار

MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تابع تانژانت سیگموئید

Epochs

با توجه به نتایج حاصله برای تعداد متغیرهای ورودی شبکه، تعداد مختلف گره‌ها، دورهای آموزش و همچنین توابع متفاوت آموزش و انتقال، بهترین شبکه‌های بدست آمده در جدول (۳-۷) گزارش شده است.

جدول (۳-۷) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا

MSE	تعداد دورهای آموزش	تعداد گره	تعداد توصیف‌کننده‌ها	تابع آموزش	تابع انتقال
۰/۰۹۷۶	۲	۹	۶	لونبرگ- مارکوارت	لگاریتم سیگموئید
۰/۰۶۵۵	۹	۸	۵	تنظیم بایزین	لگاریتم سیگموئید
۰/۰۸۴۴	۵	۳	۲	لونبرگ- مارکوارت	تانژانت سیگموئید
۰/۰۲۷۴	۷	۳	۶	تنظیم بایزین	تانژانت سیگموئید

با توجه به نتایج، تابع آموزش تنظیم بایزین نسبت به لونبرگ- مارکوارت دارای میانگین مربعات خطای کمتری است. همچنین تابع انتقال تانژانت سیگموئید به همراه تابع آموزش تنظیم

بایزین دارای خطای کمتری نسبت به لگاریتم سیگموئید می‌باشند، لذا به عنوان توابع آموزش و انتقال بهینه انتخاب گردیدند. همچنین طبق نتایج بدست آمده، تعداد ورودی های شبکه، تعداد گره در لایه‌ی پنهان و تعداد دوره‌های آموزشی نیز که بطور همزمان بهینه شدند به ترتیب برابر با ۶ متغیر ورودی، ۳ گره در لایه‌ی پنهان و تعداد دوره‌های آموزش برابر ۷ بدست آمد.

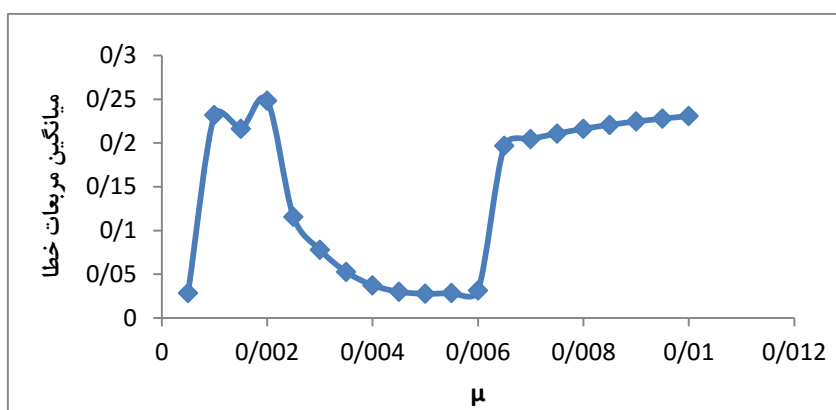
۳-۱-۵-۳- انتخاب مقدار بهینه μ

جهت بهینه کردن مقدار μ ، ساختار شبکه بهینه با ۶ متغیر ورودی، ۳ گره در لایه‌ی پنهان، تعداد دوره‌های آموزش برابر ۷، تابع انتقال تانژانت سیگموئید و تابع آموزشی تنظیم بایزین در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۰۵ تا ۰/۰۱ (حداقل میانگین مربعات خطا در این رنج مشاهده گردید) با گام‌های ۰/۰۰۰۵ تغییر داده شد و سپس برای هر مورد مقدار میانگین مربع خطای سری ارزیابی محاسبه گردید. در پایان مقدار میانگین مربع خطا بر حسب μ رسم شد. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد، همانطور که در جدول (۳-۸) و شکل (۳-۶) مشاهده می‌شود مقدار بهینه μ برابر با ۰/۰۰۵ می‌باشد (مقدار μ که در پیش فرض نرم‌افزار مطلب پیشنهاد شده است).

جدول (۳-۸) - میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ

میانگین مربع خطا	مقادیر μ	میانگین مربع خطا	مقادیر μ
------------------	--------------	------------------	--------------

۰/۰۰۰۵	۰/۰۲۸۶	۰/۰۰۶۰	۰/۰۳۱۳
۰/۰۰۱۰	۰/۲۳۱۹	۰/۰۰۶۵	۰/۱۹۶۷
۰/۰۰۱۵	۰/۲۱۶۴	۰/۰۰۷۰	۰/۲۰۴۴
۰/۰۰۲۰	۰/۲۴۸۵	۰/۰۰۷۵	۰/۲۱۰۸
۰/۰۰۲۵	۰/۱۱۵۴	۰/۰۰۸۰	۰/۲۱۶۲
۰/۰۰۳۰	۰/۰۷۸۰	۰/۰۰۸۵	۰/۲۲۰۸
۰/۰۰۳۵	۰/۰۵۲۶	۰/۰۰۹۰	۰/۲۲۴۶
۰/۰۰۴۰	۰/۰۳۷۴	۰/۰۰۹۵	۰/۲۲۷۹
۰/۰۰۴۵	۰/۰۲۹۹	۰/۰۱۰۰	۰/۲۳۰۷
۰/۰۰۵۰	۰/۰۲۷۴		
۰/۰۰۵۵	۰/۰۲۸۶		



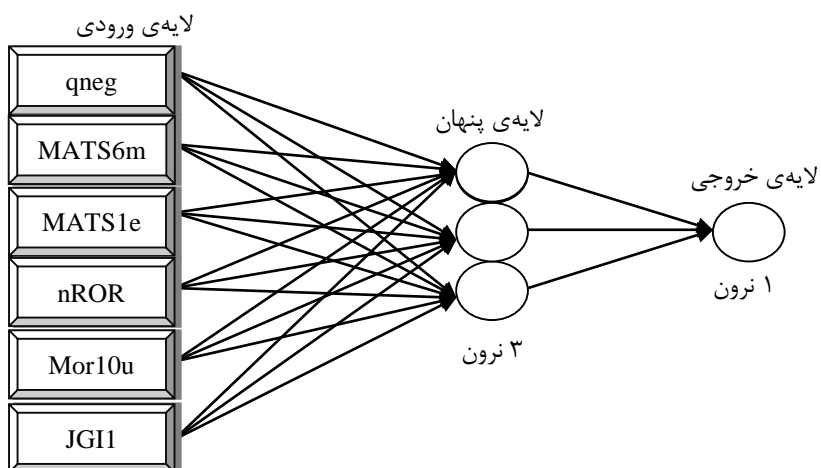
شکل (۳-۶) - نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ

۳-۱-۵-۴ - ساختار شبکه عصبی مصنوعی بهینه شده

با توجه به نتایج به دست آمده در بهینه‌سازی شبکه، شبکه عصبی سه لایه پس انتشار با

الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی، به عنوان شبکه بهینه انتخاب شد

که ساختار آن در شکل (۳-۷) نشان داده شده است.



توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۳-۹)

ارائه شده است.

جدول (۳-۹) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

تنظیم بایزین (trainbr)	تابع آموزش
تانژانت سیگموئید (tansig)	تابع انتقال لایه پنهان
purelin	تابع انتقال لایه خروجی
۳	تعداد نرون لایه پنهان (گره)
۶	تعداد متغیرهای ورودی (توصیف کننده‌ها)
۷	تعداد دوره‌های آموزش
۰/۰۰۵	پارامتر μ

۳-۱-۶ - مدل سازی شبکه عصبی مصنوعی با استفاده از توصیف کننده‌های

انتخاب شده توسط الگوریتم ژنتیک

۳-۱-۶-۱ - بهینه سازی پارامترهای شبکه

در این بخش نیز برای بهینه سازی پارامترهای شبکه، شبکه با ورودی‌های از ۲ تا ۷ ایجاد شد و با الگوریتم‌های آموزشی تنظیم بایزین و لونبرگ-مارکوارت با تعداد متفاوت گره‌های لایه پنهان (۲ تا ۱۰ گره) آموزش داده شد. در حالی که برای انتخاب بهترین تابع انتقال در لایه پنهان از توابع تانژانت سیگموئیدی و لگاریتم سیگموئیدی به عنوان تابع انتقال استفاده شد. در تمامی این شبکه‌ها تعداد دوره‌های آموزش از ۲ تا ۳۵ تغییر داده شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده گردید.

۳-۱-۶-۲- بهینه‌سازی همزمان تعداد متغیرهای ورودی شبکه، تابع آموزش، نوع تابع

انتقال، تعداد گره‌های لایه‌های پنهان و تعداد دوره‌های آموزشی

جهت بهینه‌سازی همزمان تمام پارامترها، در مجموع ۷۳۴۴ شبکه مورد ارزیابی قرار گرفته

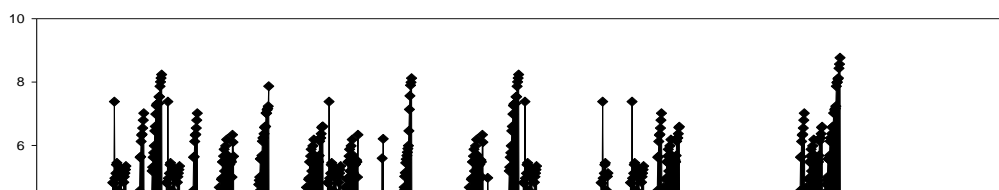
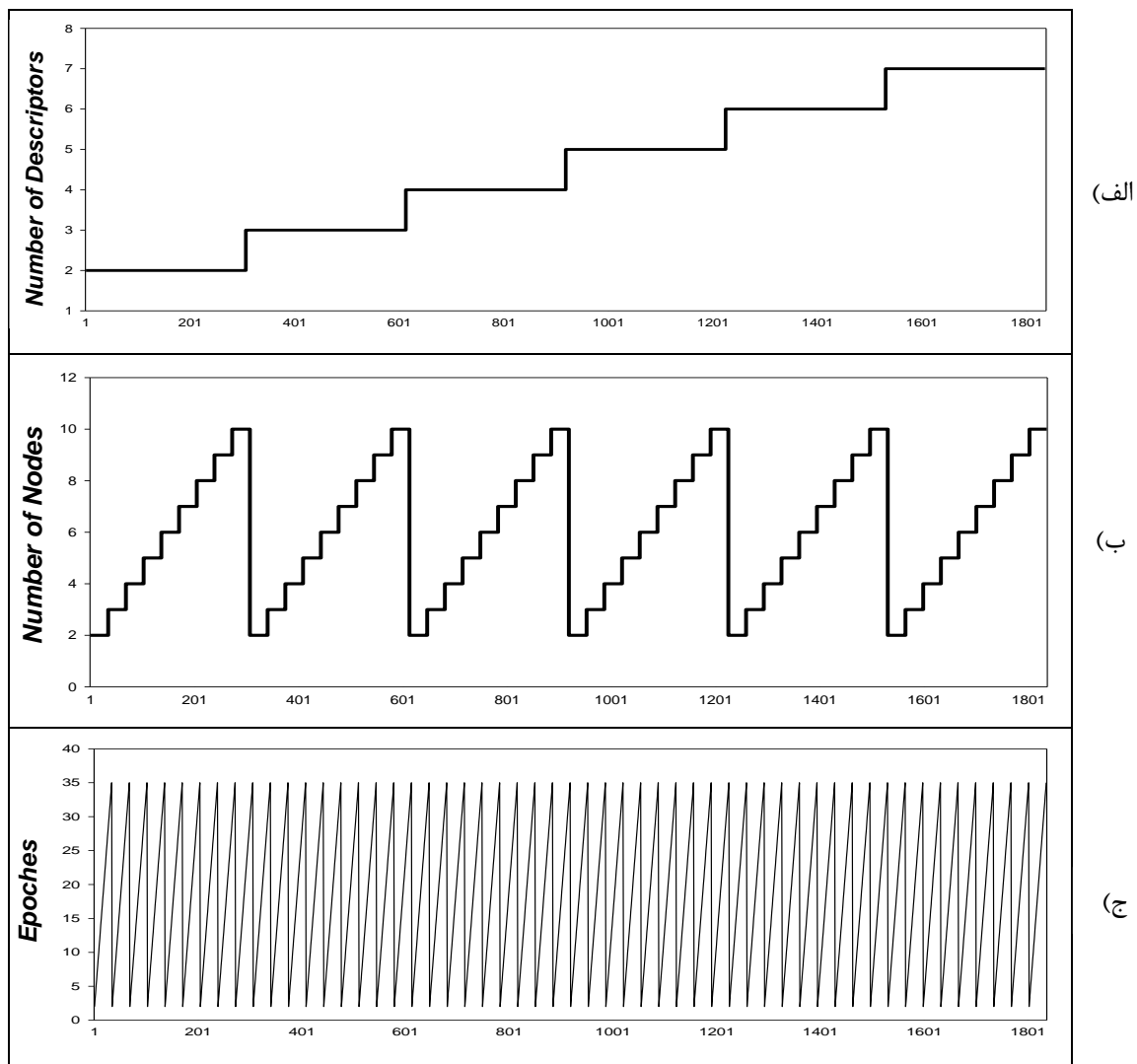
است که در زیر به تعداد تمام موارد مورد بررسی اشاره شده است.

۷۳۴۴ = دوره‌های آموزش (۳۴) × گره (۹) × تابع انتقال (۲) × تابع آموزش (۲) × توصیف‌کننده (۶)

بخشی از روند تغییرات پارامترهای شبکه در حین بهینه‌سازی همزمان پارامترها به همراه

مقادیر MSE به دست آمده به صورت نموداری بر حسب یک بردار مرجع فرضی در شکل‌های (۳-۸)،

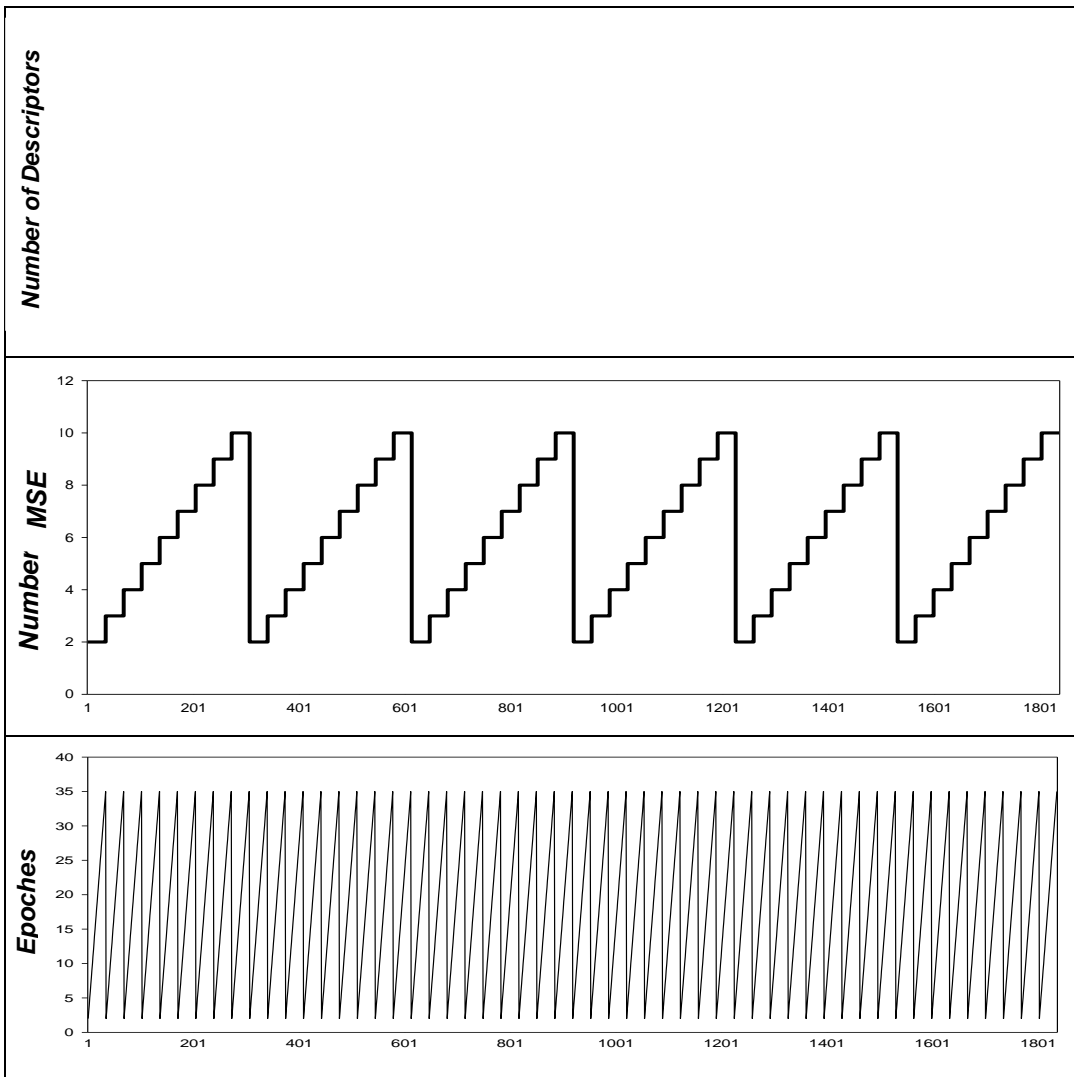
(۳-۹)، (۳-۱۰) و (۳-۱۱) آمده است.





(د)

شکل (۳-۸) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید

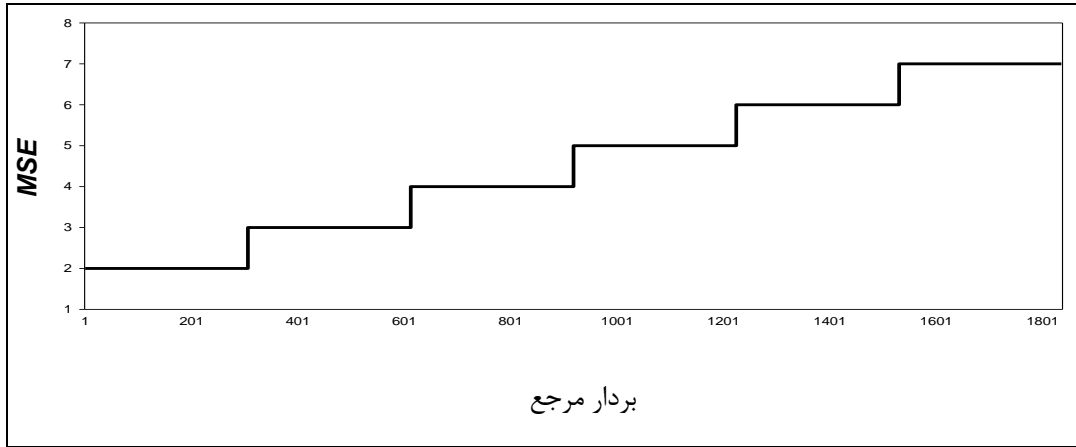


(الف)

(ب)

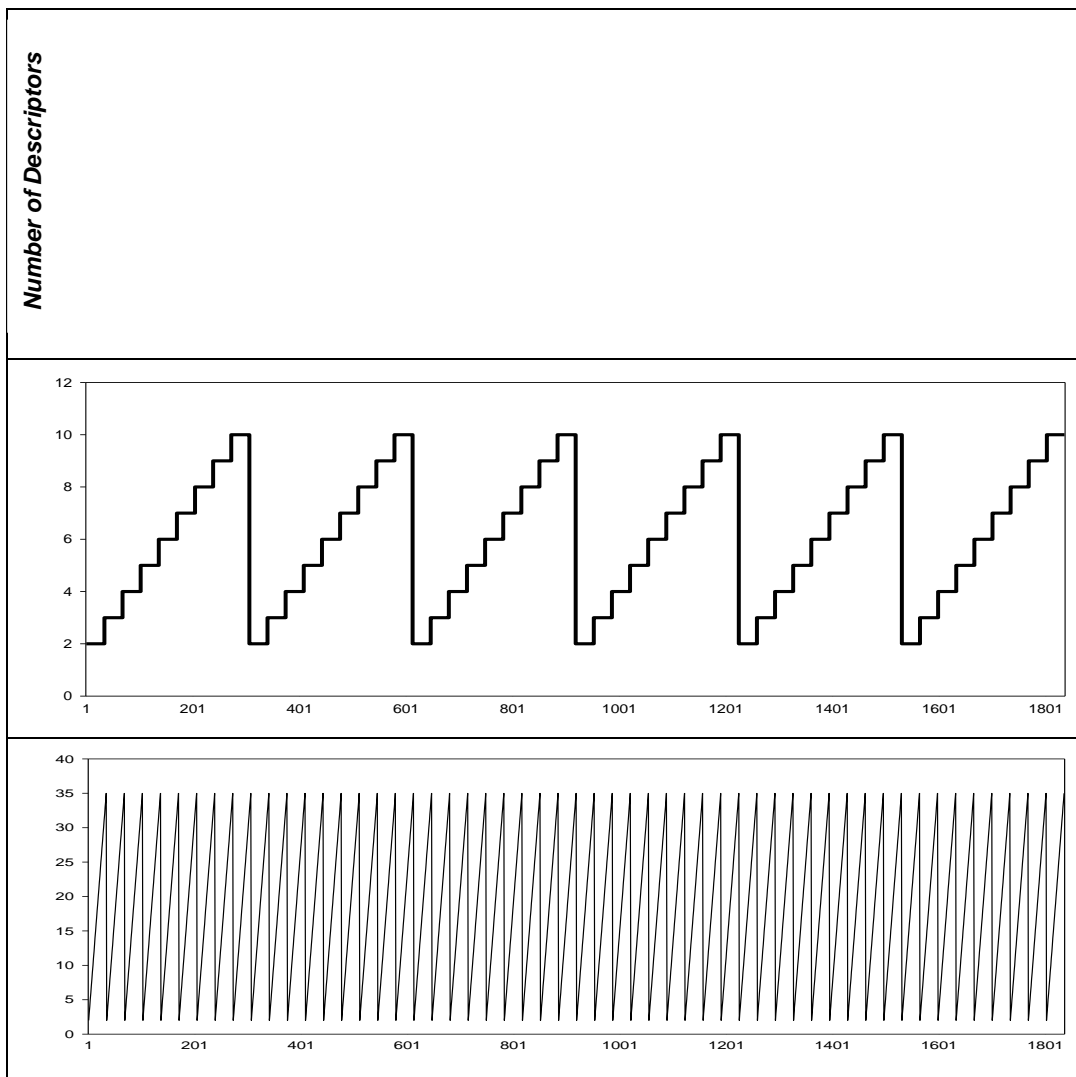
(ج)





(د)

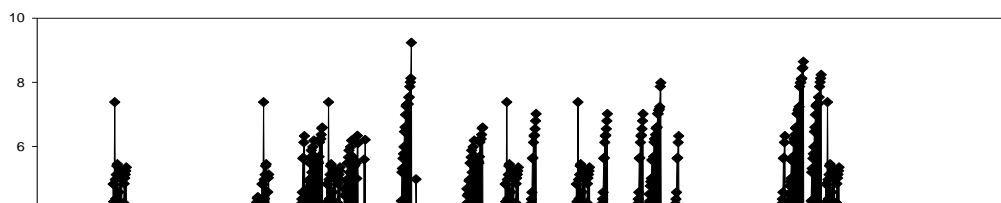
شکل (۳-۹) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئید

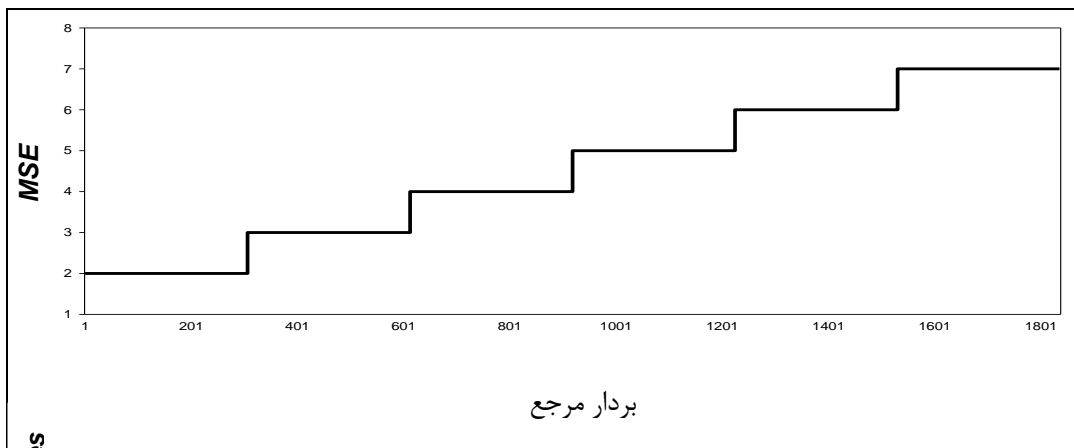


الف)

ب)

ج)





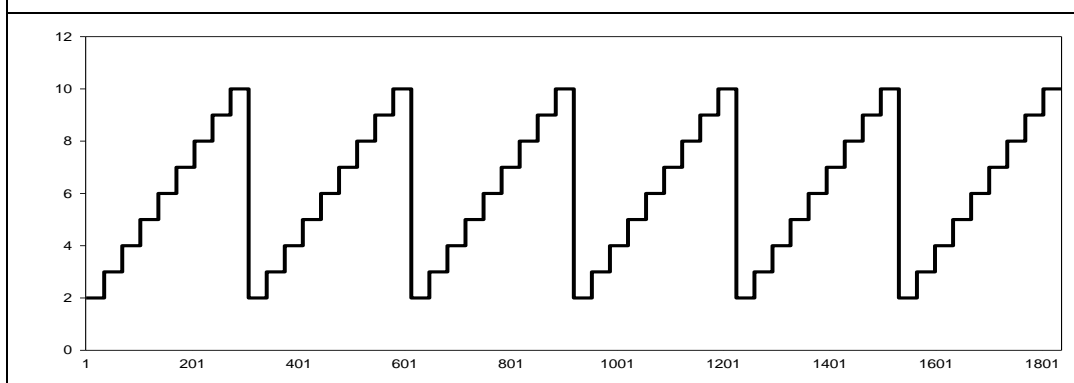
(د)

Number of Nodes

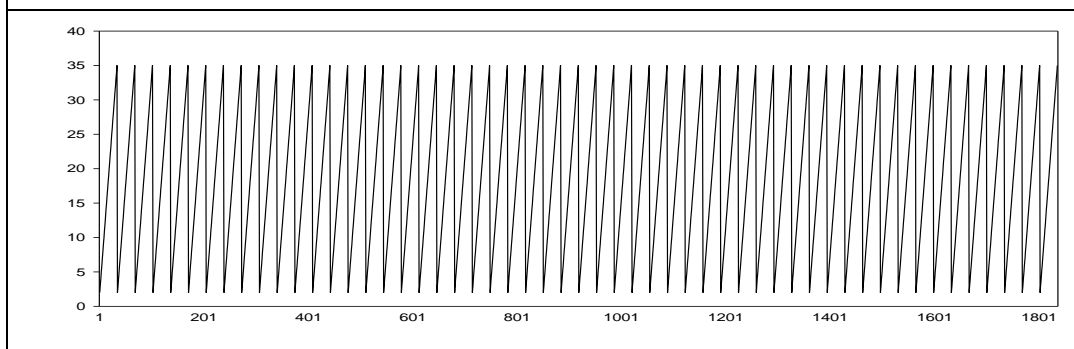
شکل (۳-۱۰) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقدار MSE برای تابع آموزش لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید



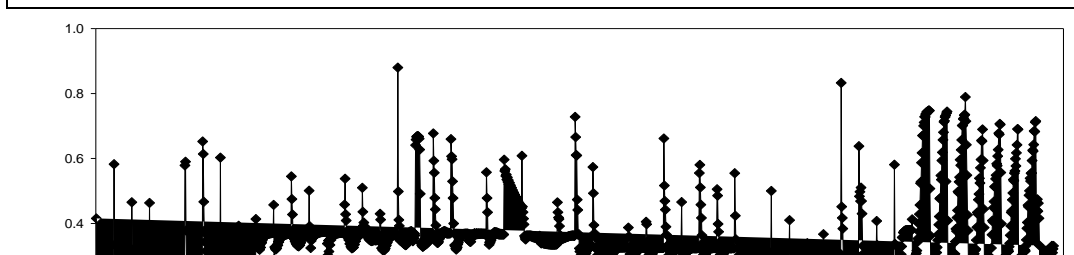
(الف)

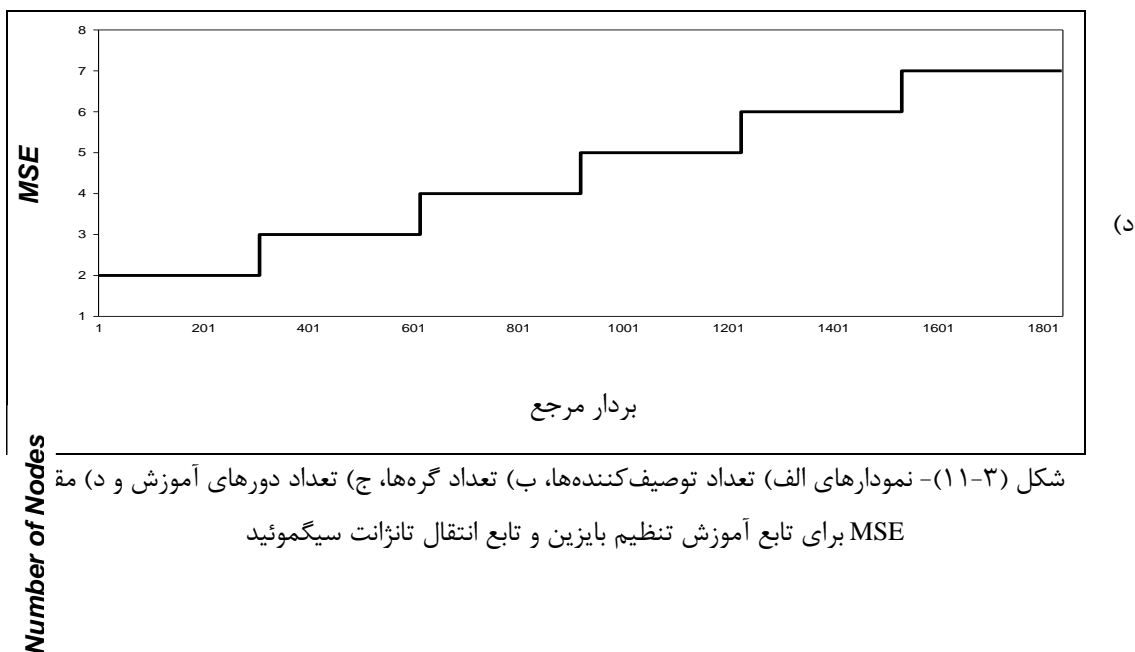


(ب)



(ج)





شکل (۳-۱۱) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقدار MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئید

جدول (۳-۱۰) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا

Epochs	MSE	تعداد دوره‌های آموزش	تعداد گره	تعداد توصیف‌کننده‌ها	تابع آموزش	تابع انتقال
	۰/۰۹۷۰	۱۱	۴	۲	لونیبرگ - مارکوارت	لگاریتم سیگموئید
۰/۰۶۸۰	۱۹	۶	۷	تنظیم بایزین	لگاریتم سیگموئید	
۰/۰۸۹۵	۳۳	۴	۲	لونیبرگ - مارکوارت	تانژانت سیگموئید	
۰/۰۷۱۷	۸	۲	۷	تنظیم بایزین	تانژانت سیگموئید	

با توجه به نتایج، تابع آموزش تنظیم بایزین نسبت به لونیبرگ- مارکوارت دارای میانگین مربعات خطای کمتری است. همچنین تابع انتقال لگاریتم سیگموئید به همراه تابع آموزش تنظیم بایزین دارای خطای کمتری نسبت به تانژانت سیگموئید می‌باشند، لذا به عنوان توابع آموزش و انتقال بهینه انتخاب گردیدند.. همچنین طبق نتایج بدست آمده، تعداد ورودی‌های شبکه، تعداد گره در لایه‌ی پنهان و تعداد دوره‌های آموزشی نیز که بطور همزمان بهینه شدند به ترتیب برابر با ۷ متغیر ورودی، ۶ گره در لایه‌ی پنهان و تعداد دوره‌های آموزش برابر ۱۹ بدست آمد.

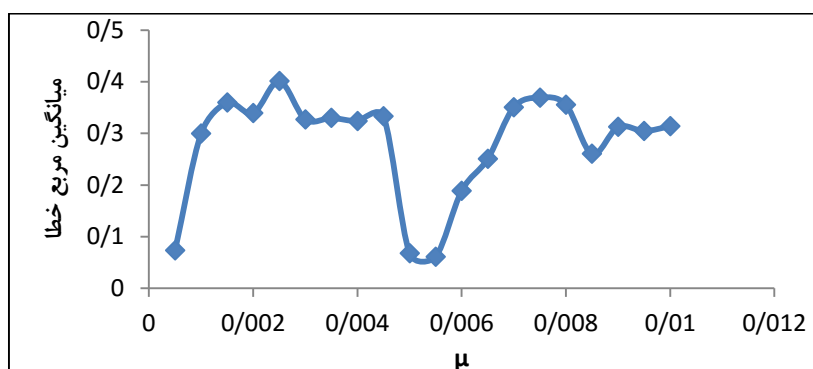
۳-۱-۶-۳- بهینه‌سازی پارامتر μ

جهت یافتن مقدار بهینه μ ، ساختار شبکه بهینه با ۷ متغیر ورودی، ۶ گره در لایه‌ی پنهان، تعداد دوره‌های آموزش برابر ۱۹، تابع انتقال لگاریتم سیگموئید و تابع آموزشی تنظیم بایزین در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۰۵ تا ۰/۰۱ (حداقل میانگین مربعات خطا در این رنج مشاهده گردید) با گام‌های ۰/۰۰۰۵ تغییر داده شد و آنگاه برای هر مورد، مقدار میانگین مربع خطای سری ارزیابی محاسبه گردید. سرانجام مقدار میانگین مربع خطا بر حسب μ ترسیم گردید. نقطه‌ای که کمترین خطا را برای سری آموزش داشت، به عنوان مقدار بهینه انتخاب شد که طبق جدول (۳-۱۱) و شکل (۳-۱۲) مقدار بهینه ۰/۰۰۵۵ به دست آمد.

جدول (۳-۱۱)- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ

مقادیر μ	میانگین مربع خطا	مقادیر μ	میانگین مربع خطا
۰/۰۰۰۵	۰/۰۷۳۱	۰/۰۰۶۰	۰/۱۸۸۸
۰/۰۰۱۰	۰/۲۹۹۵	۰/۰۰۶۵	۰/۲۵۰۷
۰/۰۰۱۵	۰/۳۵۹۴	۰/۰۰۷۰	۰/۳۵۰۴
۰/۰۰۲۰	۰/۳۳۹۲	۰/۰۰۷۵	۰/۳۶۹۲
۰/۰۰۲۵	۰/۴۰۱۰	۰/۰۰۸۰	۰/۳۵۵۳

۰/۰۰۳۰	۰/۳۲۷۰	۰/۰۰۸۵	۰/۲۶۰۳
۰/۰۰۳۵	۰/۳۲۹۸	۰/۰۰۹۰	۰/۳۱۲۵
۰/۰۰۴۰	۰/۳۲۳۹	۰/۰۰۹۵	۰/۳۰۴۷
۰/۰۰۴۵	۰/۳۳۲۸	۰/۰۱۰۰	۰/۳۱۳۸
۰/۰۰۵۰	۰/۰۶۸۰		
۰/۰۰۵۵	۰/۰۶۱۳		



شکل (۳-۱۲)- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ

۳-۱-۶-۴- ساختار شبکه عصبی مصنوعی بهینه شده

باتوجه به نتایج به دست آمده در بهینه‌سازی شبکه، مشخصات شبکه بهینه در جدول (۳-۱۲)

ارائه شده است.

جدول (۳-۱۲)- مشخصات شبکه بهینه

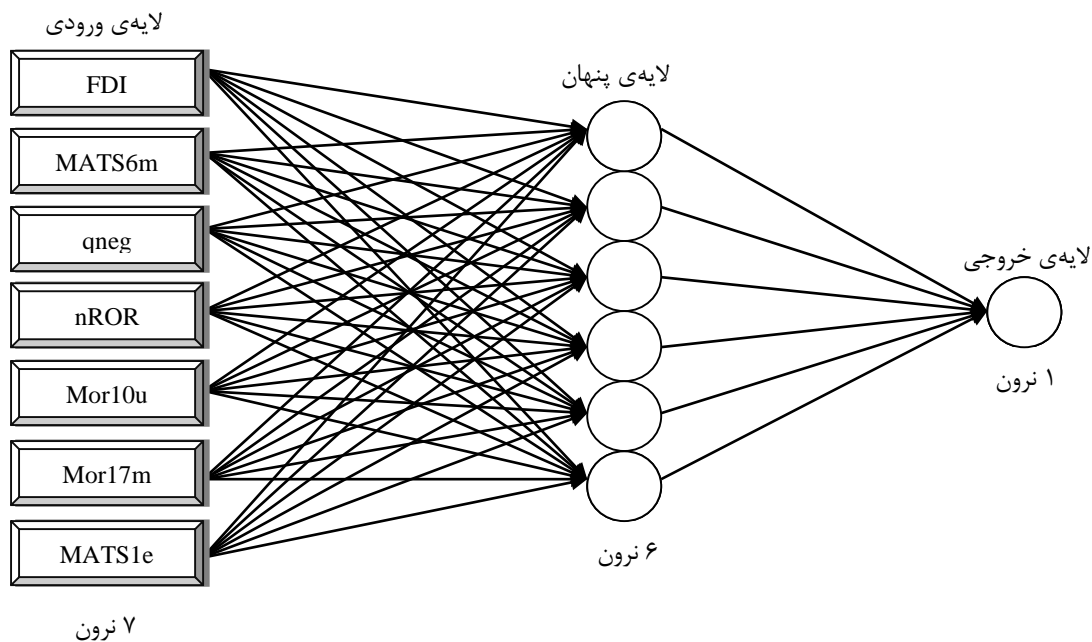
تابع آموزش	تنظیم بایزین (trainbr)
تابع انتقال لایه پنهان	لگاریتم سیگموئید (logsig)
تابع انتقال لایه خروجی	purelin
تعداد نرون لایه پنهان (گره)	۶
تعداد متغیرهای ورودی (توصیف‌کننده‌ها)	۷
تعداد دوره‌های آموزش	۱۹
پارامتر μ	۰/۰۰۵۵

۳-۱-۶-۵- معماری شبکه عصبی بهینه شده

با توجه به روند بهینه‌سازی که در قسمت‌های قبلی ذکر شد، شبکه‌ی عصبی مصنوعی بهینه

شده دارای ۷ توصیف‌کننده، ۶ گره در لایه پنهان، تابع آموزشی تنظیم بایزین و تابع انتقال لگاریتم

سیگموئیدی است. شکل (۳-۱۳) ساختار شبکه عصبی بهینه‌ی نهایی را نشان می‌دهد.



شکل (۳-۱۳) - ساختار شبکه عصبی به دست آمده پس از بهینه‌سازی

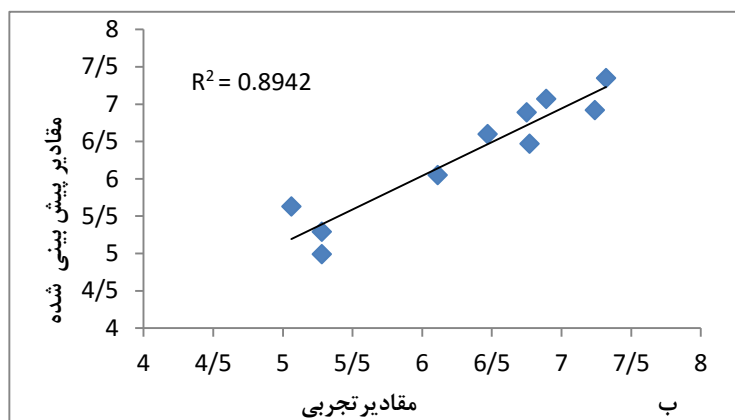
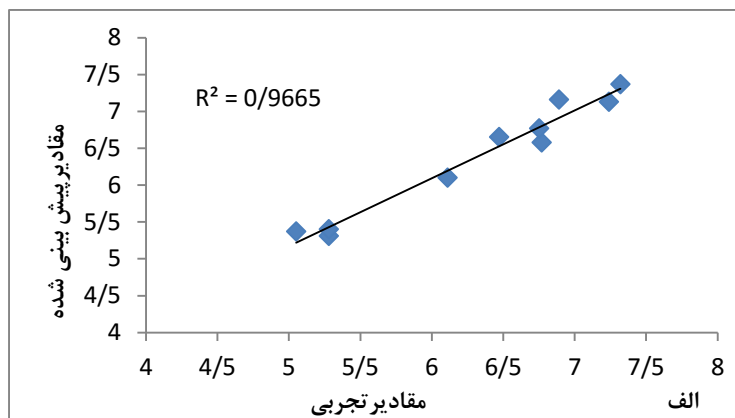
۳-۱-۷- ارزیابی مدل‌های SR-ANN و GA-ANN

۳-۱-۷-۱- ارزیابی مدل‌های SR-ANN و GA-ANN با استفاده از سری‌های ارزیابی و تست

به منظور مطالعه‌ی قدرت پیش‌بینی مدل‌های SR-ANN و GA-ANN، سری‌های ارزیابی و تست مورد استفاده قرار گرفتند. بدین ترتیب که کل داده‌ها به سه قسمت با نسبت‌های ۲۰٪، ۲۰٪ و ۶۰٪ تحت عنوان سری‌های ارزیابی، تست و آموزش تقسیم شدند. در این راستا، با استفاده از مدل‌های بدست آمده مقادیر pIC_{50} این ترکیبات پیش‌بینی شدند. نتایج حاصل از این بررسی که مجذور ضریب همبستگی را برای سری ارزیابی و تست (شکل‌های (۳-۱۴) و (۳-۱۵)) توسط دو مدل SR-ANN و GA-ANN نشان می‌دهد، بیانگر توانایی بالاتر روش SR-ANN نسبت به روش GA-ANN در پیش‌بینی فعالیت دارویی این ترکیبات می‌باشد.

جدول (۳-۱۳) - نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری ارزیابی

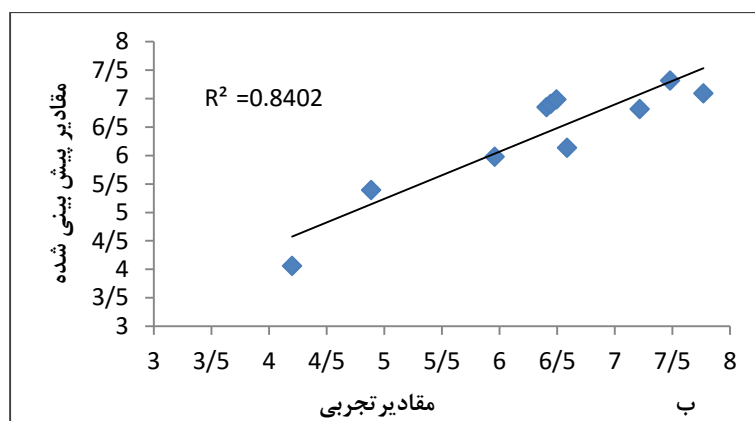
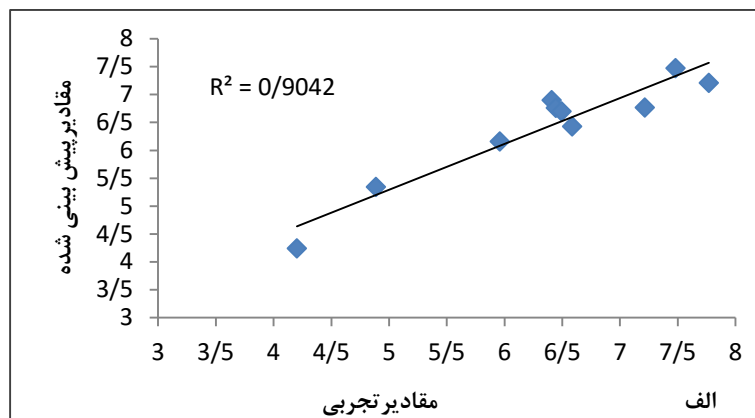
شماره ترکیب	مقدار تجربی	pIC ₅₀			
		مقدار پیش بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۲	۶/۱۱	۶/۱۰	۶/۰۵	-۰/۱۶	-۰/۹۸
۶	۵/۲۸	۵/۳۱	۴/۹۹	۰/۵۷	-۵/۴۹
۸	۵/۰۵	۵/۳۷	۵/۶۳	۶/۳۴	۱۱/۴۹
۱۵	۷/۲۴	۷/۱۳	۶/۹۳	-۱/۵۲	-۴/۲۸
۱۸	۶/۷۷	۶/۵۸	۶/۴۷	-۲/۸۱	-۴/۴۳
۲۳	۶/۷۵	۶/۷۷	۶/۸۹	۰/۳۰	۲/۰۷
۲۹	۶/۴۷	۶/۶۵	۶/۶۰	۲/۷۸	۲/۰۱
۳۷	۷/۳۲	۷/۳۷	۷/۳۵	۰/۶۸	۰/۴۱
۴۴	۵/۲۸	۵/۴۰	۵/۲۹	۲/۲۷	۰/۱۹
۴۷	۶/۸۹	۷/۱۶	۷/۰۷	۳/۹۲	۲/۶۱



شکل (۳-۱۴)- نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری ارزیابی
الف: با مدل SR-ANN ب: با مدل GA-ANN

جدول (۳-۱۴)- نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری تست

شماره ترکیب	pIC_{50}				
	مقدار تجربی	مقدار پیش‌بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۴	۴/۸۹	۵/۳۴	۵/۳۸	۹/۲۰	۱۰/۰۲
۱۱	۴/۲۰	۴/۲۴	۴/۰۶	۰/۹۵	-۳/۳۳
۱۶	۷/۲۱	۶/۷۷	۶/۸۱	-۶/۱۰	-۵/۵۵
۲۱	۶/۵۹	۶/۴۳	۶/۱۳	-۲/۴۳	-۶/۹۸
۲۴	۶/۴۱	۶/۹۰	۶/۸۵	۷/۶۴	۶/۸۶
۲۸	۶/۴۴	۶/۷۶	۶/۹۱	۴/۹۷	۷/۳۰
۳۰	۶/۵۰	۶/۷۰	۶/۹۸	۳/۰۸	۷/۳۸
۳۴	۵/۹۶	۶/۱۶	۵/۹۸	۳/۳۶	۰/۳۴
۳۸	۷/۴۸	۷/۵۲	۷/۳۲	-۰/۵۳	-۲/۱۴
۴۲	۷/۷۷	۷/۲۱	۷/۰۹	-۷/۲۱	-۸/۷۵



شکل (۳-۱۵)- نمودار مقادیر پیش‌بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری تست الف: با مدل SR-ANN ب: با مدل GA-ANN

۳-۱-۷-۲- ارزیابی مدل‌های SR-ANN و GA-ANN توسط روش رد مرحله‌ای تک‌تک

به منظور ارزیابی بیشتر مدل‌های SR-ANN و GA-ANN، از روش رد مرحله‌ای تک‌تک برای پیش‌بینی مقدار فعالیت دارویی استفاده شد که نتایج حاصله در جدول (۳-۱۵) ارائه شده است. در روش رد مرحله‌ای تک‌تک یک ترکیب به عنوان سری تست و بقیه‌ی ترکیبات به عنوان سری آموزش در نظر گرفته شده و شبکه‌ی بهینه مقدار pIC_{50} را برای تست پیش‌بینی می‌کند. پس از رسم منحنی مقدار تجربی pIC_{50} بر حسب مقدار پیش‌بینی شده‌ی آن (شکل (۳-۱۶))، مقادیر R^2 برای مدل‌های SR-ANN و GA-ANN به ترتیب ۰/۹۱۱۵ و ۰/۸۵۲۳ بدست آمد. با توجه به مقادیر R^2 بدست آمده توسط روش رد مرحله‌ای تک‌تک، برتری مدل SR-ANN نسبت به GA-ANN محسوس و مشهود می‌باشد.

همچنین شکل (۳-۱۷) نمودار باقیمانده‌ها بر حسب مقادیر تجربی pIC_{50} ترکیبات مورد مطالعه را نشان می‌دهد. اگر باقیمانده‌ها به طور یکنواخت حول محور افقی پراکنده باشند، نشان دهنده عدم وجود خطای سیستماتیک در مدل‌های ارائه شده می‌باشد و همانطور که از این شکل‌ها مشخص است پراکندگی نقاط در دو طرف محور افقی بصورت متقارن می‌باشد.

جدول (۳-۱۵) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-ANN و GA-ANN با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها

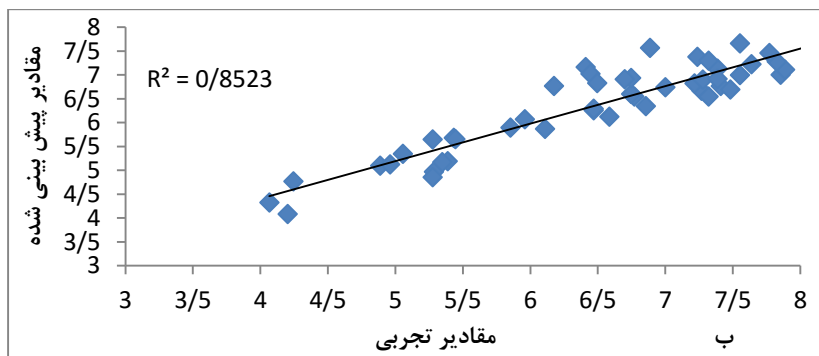
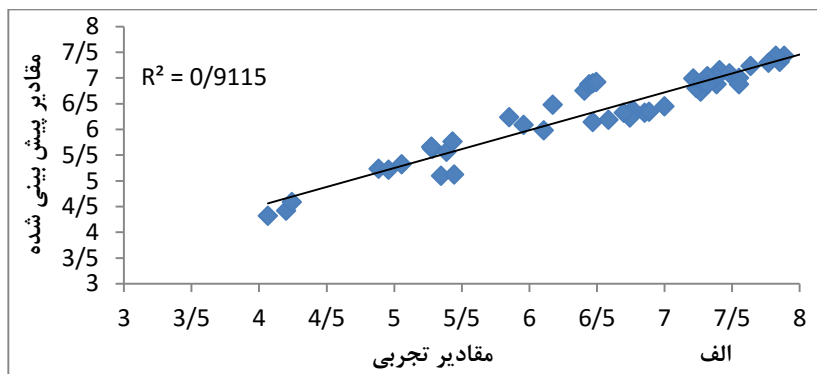
شماره ترکیب	pIC_{50}				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۱	۷/۲۷	۶/۷۳	۶/۶۷	-۷/۴۳	-۸/۲۵
۲	۶/۱۱	۵/۹۸	۵/۸۷	-۲/۱۳	-۳/۹۳
۳	۵/۳۹	۵/۵۷	۵/۱۹	۳/۳۴	-۳/۷۱
۴	۴/۸۹	۵/۲۳	۵/۰۹	۶/۹۵	۴/۰۹
۵	۵/۲۸	۵/۶۲	۴/۹۷	۶/۴۴	-۵/۸۷
۶	۵/۲۸	۵/۶۵	۴/۸۵	۷/۰۱	-۸/۱۴
۷	۴/۹۶	۵/۲۱	۵/۱۲	۵/۰۴	۳/۲۳
۸	۵/۰۵	۵/۳۲	۵/۳۵	۵/۳۵	۵/۹۴
۹	۵/۳۵	۵/۱۰	۵/۱۷	-۴/۶۷	-۳/۳۶
۱۰	۵/۴۳	۵/۷۶	۵/۶۸	۶/۰۸	۴/۶۰
۱۱	۴/۲۰	۴/۴۲	۴/۰۸	۵/۲۴	-۲/۸۶

۱۲	۵/۴۴	۵/۱۲	۵/۶۵	-۵/۸۸	۳/۹۶
۱۳	۴/۰۷	۴/۳۲	۴/۳۲	۶/۱۴	۶/۱۴
۱۴	۸/۱۵	۷/۶۵	۷/۴۳	-۶/۱۳	-۸/۸۳
۱۵	۷/۲۴	۷/۰۲	۷/۳۸	-۳/۰۴	۱/۹۳
۱۶	۷/۲۱	۶/۵۳	۶/۸۲	-۹/۴۳	-۵/۴۱
۱۷	۶/۸۵	۶/۳۲	۶/۳۴	-۷/۷۴	-۷/۴۵
۱۸	۶/۷۷	۶/۳۷	۶/۵۵	-۵/۹۱	-۳/۲۵
۱۹	۷/۸۵	۷/۳۲	۷/۰۰	-۶/۷۵	-۱۰/۸۳
۲۰	۷/۰۰	۶/۴۵	۶/۷۴	-۷/۸۶	-۳/۷۱
۲۱	۶/۵۹	۶/۱۹	۶/۱۲	-۶/۰۷	-۷/۱۳
۲۲	۶/۷۵	۶/۳۳	۶/۶۰	-۶/۲۲	-۲/۲۲
۲۳	۶/۷۵	۶/۲۳	۶/۹۳	-۷/۷۰	۲/۶۷
۲۴	۶/۴۱	۶/۷۵	۷/۱۶	۵/۳۰	۱۱/۷۰

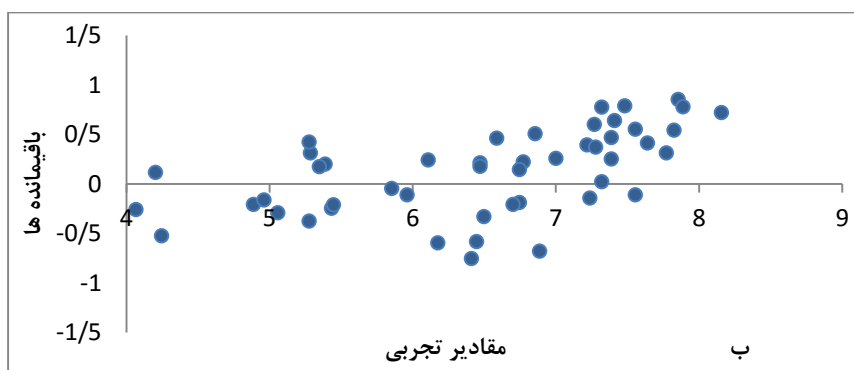
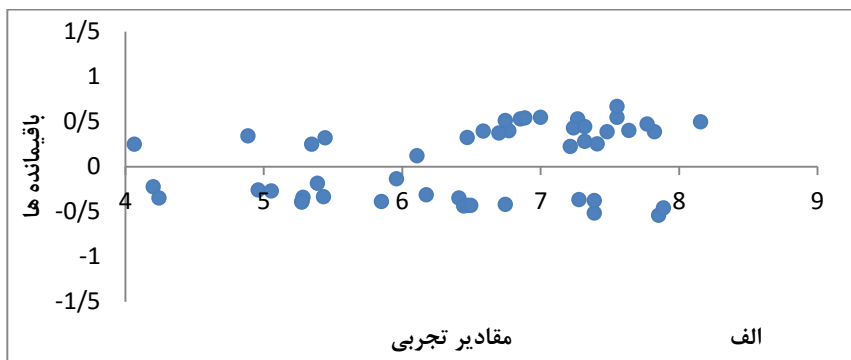
ادامه جدول (۳-۱۵)

شماره ترکیب	pIC ₅₀				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۲۵	۷/۲۸	۶/۹۱	۶/۹۰	-۵/۰۸	-۵/۲۲
۲۶	۷/۴۱	۷/۱۵	۶/۷۷	-۳/۵۱	-۸/۶۴
۲۷	۷/۳۹	۷/۰۸	۷/۱۳	-۴/۱۹	-۳/۵۲
۲۸	۶/۴۴	۶/۸۸	۷/۰۲	۶/۸۳	۹/۰۲
۲۹	۶/۴۷	۶/۹۰	۶/۲۵	۶/۶۵	-۳/۴۰
۳۰	۶/۵۰	۶/۹۲	۶/۸۲	۶/۴۶	۴/۹۲
۳۱	۷/۳۹	۷/۰۱	۶/۹۲	-۵/۱۴	-۶/۳۶
۳۲	۷/۳۲	۷/۰۴	۶/۵۴	-۳/۸۳	-۱۰/۶۶
۳۳	۶/۱۷	۶/۴۸	۶/۷۷	۵/۰۲	۹/۷۲
۳۴	۵/۹۶	۶/۰۹	۶/۰۷	۲/۱۸	۱/۸۵
۳۵	۶/۴۷	۶/۱۴	۶/۲۹	-۵/۱۰	-۲/۷۸
۳۶	۷/۶۴	۷/۲۳	۷/۲۲	-۵/۳۷	-۵/۵۰

۳۷	۷/۳۲	۶/۸۷	۷/۳۰	-۶/۱۵	-۰/۲۷
۳۸	۷/۴۸	۷/۰۹	۶/۶۹	-۵/۲۱	-۱۰/۵۶
۳۹	۷/۸۹	۷/۴۳	۷/۱۱	-۵/۸۳	-۹/۸۹
۴۰	۷/۸۲	۷/۴۳	۷/۲۸	-۴/۹۹	-۶/۹۱
۴۱	۷/۵۵	۷/۰۰	۷/۶۶	-۷/۲۸	۱/۴۶
۴۲	۷/۷۷	۷/۲۹	۷/۴۶	-۶/۱۸	-۳/۹۹
۴۳	۷/۵۵	۶/۸۸	۷/۰۰	-۸/۸۷	-۷/۲۸
۴۴	۵/۲۸	۵/۳۴	۵/۶۵	۱/۱۴	۷/۰۱
۴۵	۵/۸۵	۶/۲۴	۵/۸۹	۶/۶۷	۰/۶۸
۴۶	۴/۲۴	۴/۵۹	۴/۷۷	۸/۲۵	۱۲/۵۰
۴۷	۶/۸۹	۶/۳۴	۷/۵۷	-۷/۹۸	۹/۸۷
۴۸	۶/۷۰	۶/۳۲	۶/۹۰	-۵/۶۷	۲/۹۹



شکل (۳-۱۶) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی IC_{50} به روش رد مرحله‌ای تک تک برای کل داده‌ها الف: با مدل SR-ANN ب: با مدل GA-ANN



شکل (۳-۱۷) - نمودار باقیمانده‌ها بر حسب مقدار تجربی pIC_{50} برای کل داده‌ها
الف: با مدل SR-ANN؛ ب: با مدل GA-ANN

۳-۱-۸- ارزیابی مدل‌های برتر SR-ANN و GA-ANN با استفاده از پارامترهای

آماري

مطابق جدول (۳-۱۶) هفت پارامتر آماری، جهت ارزیابی توانایی پیش‌بینی مدل‌های ساخته شده به روش‌های SR-ANN و GA-ANN به کار گرفته شد. مقادیر این پارامترها نشان می‌دهد که مدل‌های ارائه شده از قدرت پیش‌بینی مناسبی برخوردار هستند. نتایج همچنین گویای برتری نسبی روش SR-ANN نسبت به GA-ANN در پیش‌بینی سری داده‌ها می‌باشد.

جدول (۳-۱۶) - پارامترهای آماری برای مدل‌های SR-ANN و GA-ANN

پارامتر	سری تست (N=10)		سری ارزیابی (N=10)		کل داده‌ها (N=48)	
	SR-ANN	GA-ANN	SR-ANN	GA-ANN	SR-ANN	GA-ANN
MAE	۰/۲۹۰۰	۰/۳۷۴۰	۰/۱۳۰۰	۰/۲۰۳۰	۰/۳۷۶۵	۰/۳۸۴۴
MSE	۰/۱۱۶۱	۰/۱۷۶۵	۰/۰۲۷۴	۰/۰۶۸۰	۰/۱۵۹۰	۰/۱۸۹۹

PRESS	۱/۱۶۱۰	۱/۷۶۵۰	۰/۲۷۴۰	۰/۶۸۰۰	۷/۶۳۲۰	۹/۱۱۵۲
SEP	۰/۳۴۰۷	۰/۴۲۰۱	۰/۱۶۵۵	۰/۲۶۰۸	۰/۳۹۸۷	۰/۴۳۵۸
R ²	۰/۹۰۴۲	۰/۸۴۰۲	۰/۹۶۶۵	۰/۸۹۴۲	۰/۹۱۱۵	۰/۸۵۲۳
REP (%)	۵/۳۶۵۴	۶/۶۱۵۷	۲/۶۱۸۷	۴/۱۲۶۶	۶/۱۵۲۷	۶/۷۲۵۳
MRE	۴/۵۴۷۰	۵/۸۶۶۰	۲/۱۳۵۰	۳/۳۹۶۰	۵/۶۸۰۲	۵/۷۱۰۴

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

در تکنیک استفاده از آزمون Y-تصادفی، مقادیر تصادفی از متغیر وابسته تولید و ارزیابی مدل با هدف بررسی هرگونه ارتباط تصادفی بین داده‌ها انجام می‌گردد. مدل QSAR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر وابسته توسعه یافت. اگر مدل اصلی هیچ گونه ارتباط تصادفی نداشته باشد، تفاوت قابل توجهی بین مقدار ضریب تعیین مدل اصلی و مدل QSAR که با پاسخ‌های تصادفی توسعه یافته، وجود دارد. نتایج حاصل از چندین بار اجرای آزمون Y-تصادفی در جداول (۳-۱۷) و (۳-۱۸) نشان داده شده است. مقادیر کوچک ضریب تعیین (R^2) بیانگر عدم وجود ارتباط شانس‌ی یا وابستگی ساختاری به سری آموزش در مدل‌های توسعه یافته توسط شبکه می‌باشد.

جدول (۳-۱۷)- مقادیر R^2 برای سری تست و ارزیابی مدل SR-ANN با استفاده از آزمون Y-تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2 سری تست	۰/۰۰۳	۰/۰۲۱	۰/۱۴۰	۰/۰۸۸	۰/۰۳۵	۰/۰۵۰	۰/۰۴۵	۰/۰۱۶	۰/۰۰۰	۰/۰۲۱
R^2 سری ارزیابی	۰/۱۳۰	۰/۰۸۷	۰/۰۳۵	۰/۰۱۸	۰/۰۰۱	۰/۱۸۷	۰/۰۵۴	۰/۰۳۲	۰/۰۰۵	۰/۰۴۰

جدول (۳-۱۸)- مقادیر R^2 برای سری تست و ارزیابی مدل GA-ANN با استفاده از آزمون Y-تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2 سری تست	۰/۰۴۲	۰/۰۲۰	۰/۱۸۰	۰/۰۹۴	۰/۰۳۹	۰/۰۳۰	۰/۰۳۲	۰/۰۶۶	۰/۰۳۸	۰/۰۹۷
R^2 سری ارزیابی	۰/۰۷۷	۰/۰۰۰	۰/۰۵۴	۰/۰۱۹	۰/۰۵۰	۰/۱۰۰	۰/۱۹۰	۰/۰۰۳	۰/۰۶۱	۰/۰۹۴

۳-۱-۱۰- بررسی ارتباط توصیف‌کننده‌های منتخب با فعالیت ضد اختلال خواب

مهمتر از قدرت پیش‌بینی میزان فعالیت ترکیبات جدید، رسالت دیگر مدل‌های QSAR ایجاد ارتباط بین فعالیت مورد نظر و پارامترهای ساختاری یا توصیف‌کننده‌هاست که بتوان به کمک این ارتباط به بررسی تئوری و در نهایت عملی ترکیباتی جدید با خواص بهتر پرداخت. در این بخش به بررسی ارتباط بین توصیف‌کننده‌های منتخب با فعالیت ترکیبات ضد اختلال خواب می‌پردازیم و سعی می‌کنیم تا حد امکان توضیحی برای ارتباط بین این توصیف‌کننده‌ها و فعالیت مورد نظر بیابیم. جدول زیر مقادیر اثر متوسط^۱ هر توصیف‌کننده را نشان می‌دهد. اثر متوسط یک متغیر مستقل با استفاده از فرمول (۱-۳) به دست می‌آید:

$$t = \frac{\beta_x \times \sum x_n}{\sum RI} \quad (1-3)$$

که در آن، β_x ضریب متغیر مستقل x در مدل، $\sum x_n$ مقدار متغیر مستقل x مورد نظر برای ترکیب n ام و RI نیز شاخص بازداری تجربی می‌باشد.

اثرات متوسط توصیف‌کننده‌های منتخب بدست آمده از روش رگرسیون مرحله‌ای در جدول (۱۹-۳) گزارش شده است.

جدول (۱۹-۳) - اثر متوسط توصیف‌کننده‌های بدست آمده توسط روش رگرسیون مرحله‌ای

توصیف‌کننده	qneg	MATS6m	MATS1e	nROR	Mor10u	JGI1
اثر متوسط	+۱۴/۸۸۸	-۲۰/۳۰۷	-۲۱/۶۸۷	-۰/۳۶۴	-۰/۹۳۹	-۸/۹۸۰

الف- توصیف‌کننده‌های 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 توصیف‌کننده‌ها که در مدل برگزیده شده، توصیف‌کننده الکترونی توپوگرافیکی^۲ است که می‌تواند توزیع بار را در بازدارنده توصیف کند. qneg^۳ نیز از این گروه توصیف‌کننده‌ها می‌باشد.

^۱- Mean effect

^۲- Topographic electronic descriptor

^۳- Maximum negative charge

باشد که وارد مدل SR-ANN شده است. علامت مثبت 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 [۶۵].

ب- توصیف‌کننده‌های 2D autocorrelation

توصیف‌کننده‌های 2D autocorrelation بر اساس تابع خود ارتباطی AC_t تعریف می‌شوند:

$$AC_t = \int_b^a f(x) \cdot f(x+t) dx \quad (۲-۳)$$

در رابطه‌ی بالا $f(x)$ تابعی از متغیر x و t نیز نشان‌دهنده‌ی فواصل x می‌باشد، همچنین a و b کل فاصله مورد بررسی برای تابع $f(x)$ را بیان می‌کنند. $f(x)$ معمولاً یک تابع وابسته به زمان مانند یک سیگنال الکتریکی یا وابسته به مکان مانند دانسیته جمعیت در فضا می‌باشد.

نوعی از این توصیف‌کننده‌ها گروه Autocorrelation Geary است که بدین صورت محاسبه

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

می‌شود:

w یک ویژگی اتم، \bar{w} میانگین مقدار آن ویژگی روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است. δ_{ij} نیز به تابع کرونگر^۱ معروف است و Δ هم مجموعه‌ی δ هاست. مقدار این توصیف‌کننده که از جنس فاصله است، از صفر تا بی‌نهایت متغیر است [۶۶]. از این گروه توصیف‌کننده‌ها که دارای ضریب منفی می‌باشند، ${}^2\text{MATS6m}$ و ${}^2\text{MATS1e}$ وارد مدل SR-ANN شده‌اند که نقش الکترون‌گاتیویته‌ی ساندerson را در ترکیبات نشان می‌دهند و افزایش این توصیف‌کننده‌ها باعث کاهش pIC_{50} می‌شوند.

¹ - Kronecker

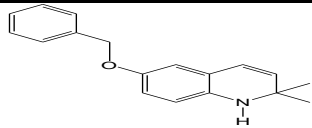
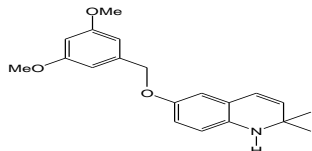
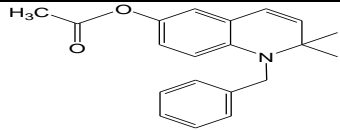
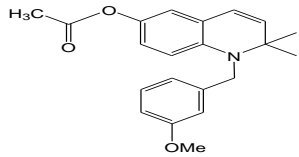
² - Moran auto correlation - lag6

³ - Moran auto correlation - lag1

پ- توصیف‌کننده‌های Functional group

توصیف‌کننده‌های Functional group از دسته توصیف‌کننده‌های گروه عاملی^۱ می‌باشند. nROR^۲ از جمله این توصیف‌کننده‌هاست که مشخص‌کننده‌ی نقش گروه‌های اتری در ترکیبات می‌باشد و دارای ضریب منفی در مدل SR-ANN می‌باشد. این توصیف‌کننده بیان‌گر این موضوع می‌باشد که افزایش تعداد استخلاف‌های اتری باعث کاهش مقادیر pIC₅₀ ترکیبات می‌شوند. چند مثال از اثر این توصیف‌کننده در جدول (۳-۲۰)، ارائه شده است.

جدول (۳-۲۰) - مثال‌هایی از اثر nROR بر فعالیت دارویی

ترکیب	nROR	pIC ₅₀
	۰	۵/۳۹
	۲	۴/۸۹
	۰	۷/۸۵
	۱	۶/۵۰

^۱ - Functional group

^۲ - Number of ethers

ت- توصیف‌کننده‌های 3D – MORSE¹

ایده اصلی توصیف‌کننده‌های 3D – Morse، به دست آوردن اطلاعاتی از مختصات اتمی سه بعدی با استفاده از تبدیلی است که در مطالعات پراش الکترونی، برای فراهم آوردن منحنی‌های نظری تفرق کاربرد دارد:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad (4-3)$$

در این رابطه I شدت الکترون پراکنده شده، A یک خاصیت اتمی، S زاویه پراکندگی، r_{ij} فاصله بین اتم‌های i، 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.

این توصیف‌کننده‌ها قادرند ارتباط بین ساختار سه بعدی ترکیبات آلی و خصوصیات فیزیکی، شیمیایی و بیولوژیکی را فراهم سازند. چون این توصیف‌کننده‌ها آرایش سه بعدی اتم‌ها را بدون اینکه به اندازه مولکول ارتباط داشته باشند، بیان می‌کنند؛ بنابراین برای تعداد زیادی مولکول با تفاوت‌های ساختاری زیاد قابل کاربرد است [67]. از این گروه توصیف‌کننده‌ها، توصیف‌کننده‌ی سه بعدی Mor10u² در مدل SR-ANN دیده می‌شود که دارای ضریب منفی می‌باشد.

ث- توصیف‌کننده‌های Galvez topol- charge indices

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

$$M = A \times D^{-2} \quad (5-3)$$

¹ - 3D- Molecular Representation of Structure based on Electron diffraction

² - 3D –MORSE-signal10

که در آن A ماتریس مجاورت^۱ و D^{-2} ماتریس عکس مربع فاصله^۲ می‌باشد همچنین M (ماتریس گالوز)^۳ یک ماتریس مربعی نامتقارن $A \times A$ می‌باشد که A مشخص‌کننده‌ی تعداد اتم‌های مولکول است و به صورت زیر تعریف می‌شود:

$$CT = \partial_i \quad \text{if } i = j \quad (۶-۳)$$

$$CT = m_{ij} - m_{ji} \quad \text{if } i \neq j \quad (۷-۳)$$

$$CT_{ij} = D(M) + [M - M^T] \quad (۸-۳)$$

که در آن m_{ij} عناصر M و ∂_i درجه‌ی بردار اتم i می‌باشد. ورودی‌های قطری CT ظرفیت توپولوژیکی اتم‌ها و ورودی‌های قطری CT_{ij} مقدار بار منتقل شده از اتم j به اتم i می‌باشد. $JGHI$ از این گروه توصیف‌کننده‌ها با ضریب منفی می‌باشد که وارد مدل SR-ANN شده است.

۳-۱-۱۱- بررسی میزان مشارکت توصیف‌کننده‌های منتخب در شبکه عصبی

به منظور بررسی بیشتر اثر توصیف‌کننده‌های به کار گرفته شده در مدل شبکه عصبی مصنوعی، در صد مشارکت هر یک از توصیف‌کننده‌ها به صورت زیر محاسبه شد:

- ۱- توصیف‌کننده‌ی مورد نظر از سری توصیف‌کننده‌های ارائه به شبکه بهینه حذف شد.
- ۲- با استفاده از بقیه توصیف‌کننده‌ها مقدار متغیر وابسته (pIC_{50}) برای داده‌های سری ارزیابی محاسبه شد.
- ۳- با استفاده از مقادیر پاسخ محاسبه شده و مقادیر پاسخ تجربی، میانگین خطای مطلق در غیاب توصیف‌کننده‌ی مورد نظر محاسبه شد.
- ۴- مراحل ۱ تا ۳ برای هریک از توصیف‌کننده‌ها تکرار شد.
- ۵- در پایان در صد مشارکت هر یک از توصیف‌کننده‌ها به کمک رابطه‌ی (۸-۳) بدست آمد.

¹ - Adjacency

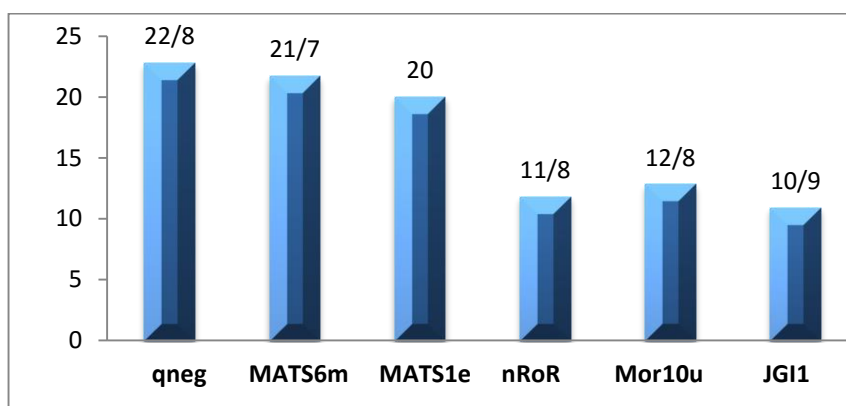
² - Reciprocal square distance

³ - Galvez matrix

⁴ - Mean topological charge

$$c_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad (9-3)$$

در رابطه‌ی بالا c_i درصد مشارکت توصیف‌کننده‌ی حذف شده نام، N تعداد توصیف‌کننده‌های استفاده شده در مدل و Δm_i میانگین خطای مطلق حاصل از سری ارزیابی در غیاب نام را نشان می‌دهد [۶۸]. بر این اساس درصد مشارکت توصیف‌کننده‌های منتخب در ترکیبات مورد بررسی به شکل زیر می‌باشد.



شکل (۳-۱۸) - مشارکت توصیف‌کننده‌ها در شبکه عصبی بهینه

همان‌طور که مشاهده می‌شود، توصیف‌کننده‌های qneg، MATS6m و MATS1e بیشترین اثر مشارکت را در مدل ارائه شده دارند. توصیف‌کننده‌های MATS6m و MATS1e در برگیرنده اطلاعات دو بعدی مولکول هستند. بنابراین می‌توان گفت که فعالیت بازدارنده این ترکیبات وابستگی قابل ملاحظه‌ای به توپولوژی ترکیب بازدارنده دارد. همچنین توصیف‌کننده‌ی qneg از دسته توصیف‌کننده‌هایی است که توزیع بار را در بازدارنده توصیف می‌کند.

۳-۱-۱۲ - نتیجه‌گیری نهایی

هدف اصلی از کار حاضر، ارائه مدل‌های SR-ANN و GA-ANN جهت مطالعه ارتباط کمی بین ساختار و فعالیت ضد اختلال خواب ترکیبات ۱ و ۲-دی هیدرو کینولین و همچنین مشتقات استری آن‌ها می‌باشد، بنابراین این مدل‌ها می‌توانند فعالیت ضد اختلال خواب، برای همین طبقه از

ترکیبات که هنوز سنتز نشده‌اند را پیش‌بینی کنند. نتایج نشان می‌دهند که مدل‌های SR-ANN و GA-ANN می‌توانند به عنوان روش‌های موفق برای مدل‌سازی و پیش‌بینی فعالیت بازدارندگی ترکیبات مورد مطالعه باشند و این مدل‌ها می‌توانند به محققان در طراحی داروهایی با فعالیت بازدارندگی قویتر کمک کنند.

فصل چهارم

- مطالعه ارتباط کمی ساختار-فعالیت برخی مشتقات پیریدوپیریدازین به عنوان بازدارنده‌های p38 MAP کیناز

با استفاده از SR-ANN و GA-ANN

زین به عنوان بازدارنده

۴-۱- مدل سازی فعالیت

های p38 MAP کیناز

؛ پیریدازین به عنوان بازدارنده

در راستای مطالعه ارتباط

G استفاده شد و قدرت پیش

های p38 MAP کیناز، از روش

بینی مدل های بدست آمده از هر روش پیریدازین به عنوان بازدارنده سنجش قرار گرفت.

۴-۱-۱- انتخاب سری داده ها

در این پروژه فعالیت دارویی (IC_{50}) ۴۲ ترکیب از مشتقات پیریدوپیریدازین به عنوان سری

داده ها مورد بررسی قرار گرفت [۶۹]. برای مدل سازی، مقادیر فعالیت به صورت $\text{Log}(IC_{50} \times 10^{-9})$

(برحسب نانومولار) مورد استفاده قرار گرفتند. در جدول (۴-۱) جزئیات استخلاف ها و مقادیر

عددی pIC_{50} برای هر ترکیب آورده شده است. داده های جدول نشان می دهد که نوع استخلاف و

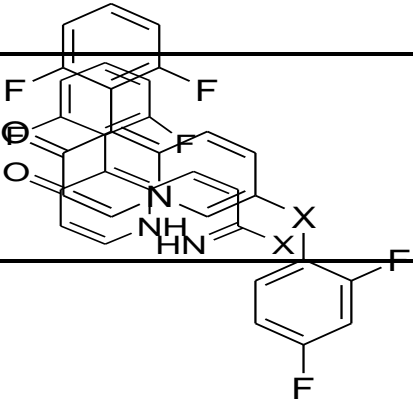
نحوه آرایش آن ها در اطراف اسکلت اصلی مولکولی نقش مهمی در اثر بازدارندگی ترکیب دارد.

بنابراین در این تحقیق تلاش شده است تا مدلی ارائه گردد که بتواند میزان بازدارندگی یک ترکیب را

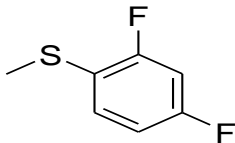
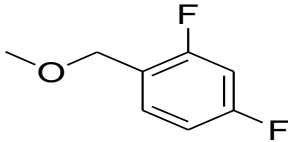
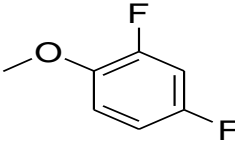
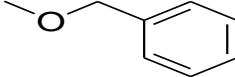
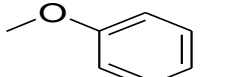
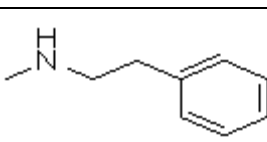
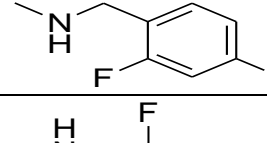
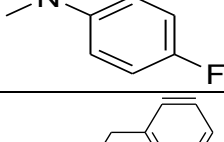
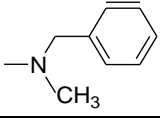
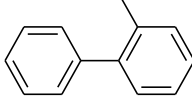
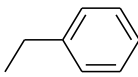
با هر نوع استخلاف در هر موقعیتی پیش بینی کند.



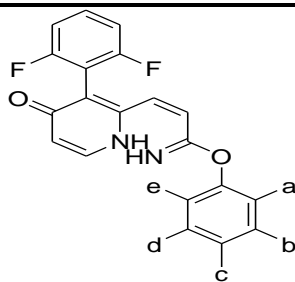
جدول (۱-۴) - ساختار و مقادیر pIC_{50} ترکیبات مورد مطالعه

				
شماره ترکیب	سری داده	X	IC ₅₀ (nM)	pIC ₅₀
۱	آموزش	-	۷/۰	۸/۱۶
۲	ارزیابی	O	۱۹۰/۰	۶/۷۲
۳	تست	CH ₂	۹۳/۰	۷/۰۳
۴	آموزش	NHCH ₂	۸۷۰/۰	۶/۰۶
۵	آموزش	C ₂ H ₂	۸۹۰/۰	۶/۰۵
۶	ارزیابی	C ₂ H ₄	۴۹۰/۰	۶/۳۱

ادامه جدول (۱-۴)

شماره ترکیب	سری داده	X	IC ₅₀ (nM)	pIC ₅₀
۷	آموزش		۳/۲	۸/۵۰
۸	تست		۷۲۰/۰	۶/۱۴
۹	آموزش		۴/۷	۸/۳۳
۱۰	آموزش		۱۳۰۰/۰	۵/۸۹
۱۱	ارزیابی		۱۶۰/۰	۶/۸۰
۱۲	آموزش		۱۳۰۰/۰	۵/۸۹
۱۳	تست		۷۱۹/۰	۶/۱۴
۱۴	آموزش		۳/۵	۶/۵۰
۱۵	آموزش		۶۵۰/۰	۶/۱۹
۱۶	ارزیابی		۴۸۰/۰	۶/۳۲
۱۷	آموزش		۱۴۳۰/۰	۵/۸۵

ادامه جدول (۴-۱)



شماره ترکیب	سری داده	a	b	c	d	e	IC ₅₀ (nM)	pIC ₅₀
۱۸	تست	F	-	-	-	-	۱۲/۰	۷/۹۲
۱۹	آموزش	Cl	-	-	-	-	۴۴/۰	۷/۳۶
۲۰	ارزیابی	Me	-	-	-	-	۹۲/۰	۷/۰۴
۲۱	آموزش	OMe	-	-	-	-	۸۲/۰	۷/۰۹
۲۲	آموزش	CF ₃	-	-	-	-	۲۴۰۰/۰	۵/۶۲
۲۳	تست	-	F	-	-	-	۴۲۰/۰	۶/۳۸
۲۴	ارزیابی	-	Cl	-	-	-	۲۴۰/۰	۶/۶۲
۲۵	آموزش	-	Me	-	-	-	۷۰/۰	۷/۱۶
۲۶	آموزش	-	-	F	-	-	۴۸/۰	۷/۳۲
۲۷	آموزش	-	-	Cl	-	-	۷۹/۰	۷/۱۰
۲۸	تست	-	-	Me	-	-	۲۶۰/۰	۶/۵۹
۲۹	ارزیابی	F	F	-	-	-	۳۲/۰	۷/۵۰
۳۰	آموزش	Cl	Cl	-	-	-	۳۲/۰	۷/۵۰
۳۱	آموزش	F	-	Cl	-	-	۱۱/۰	۷/۹۶
۳۲	تست	Cl	-	F	-	-	۱۴/۰	۷/۸۵
۳۳	آموزش	Cl	-	Cl	-	-	۳۰/۰	۷/۵۲
۳۴	ارزیابی	F	-	-	F	-	۱۳۰/۰	۶/۸۹
۳۵	آموزش	F	-	-	-	F	۹۵/۰	۷/۰۲
۳۶	آموزش	Cl	-	-	-	Cl	۹۰/۰	۷/۰۵
۳۷	تست	F	-	F	-	F	۲۰۰/۰	۶/۷۰
۳۸	آموزش	Cl	-	F	-	Cl	۶۸/۰	۷/۱۷
۳۹	ارزیابی	F	F	F	F	F	۴۳۰۰/۰	۵/۳۷
۴۰	آموزش	-	F	F	-	-	۲۷۰/۰	۶/۵۷
۴۱	تست	-	Cl	F	-	-	۱۴۰/۰	۶/۸۵
۴۲	آموزش	-	Cl	Cl	-	-	۲۶۰/۰	۶/۵۹

۴-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها

در این مرحله از مطالعه، باید ساختار ترکیبات مورد بررسی رسم و بهینه‌سازی شود. جهت انجام این کار از نرم‌افزار Hyperchem 8.0 استفاده گردید. ساختارها ابتدا در این نرم‌افزار رسم شدند و سپس ساختار نهایی به روش نیمه تجربی AM1 بهینه گردید. بهینه‌سازی تا رسیدن به گرادیان انرژی ۰/۰۰۱ کیلوکالری بر مول ادامه یافت.

۴-۱-۳- محاسبه و انتخاب توصیف‌کننده‌های مهم

ساختارهای سه بعدی رسم شده برای ۴۲ ترکیب موجود به عنوان داده‌های ورودی به نرم‌افزار Dragon2.1 داده شد که این نرم‌افزار تعداد ۱۴۸۱ توصیف‌کننده را در ۱۸ گروه مختلف برای هر مولکول محاسبه نمود.

به منظور انتخاب توصیف‌کننده‌های مهم از میان تمام توصیف‌کننده‌های محاسبه شده، توصیف‌کننده‌هایی که بیش از ۹۰٪ از مقادیر آن‌ها یکسان بودند، حذف شدند. سپس با بررسی همبستگی بین متغیرها، از میان هر دو توصیف‌کننده‌ای که ضریب همبستگی بزرگتر از ۰/۹ داشتند، آنکه که همبستگی کمتری با متغیر وابسته داشت، حذف شد. بدین ترتیب در پایان این مرحله تعداد ۳۵۷ توصیف‌کننده باقی ماند. در این تحقیق برای انتخاب توصیف‌کننده‌ها از روش خطی رگرسیون مرحله‌ای (SR) و روش الگوریتم ژنتیک (GA) استفاده شده است.

۴-۱-۳-۱- انتخاب توصیف‌کننده‌های مناسب توسط رگرسیون مرحله‌ای در SPSS

در روش رگرسیون مرحله‌ای، متغیرها یکی پس از دیگری وارد مدل شده و ابتدا متغیری وارد مدل می‌شود که بالاترین میزان همبستگی را با متغیر وابسته دارد. با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آن‌ها سطح معناداری خود را از دست داده باشند، قبل از ورود متغیر جدید از مدل خارج می‌شوند. بدین ترتیب از میان ۳۵۷ توصیف‌کننده‌ای

ارائه شده تعداد ۱۲ توصیف‌کننده انتخاب شدند. نام این توصیف‌کننده‌ها همراه با معنی و طبقه آن‌ها در جدول (۲-۴) ارائه شده است.

جدول (۲-۴) - توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای

No	Symbol	Class	Meaning
1	MATS4e	2D auto correlation	Moran auto correlation – lag4
2	H8v	GETAWAY	H autocorrelation of lag8
3	MDDD	Topological	Mean distance degree deviation
4	VAR	Topological	Variation
5	PW ₄	Topological	Path/walk4- Randic shape index
6	IDDE	Topological	Mean information on the distance degree equality
7	G3m	WHIM	3st component symmetry directional WHIM index
8	GATS3p	2D autocorrelation	Geary autocorrelation – lag3
9	Mor27m	3D MORSE	3D –MORSE-signal27
10	MATS8e	2D auto correlation	Moran auto correlation – lag8
11	G2u	WHIM	3st component symmetry directional WHIM index
12	GATS4e	2D autocorrelation	Geary autocorrelation – lag4

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

ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول (۳-۴) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای

	MATS4e	H8v	MDDD	VAR	PW ₄	IDDE	G3m	GATS3p	Mor27m	MATS8e	G2u	GATS4e
MATS4e	۱/۰۰۰											
H8v	۰/۰۱۷	۱/۰۰۰										
MDDD	۰/۳۸۱	-۰/۱۱۶	۱/۰۰۰									
VAR	۰/۲۴۱	۰/۱۰۰	۰/۴۵۸	۱/۰۰۰								
PW ₄	۰/۲۴۸	-۰/۴۱۶	۰/۰۴۷	-۰/۳۳۰	۱/۰۰۰							
IDDE	۰/۰۳۹	-۰/۰۵۱	۰/۴۷۹	۰/۳۲۲	-۰/۰۵۶	۱/۰۰۰						
G3m	-۰/۰۸۰	-۰/۱۷۲	-۰/۲۵۸	-۰/۲۵۲	۰/۱۶۳	-۰/۰۲۴	۱/۰۰۰					
GATS3p	-۰/۳۰۷	۰/۱۳۹	۰/۳۱۱	-۰/۳۲۳	-۰/۱۲۸	-۰/۱۵۶	-۰/۱۶۳	۱/۰۰۰				
Mor27m	-۰/۵۲۳	۰/۳۲۷	-۰/۳۵۲	-۰/۱۶۲	-۰/۳۹۵	-۰/۰۶۰	-۰/۰۰۲	-۰/۱۹۰	۱/۰۰۰			
MATS8e	-۰/۴۹۰	۰/۴۲۷	-۰/۵۵۰	-۰/۳۰۷	-۰/۳۷۱	-۰/۲۳۳	-۰/۱۷۵	-۰/۲۰۸	۰/۳۸۰	۱/۰۰۰		
G2u	-۰/۰۲۱	-۰/۱۳۱	۰/۰۱۸	-۰/۰۳۶	-۰/۰۳۵	۰/۱۸۵	-۰/۱۰۹	-۰/۲۵۷	۰/۰۴۵	-۰/۰۸۳	۱/۰۰۰	
GATS4e	-۰/۲۶۲	۰/۲۴۱	-۰/۵۵۰	-۰/۲۹۱	-۰/۳۶۴	-۰/۱۸۹	-۰/۰۵۴	-۰/۲۷۳	۰/۵۷۷	۰/۵۱۰	-۰/۰۳۰	۱/۰۰۰

۴-۱-۳-۲- انتخاب توصیف‌کننده‌های مناسب توسط الگوریتم ژنتیک در MATLAB

بر اساس روشی که در بخش (۳-۱-۴-۲) توضیح داده شد، انتخاب توصیف‌کننده‌های مناسب توسط الگوریتم ژنتیک در پیش‌بینی فعالیت دارویی موردنظر برآورد شد که نتایج آن در این بخش ارائه شده است. در فرآیند اجرای الگوریتم ژنتیک، ابتدا به صورت اتفاقی یک جمعیت اولیه از توصیف‌کننده‌ها تولید می‌شود سپس برآزندگی یا شایستگی تک‌تک اعضای هر نسل یعنی توصیف‌کننده‌ها، محاسبه و با توجه به شایستگی‌ها، نسل‌های بعدی با اعمال سه پارامتر جمعیت، پیوند و جهش تولید یا باز ترکیب می‌شوند.

جمعیت

همان‌طور که در فصل سوم بیان شد، اکثر محققان برای اندازه جمعیت عددی بین ۳۰ تا ۱۰۰ پیشنهاد کرده‌اند. این مجموعه از جمعیت‌های انتخاب شده اولین نسل را ایجاد نموده و از نسل اولیه نیز نسل‌های بعدی ایجاد می‌شود. سرانجام در نسل نهایی، یک جمعیت انتخاب شده که موجب حل بهینه‌ی مسأله است، به دست می‌آید [۶۲]. در این مطالعه، به منظور کاهش زمان همگرایی به پاسخ مناسب، به طور حدسی از یک جامعه‌ی اولیه که دارای اندازه‌ی بزرگتری نسبت به جامعه تولیدی بعدی می‌باشد، استفاده شده است. در تحقیق حاضر، اندازه جمعیت عدد ۳۰ در نظر گرفته شده است [۶۱].

عملگر پیوند

عملگر پیوند می‌تواند دو نقطه‌ای باشد که در این صورت، دو نقطه به طور تصادفی انتخاب شده و پیوند بین آن دو اتفاق می‌افتد. این عملیات در چند نقطه انجام شده که به آن بازترکیبی چند نقطه‌ای می‌گویند. معمولاً احتمال پیوند برای هر زوج کروموزوم ۰/۱ تا ۰/۹ در نظر گرفته می‌شود.

شود [۶۳]. در این تحقیق این مقدار ۰/۵ فرض شده است، که مقدار پیش فرض برنامه بکار برده شده می باشد [۶۱].

عملگر جهش

جهش در جمعیت تنوع ایجاد نموده و به پرهیز از همگرایی زودرس کمک می کند. احتمال وقوع این عملگر بین ۰/۰۰۵ تا ۰/۰۱ انتخاب می شود (۰/۵ تا ۱ درصد) [۶۴]. در این مطالعه مقدار ۰/۰۱ در نظر گرفته شده است [۶۱].

به منظور اجرای برنامه الگوریتم ژنتیک تعداد ۳۵۷ توصیف کننده ای که بالاترین میزان همبستگی را با متغیر وابسته داشتند به عنوان ورودی به این برنامه داده شدند. در جدول (۴-۴) توصیف کننده های انتخاب شده در هر بار اجرای الگوریتم ژنتیک گزارش شده است. همانطور که در

تعداد تکرار	توصیف کننده های انتخاب شده توسط الگوریتم ژنتیک
1	MATS1e- MAXDN- H8v- HATS2u- G3m - Mor12u - MDDD- Mor27m- Mor22m- RDF030m
2	MATS1e- HATS2u- H8v- MAXDN- G3m- Mor12u- MSD- GATS3p- GATS6e

فصل سوم بیان شد، بدلیل اینکه توصیف کننده های تصادفی حذف گردند و نتایج خوبی حاصل گردد این روش ۵ بار تکرار گردید.

3	MATS1e- HATS2u- MAXDN- H8v- G3m- Mor12u- VAR- Mor27m- Mor11u
4	MATS1e- H8v- HATS2u- MAXDN- Mor27m- G3m- G2u
5	MATS1e- HATS2u- H8v- G3m- Mor12u- MAXDN- Mor27m- Mor17m- MATS8e

سپس توصیف‌کننده‌های انتخاب شده در این ۵ بار تکرار را مجدداً وارد برنامه GA نموده و توصیف‌کننده‌های انتخاب شده در این مرحله به‌عنوان توصیف‌کننده‌های انتخاب شده نهایی در نظر گرفته شد که به ترتیب زیر می‌باشند:

MATS1e - HATS2u - H8v - MAXDN - G3m - Mor27m - Mor12u

با توجه به موارد ذکر شده، توصیف‌کننده‌های انتخاب شده نهایی توسط این روش همراه با گروه و معنی آن‌ها در جدول (۴-۵) آورده شده است.

جدول (۴-۵) - توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک

No	Symbol	Class	Meaning
1	MATS1e	2D auto correlation	Moran auto correlation – lag1
2	HATS2u	GETAWAY	Leverage-weighted correlation – lag2
3	H8v	GETAWAY	H autocorrelation of lag8
4	MAXDN	Geometrical	Maximal electropological negative variation
5	G3m	WHIM	3st component symmetry directional WHIM index
6	Mor27m	3D MORSE	3D-MORSE-signal27
7	Mor12u	3D MORSE	3D-MORSE-signal12

همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۴-۶) ارائه شده که این ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول (۴-۶) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده نهایی توسط الگوریتم ژنتیک

	MATS1e	HATS2u	H8v	MAXDN	G3m	Mor27m	Mor12u
MATS1e	۱/۰۰۰						
HATS2u	۰/۰۸۴	۱/۰۰۰					

H8v	-۰/۲۱۸	-۰/۲۳۲	۱/۰۰۰				
MAXDN	۰/۰۰۲	۰/۰۴۸	-۰/۰۲۵	۱/۰۰۰			
G3m	-۰/۱۹۸	۰/۲۶۷	-۰/۱۷۲	۰/۲۴۱	۱/۰۰۰		
Mor27m	-۰/۲۵۳	-۰/۲۸۸	۰/۳۲۷	-۰/۰۲۸	-۰/۰۰۲	۱/۰۰۰	
Mor12u	-۰/۰۸۱	۰/۵۸۷	-۰/۴۲۲	۰/۲۸۰	۰/۲۸۰	-۰/۴۳۶	۱/۰۰۰

۴-۱-۴- مدل‌سازی شبکه عصبی مصنوعی با استفاده از توصیف‌کننده‌های

انتخاب شده توسط رگرسیون مرحله‌ای

در این پروژه، شبکه‌ی عصبی مصنوعی جهت پیش‌بینی فعالیت بیولوژیکی سری داده‌ها مورد استفاده قرار گرفت. برنامه‌نویسی و اجرای یک شبکه‌ی عصبی پیش‌خور سه لایه با الگوی آموزشی پس انتشار در محیط نرم‌افزاری مطلب انجام شد. سری داده‌ها به طور تصادفی به سه قسمت آموزش (۲۴ ترکیب)، سری ارزیابی (۹ ترکیب) و سری تست (۹ ترکیب) تقسیم شدند. سپس پارامترهای مؤثر بر آموزش موفق شبکه عصبی پس‌انتشار از قبیل تعداد ورودی‌ها، تعداد لایه‌های پنهان، تعداد گره‌های لایه پنهان، نوع تابع انتقال، نوع تابع آموزش، تعداد دوره‌های آموزش و پارامتر μ بهینه شدند.

۴-۱-۴-۱- بهینه‌سازی پارامترهای شبکه

برای بهینه‌سازی پارامترهای شبکه، شبکه‌هایی با تعداد ورودی‌های از ۲ تا ۱۲ توصیف‌کننده ایجاد شدند. هر شبکه با تابع آموزش لونیگ-مارکوارت و تنظیم بایزین با تعداد گره متفاوت از ۲ تا ۱۰ گره در لایه پنهان آموزش داده شد. همچنین تعداد دوره‌های آموزش نیز بطور همزمان از ۲ تا ۳۰ تغییر داده شد. برای انتخاب بهترین تابع انتقال از توابع تانژانت سیگموئیدی ($\tan \text{sig}$) و لگاریتم سیگموئیدی ($\log \text{sig}$) به عنوان تابع انتقال استفاده شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده گردید. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مجموع مربعات خطا (MSE) به عنوان معیار انتخاب شد.

۲-۴-۱-۴- بهینه‌سازی همزمان تعداد متغیرهای ورودی شبکه، تابع آموزش، نوع تابع

انتقال، تعداد گره‌های لایه‌های پنهان و تعداد دوره‌های آموزشی

در این قسمت، بهینه‌سازی و انتخاب بهترین تعداد متغیرهای ورودی شبکه، تابع آموزشی، نوع

تابع انتقال، تعداد نرون‌های لایه پنهان و تعداد دوره‌های آموزش به طور همزمان انجام شد.

جهت بهینه‌سازی همزمان تمام پارامترها، در مجموع ۱۱۴۸۴ شبکه مورد ارزیابی قرار گرفته

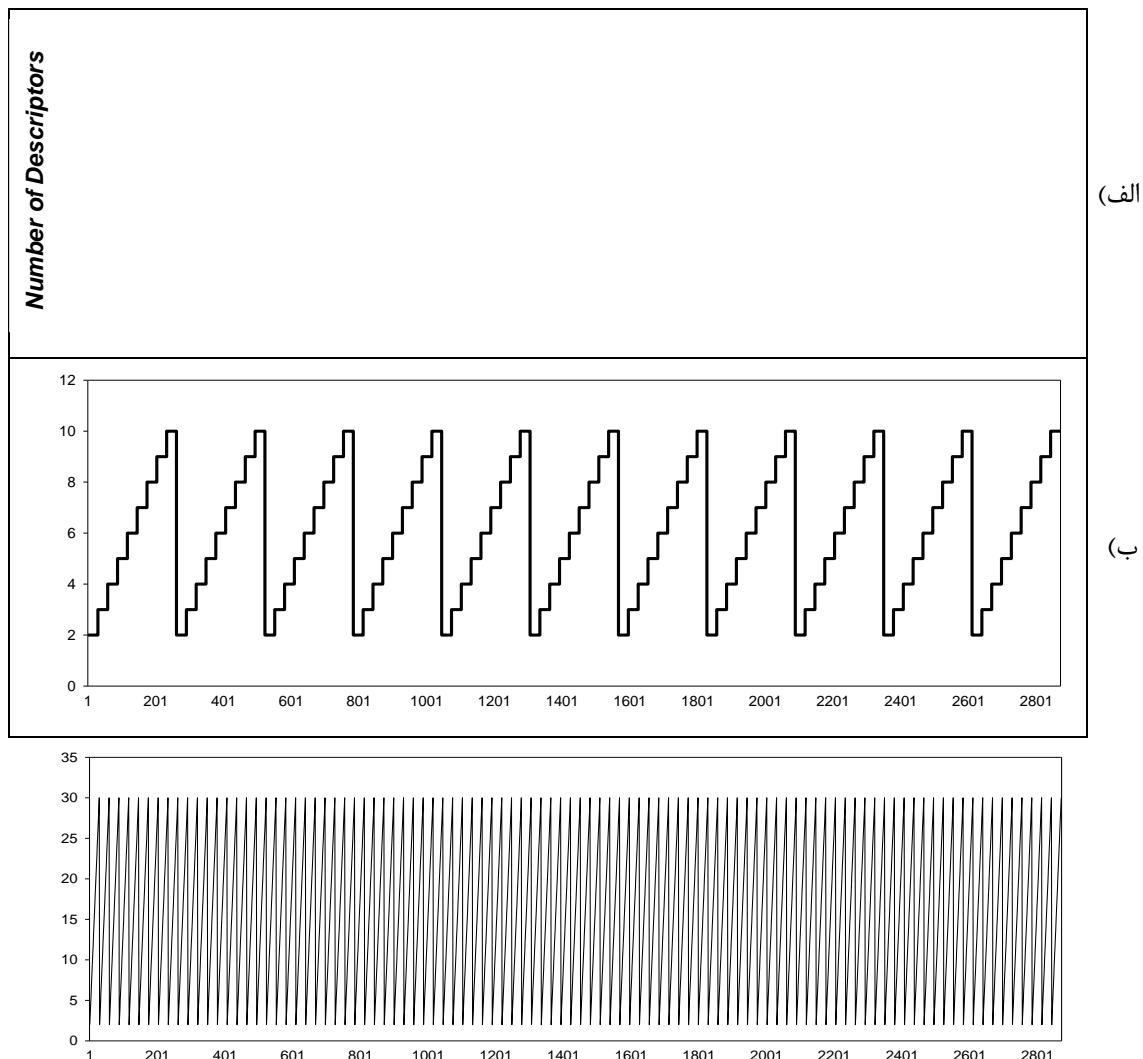
است که در زیر به تعداد تمام موارد مورد بررسی اشاره شده است.

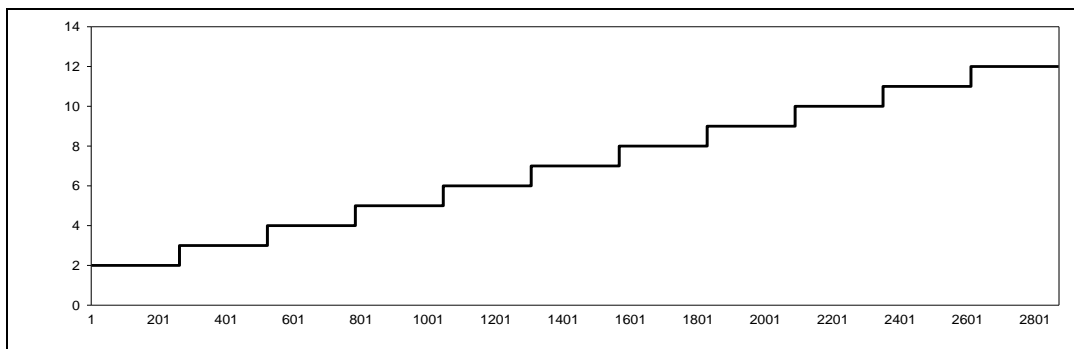
$11484 = 29 \times 9 \times 2 \times 2 \times 2 \times 11$

بخشی از روند تغییرات پارامترهای شبکه در حین بهینه‌سازی همزمان پارامترها به همراه

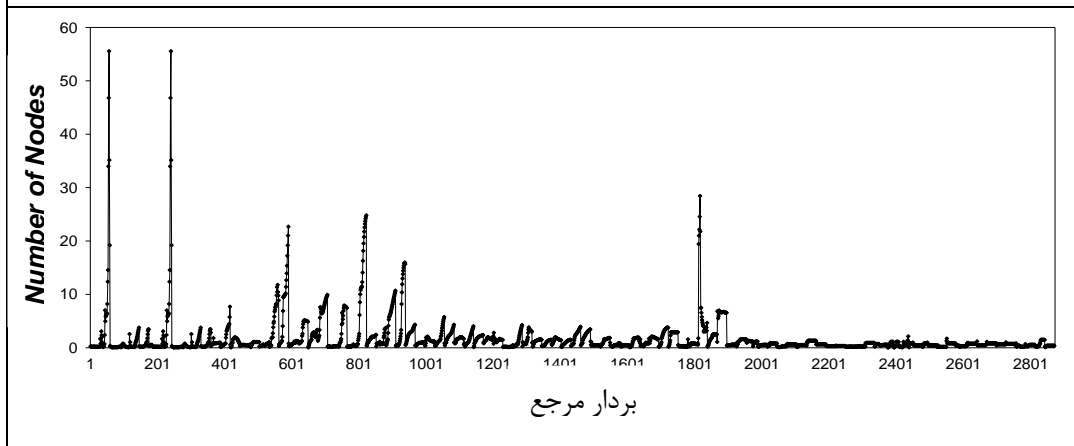
مقادیر MSE به دست آمده به صورت نموداری بر حسب یک بردار مرجع فرضی در شکل‌های (۴-۱)،

(۲-۴)، (۳-۴) و (۴-۴) آمده است.





(ج)

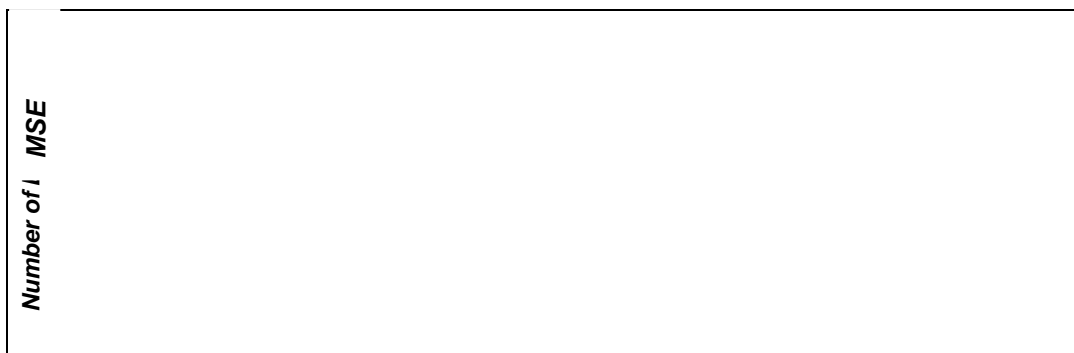


(د)

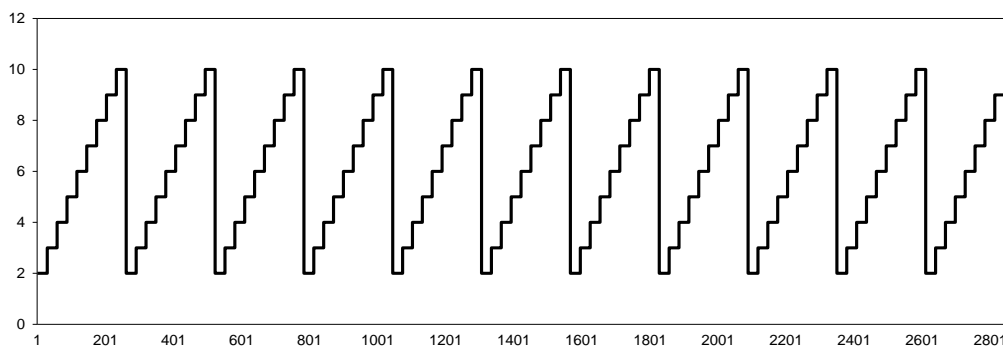
شکل (۴-۱) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقدار

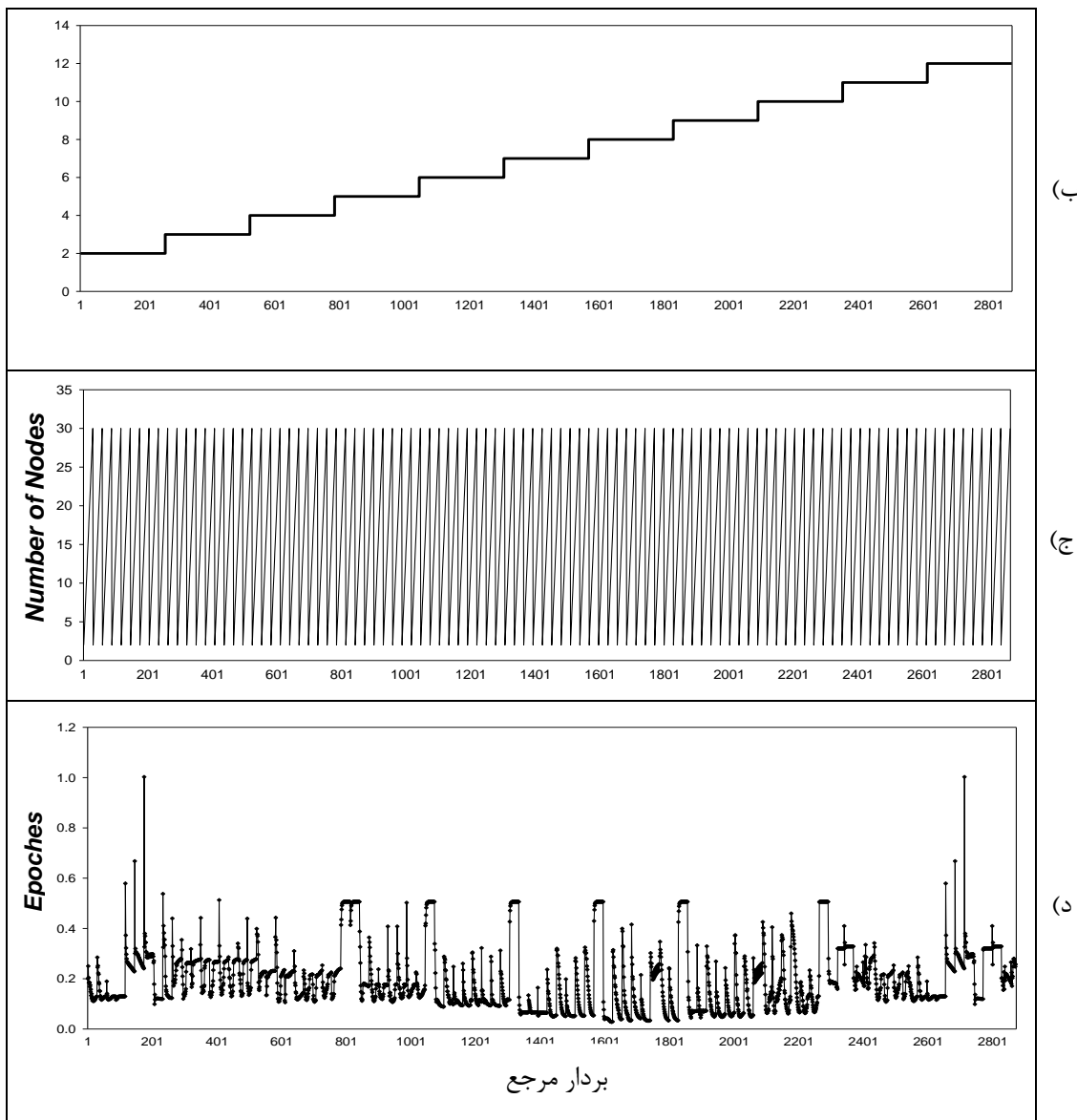
MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید

Epochs



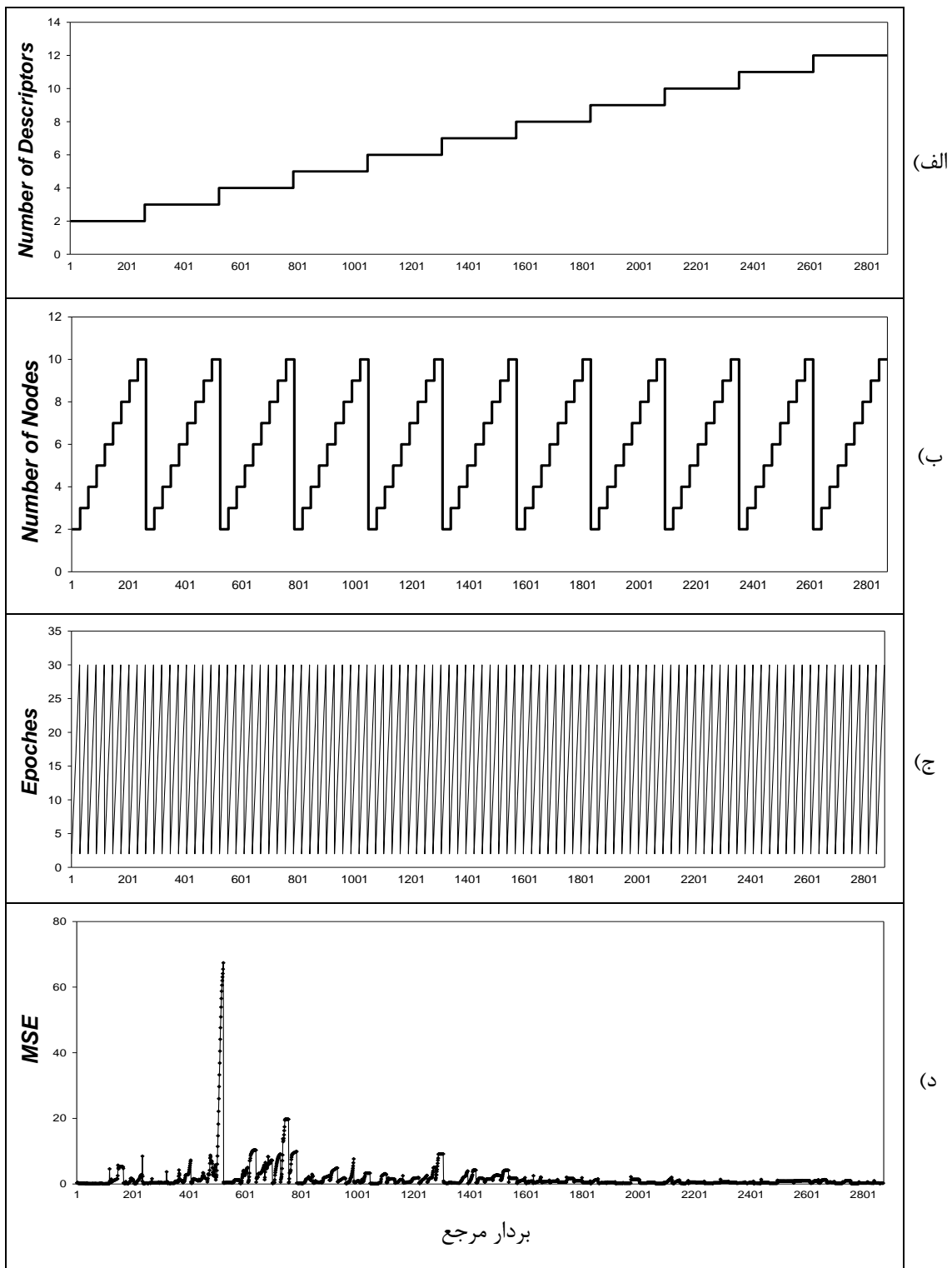
(الف)



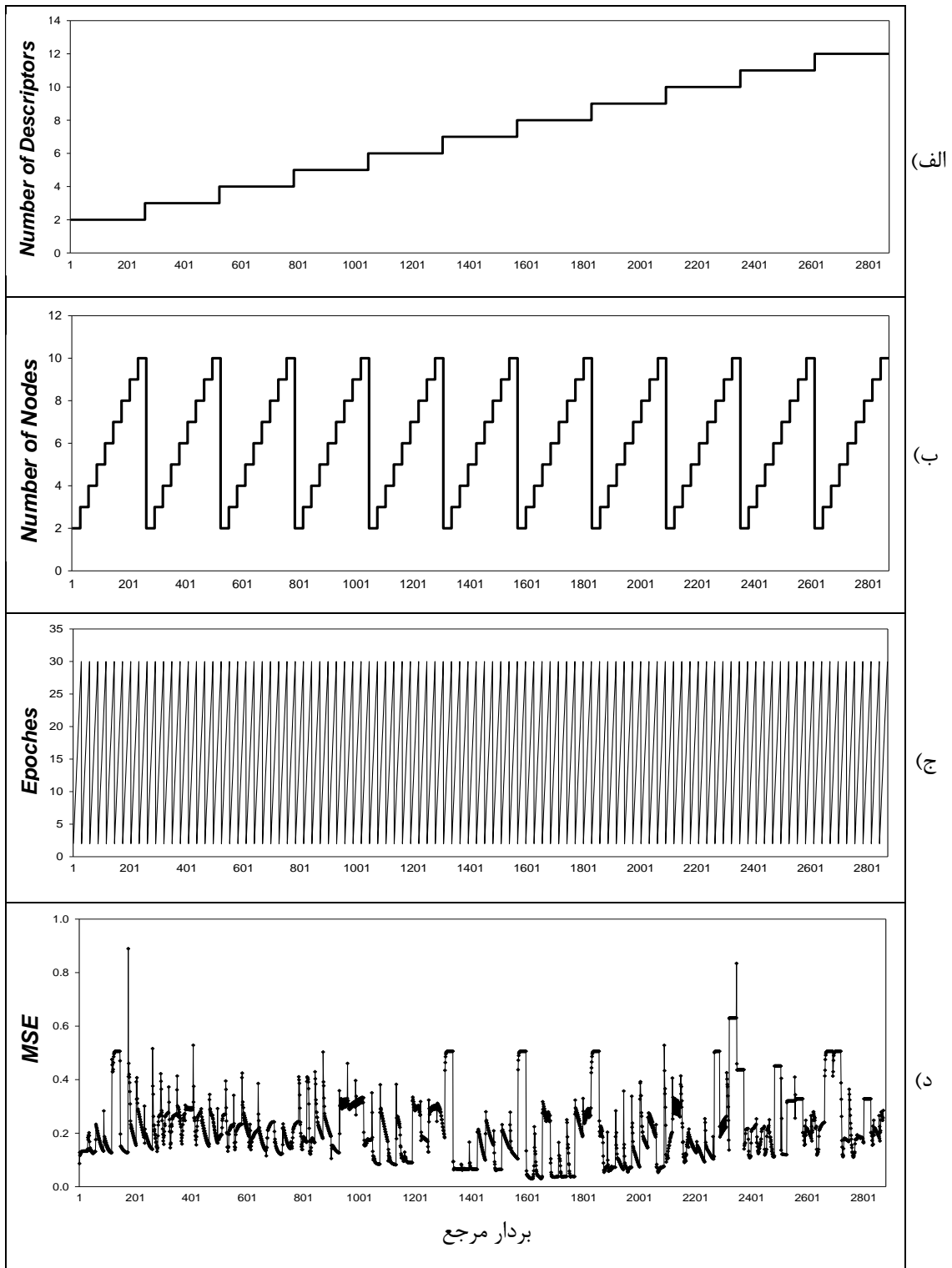


شکل (۴-۲) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقایسه MSE برای تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئید

MSE



شکل (۴-۳) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید



شکل (۴-۴) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئید

جدول (۴-۷) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا

MSE	تعداد دوره‌های آموزش	تعداد گره	تعداد توصیف کننده‌ها	تابع آموزش	تابع انتقال
۰/۰۵۲۸	۴	۵	۱۱	لونبرگ - مارکوارت	لگاریتم سیگموئید
۰/۰۲۸۰	۲۷	۳	۸	تنظیم بایزین	لگاریتم سیگموئید
۰/۰۵۱۲	۷	۵	۲	لونبرگ - مارکوارت	تانژانت سیگموئید
۰/۰۲۹۴	۲۴	۳	۸	تنظیم بایزین	تانژانت سیگموئید

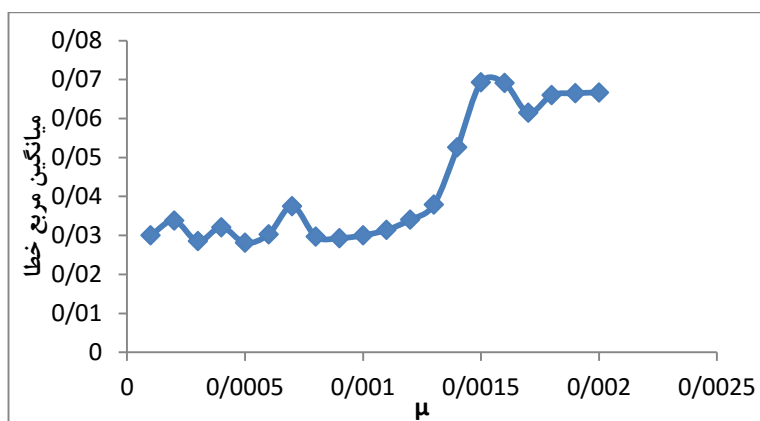
با توجه به نتایج، تابع آموزش تنظیم بایزین نسبت به لونبرگ- مارکوارت دارای میانگین مربعات خطای کمتری است. همچنین تابع انتقال لگاریتم سیگموئید به همراه تابع آموزش تنظیم بایزین دارای خطای کمتری نسبت به تانژانت سیگموئید می‌باشند، لذا به عنوان توابع آموزش و انتقال بهینه انتخاب گردیدند. همچنین طبق نتایج بدست آمده، تعداد ورودی‌های شبکه، تعداد گره در لایه‌ی پنهان و تعداد دوره‌های آموزشی نیز که بطور همزمان بهینه شدند به ترتیب برابر با ۸ متغیر ورودی، ۳ گره در لایه‌ی پنهان و تعداد دوره‌های آموزش برابر ۲۷ بدست آمد.

۴-۱-۳- بهینه‌سازی پارامتر μ

جهت یافتن مقدار بهینه μ ، ساختار شبکه بهینه با ۸ متغیر ورودی، ۳ گره در لایه‌ی پنهان، تعداد دوره‌های آموزش برابر ۲۷، تابع انتقال لگاریتم سیگموئید و تابع آموزشی تنظیم بایزین در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۰۱ تا ۰/۰۰۰۲ (حداقل میانگین مربعات خطا در این رنج مشاهده گردید) با گام‌های ۰/۰۰۰۱ تغییر داده شد و آنگاه برای هر مورد، مقدار میانگین مربع خطای سری ارزیابی محاسبه گردید. سرانجام مقدار میانگین مربع خطا بر حسب μ ترسیم گردید. نقطه‌ای که کمترین خطا را برای سری آموزش داشت، به عنوان مقدار بهینه انتخاب شد که طبق شکل (۴-۵) مقدار بهینه ۰/۰۰۰۵ به دست آمد.

جدول (۴-۸) - میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ

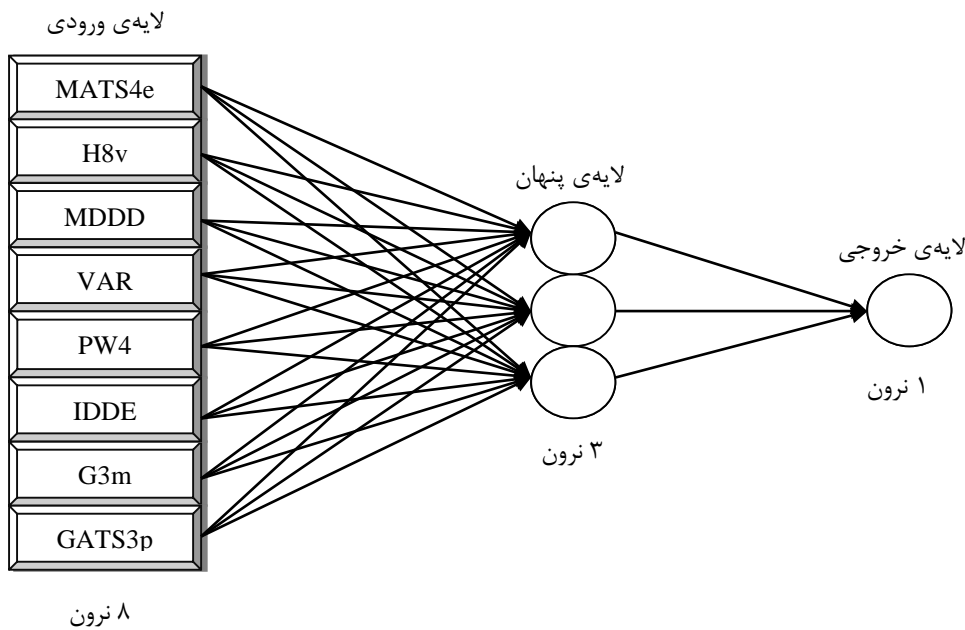
مقادیر μ	میانگین مربع خطا	مقادیر μ	میانگین مربع خطا
۰/۰۰۰۱	۰/۰۳۰۰	۰/۰۰۱۲	۰/۰۳۴۰
۰/۰۰۰۲	۰/۰۳۳۸	۰/۰۰۱۳	۰/۰۳۷۹
۰/۰۰۰۳	۰/۰۲۸۰	۰/۰۰۱۴	۰/۰۵۲۶
۰/۰۰۰۴	۰/۰۳۲۱	۰/۰۰۱۵	۰/۰۶۹۳
۰/۰۰۰۵	۰/۰۲۷۶	۰/۰۰۱۶	۰/۰۶۹۱
۰/۰۰۰۶	۰/۰۳۰۳	۰/۰۰۱۷	۰/۰۶۱۵
۰/۰۰۰۷	۰/۰۳۷۵	۰/۰۰۱۸	۰/۰۶۶۰
۰/۰۰۰۸	۰/۰۲۹۷	۰/۰۰۱۹	۰/۰۶۶۵
۰/۰۰۰۹	۰/۰۲۹۳	۰/۰۰۲۰	۰/۰۶۶۷
۰/۰۰۱۰	۰/۰۳۰۰		
۰/۰۰۱۱	۰/۰۳۱۴		



شکل (۴-۵) - نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ

۴-۴-۱-۴- ساختار شبکه عصبی مصنوعی بهینه شده

با توجه به روند بهینه‌سازی که در قسمت‌های قبلی ذکر شد، شبکه عصبی مصنوعی که بتواند فعالیت ترکیب مورد مطالعه را پیش‌بینی کند، دارای ۸ توصیف‌کننده، ۳ گره در لایه پنهان، تابع آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی است. شکل (۴-۶) ساختار شبکه عصبی بهینه نهایی را نشان می‌دهد.



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

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۴-۹)

ارائه شده است.

جدول (۴-۹) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

تنظیم بایزین (train br)	تابع آموزش
لگاریتم سیگموئید (logsig)	تابع انتقال لایه پنهان
purelin	تابع انتقال لایه خروجی
۳	تعداد نرون لایه پنهان (گره)
۸	تعداد متغیرهای ورودی (توصیف‌کننده‌ها)
۲۷	تعداد دوره‌های آموزش
۰/۰۰۰۵	پارامتر μ

۴-۱-۵- مدل سازی شبکه عصبی مصنوعی با استفاده از توصیف کننده های

انتخاب شده توسط الگوریتم ژنتیک

۴-۱-۵-۱- بهینه سازی پارامترهای شبکه

در این قسمت برای بهینه سازی تعداد ورودی های شبکه، نوع تابع آموزش، نوع تابع انتقال و تعداد گره های لایه پنهان، شبکه ای با تعداد ورودی های از ۲ تا ۷ توصیف کننده ایجاد شد و با الگوریتم های آموزشی تنظیم بایزین و لونبرگ مارکوارت با تعداد متفاوت گره های لایه پنهان (۲ تا ۱۰ گره) آموزش داده شد. همچنین تعداد دوره های آموزش نیز بطور همزمان از ۲ تا ۱۵ تغییر داده شد. برای انتخاب بهترین تابع انتقال در لایه پنهان از توابع تانژانت سیگموئیدی و لگاریتم سیگموئیدی به عنوان تابع انتقال استفاده شد. همچنین از تابع انتقال خطی در لایه خروجی استفاده گردید. در روند بهینه سازی شبکه، به حداقل رساندن مقدار میانگین مجموع مربعات خطا (MSE) به عنوان معیار انتخاب شد.

۴-۱-۵-۲- بهینه سازی همزمان تعداد متغیرهای ورودی شبکه، الگوریتم آموزش، نوع تابع

انتقال، تعداد گره های لایه های پنهان و تعداد دوره های آموزشی

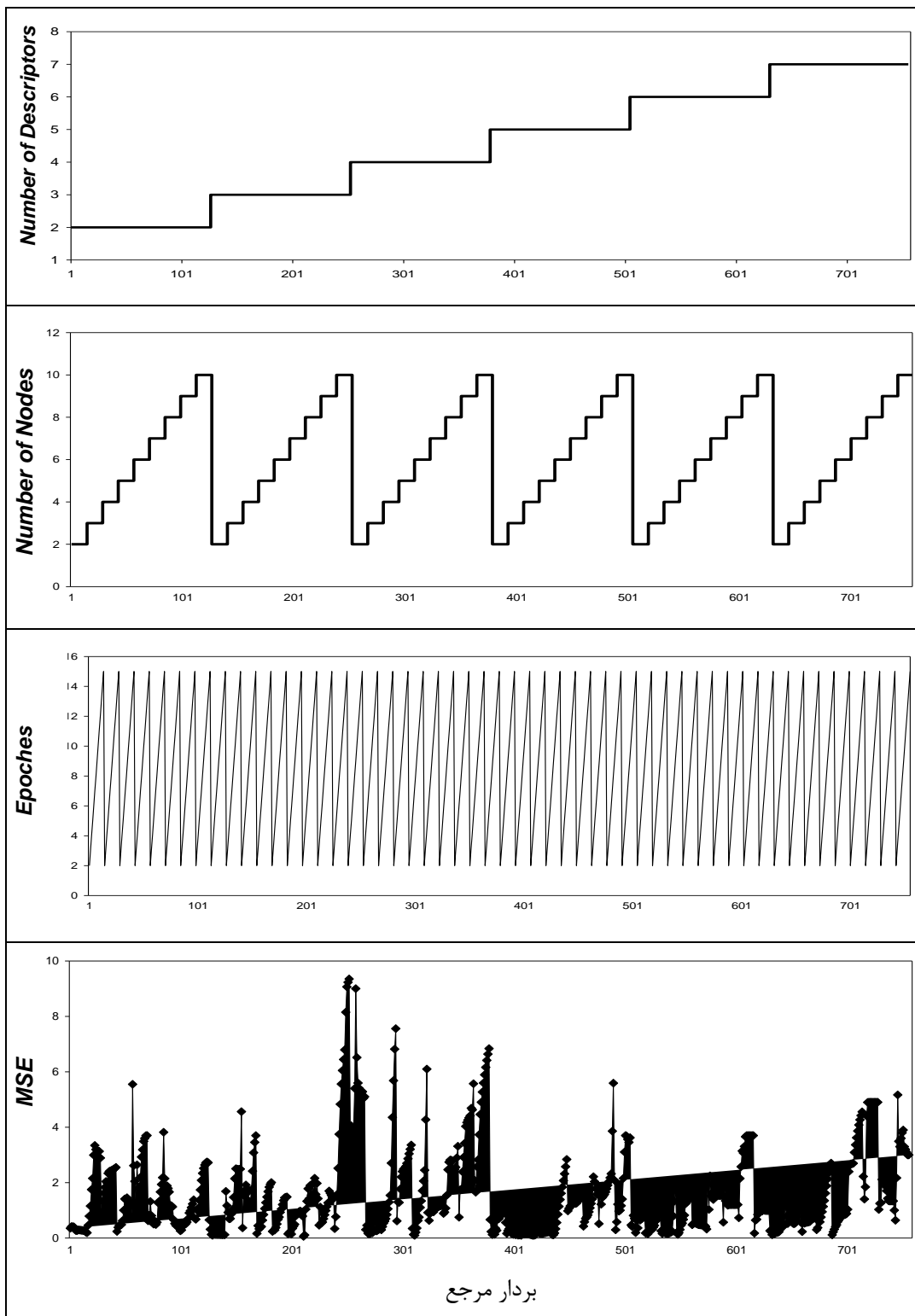
جهت بهینه سازی همزمان تمام پارامترها، در مجموع ۳۰۲۴ شبکه مورد ارزیابی قرار گرفته است که در زیر به تعداد تمام موارد مورد بررسی اشاره شده است.

$$۳۰۲۴ = \text{دوره های آموزش (۱۴)} \times \text{گره (۹)} \times \text{تابع انتقال (۲)} \times \text{تابع آموزش (۲)} \times \text{توصیف کننده (۶)}$$

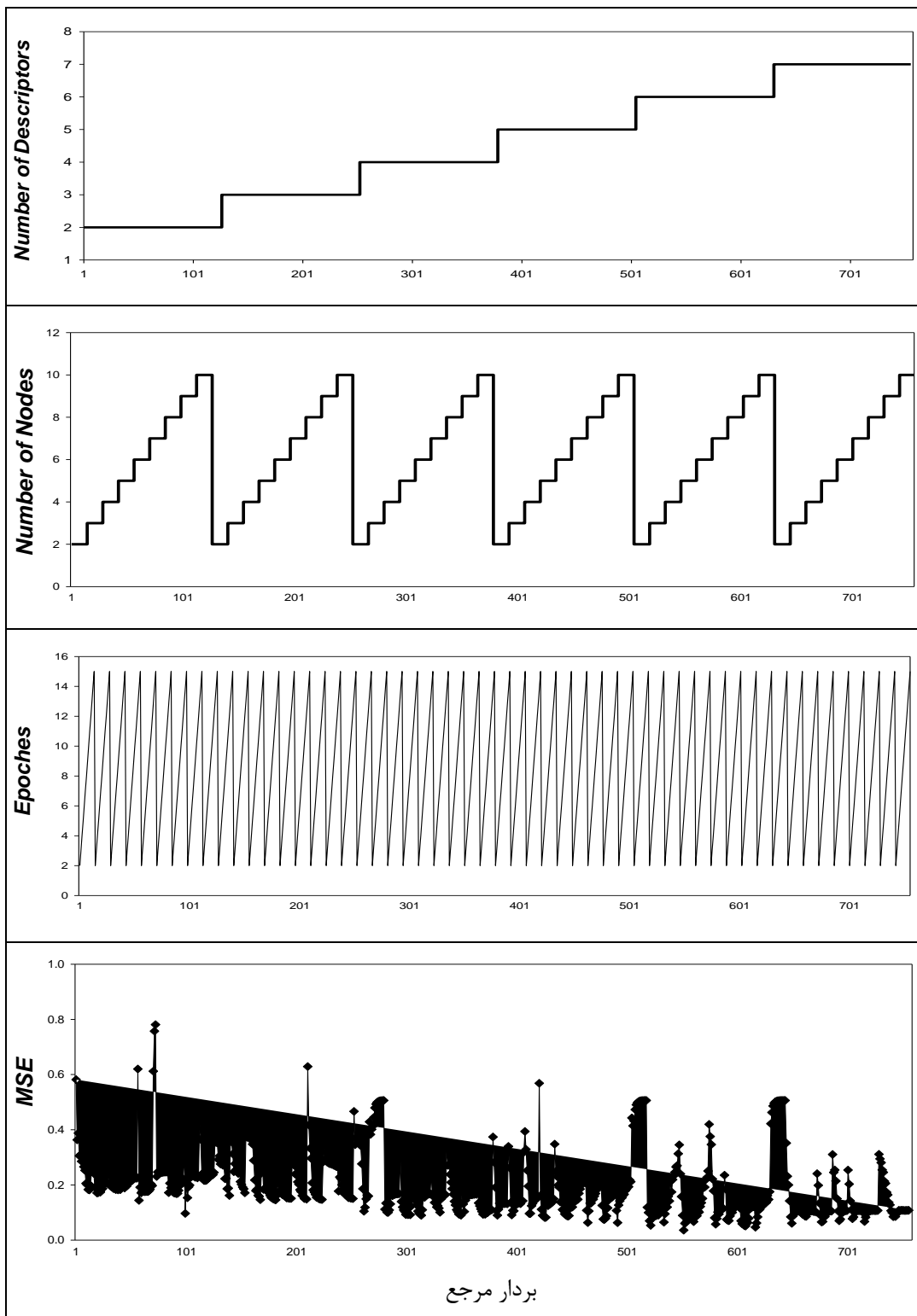
بخشی از روند تغییرات پارامترهای شبکه در حین بهینه سازی همزمان پارامترها به همراه

مقادیر MSE به دست آمده به صورت نموداری بر حسب یک بردار مرجع فرضی در شکل های (۴-۷)،

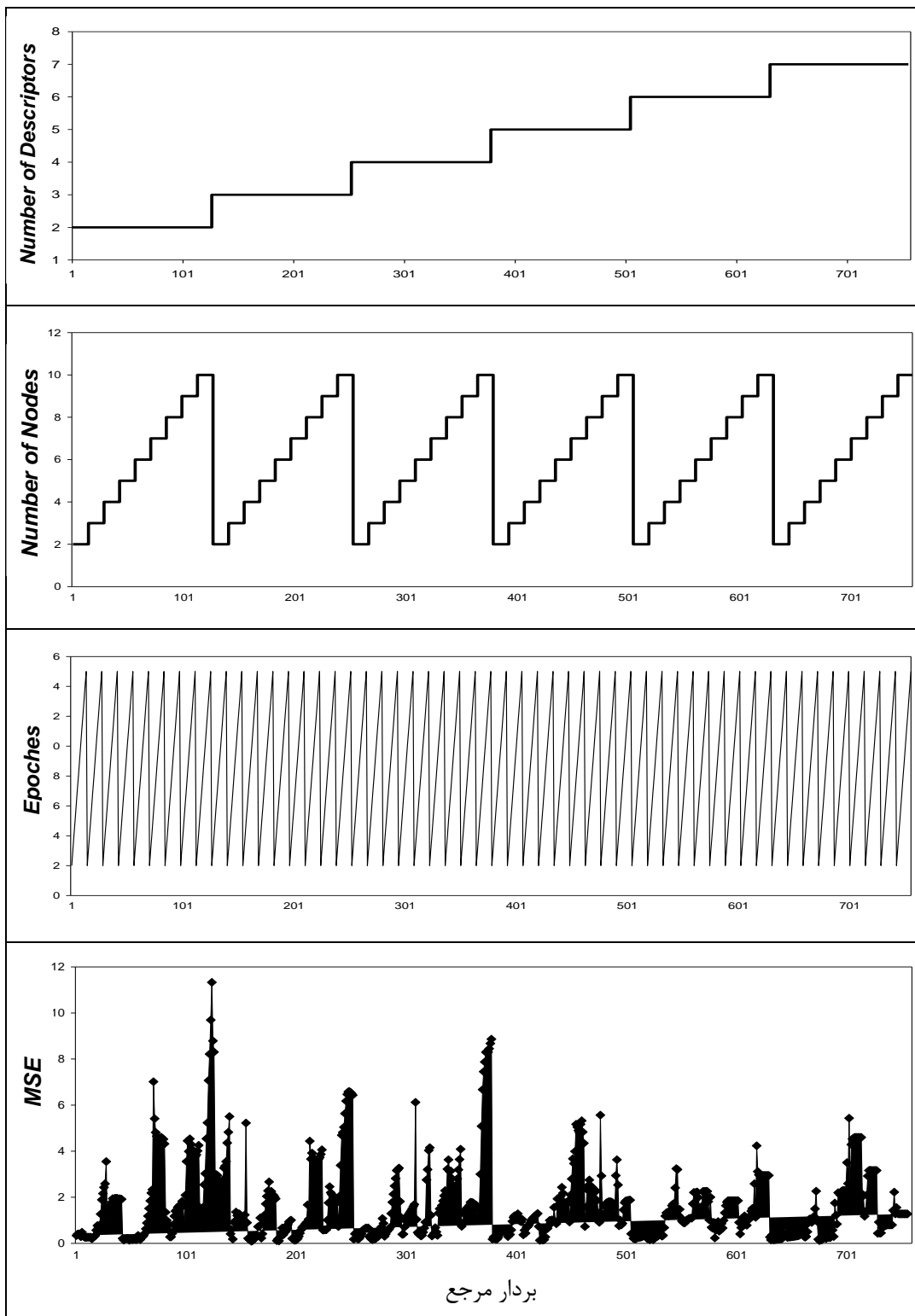
(۴-۸)، (۴-۹) و (۴-۱۰) آمده است.



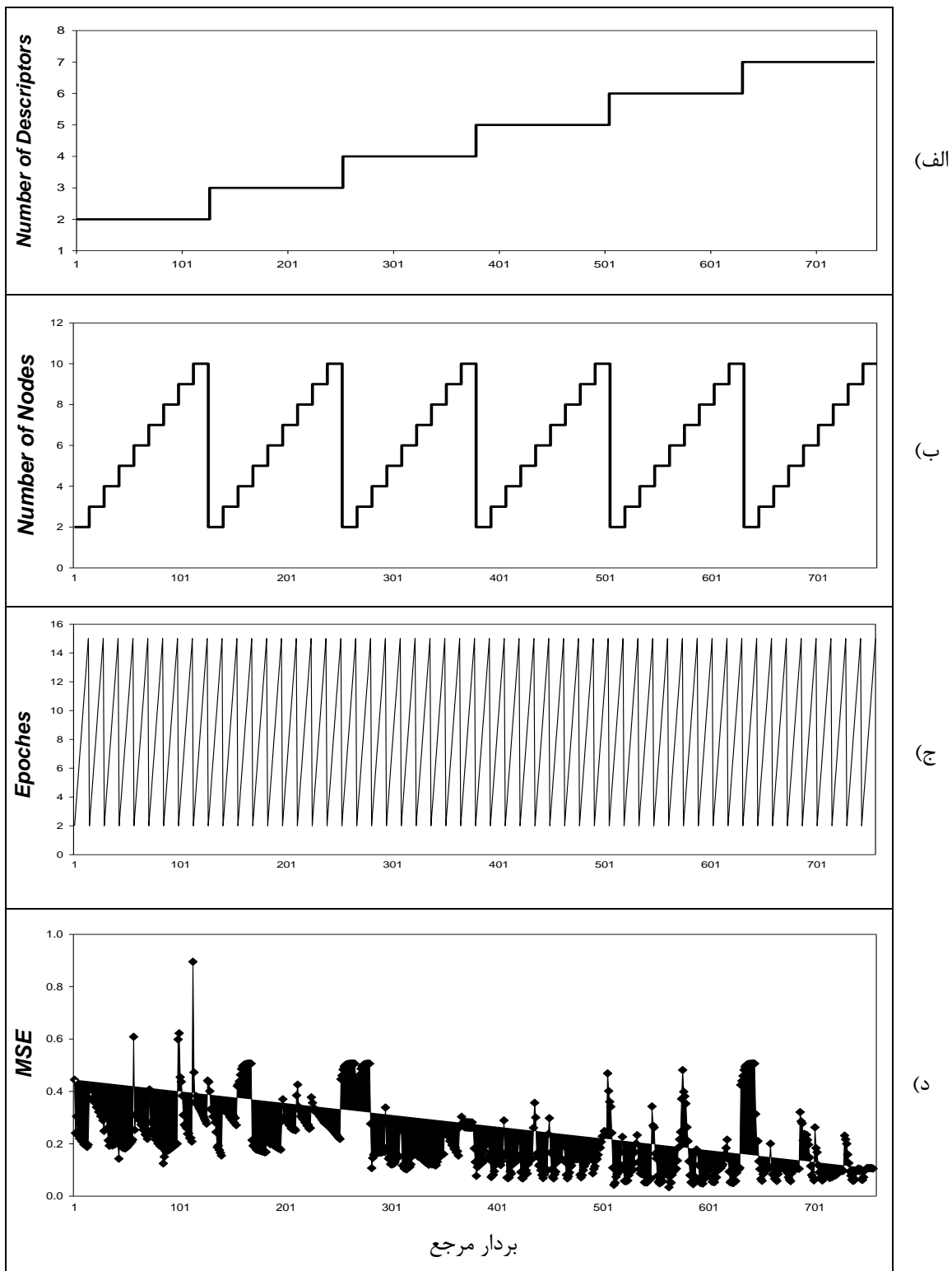
شکل (۴-۷) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید



شکل (۴-۸) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئید



شکل (۴-۹) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دوره‌های آموزش و د) مقادیر MSE برای تابع آموزش لونیبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید



شکل (۴-۱۰) - نمودارهای الف) تعداد توصیف‌کننده‌ها، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئید

جدول (۴-۱۰) - توابع و پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا

MSE	تعداد دوره‌های آموزش	تعداد گره	تعداد توصیف کننده‌ها	تابع آموزش	تابع انتقال
۰/۰۵۹۴	۲	۸	۳	لونبرگ - مارکوارت	لگاریتم سیگموئید
۰/۰۳۵۹	۷	۵	۶	تنظیم بایزین	لگاریتم سیگموئید
۰/۰۹۱۷	۷	۴	۳	لونبرگ - مارکوارت	تانژانت سیگموئید
۰/۰۳۳۵	۴	۶	۶	تنظیم بایزین	تانژانت سیگموئید

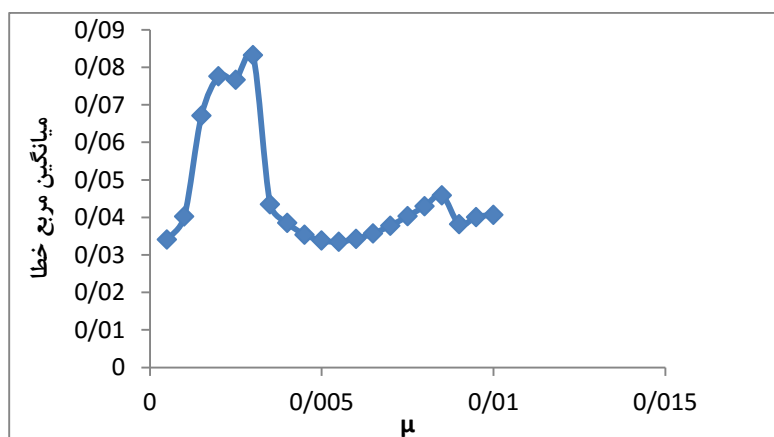
با توجه به نتایج، تابع آموزش تنظیم بایزین نسبت به لونبرگ- مارکوارت دارای میانگین مربعات خطای کمتری است. همچنین تابع انتقال تانژانت سیگموئید به همراه تابع آموزش تنظیم بایزین دارای خطای کمتری نسبت به لگاریتم سیگموئید می‌باشند، لذا به عنوان توابع آموزش و انتقال بهینه انتخاب گردیدند. همچنین طبق نتایج بدست آمده، تعداد ورودی‌های شبکه، تعداد گره در لایه‌ی پنهان و تعداد دوره‌های آموزشی نیز که بطور همزمان بهینه شدند به ترتیب برابر با ۶ متغیر ورودی، ۶ گره در لایه‌ی پنهان و تعداد دوره‌های آموزش برابر ۴ بدست آمد.

۴-۱-۵-۳- بهینه‌سازی پارامتر μ

جهت بهینه کردن مقدار μ ، ساختار شبکه بهینه با ۶ متغیر ورودی، ۶ گره در لایه پنهان، تعداد دوره‌های آموزش برابر ۴، تابع آموزش تانژانت سیگموئید و تابع آموزشی تنظیم بایزین در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۰۵ تا ۰/۰۱ (حداقل میانگین مربعات خطا در این رنج مشاهده گردید) با گام های ۰/۰۰۰۵ تغییر داده شد. سپس برای هر مورد مقدار میانگین مربع خطای سری ارزیابی محاسبه گردید. در نهایت مقدار میانگین مربع خطا بر حسب μ رسم گردید. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد که طبق جدول (۴-۱۱) و شکل (۴-۱۱) مقدار بهینه ۰/۰۰۵۵ به دست آمد.

جدول (۴-۱۱)- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ

مقادیر μ	میانگین مربع خطا	مقادیر μ	میانگین مربع خطا
۰/۰۰۰۵	۰/۰۳۴۱	۰/۰۰۶۰	۰/۰۳۴۳
۰/۰۰۱۰	۰/۰۴۰۲	۰/۰۰۶۵	۰/۰۳۵۷
۰/۰۰۱۵	۰/۰۶۷۱	۰/۰۰۷۰	۰/۰۳۷۸
۰/۰۰۲۰	۰/۰۷۷۶	۰/۰۰۷۵	۰/۰۴۰۳
۰/۰۰۲۵	۰/۰۷۶۷	۰/۰۰۸۰	۰/۰۴۳۰
۰/۰۰۳۰	۰/۰۸۳۲	۰/۰۰۸۵	۰/۰۴۵۹
۰/۰۰۳۵	۰/۰۴۳۵	۰/۰۰۹۰	۰/۰۳۸۲
۰/۰۰۴۰	۰/۰۳۸۵	۰/۰۰۹۵	۰/۰۴۰۱
۰/۰۰۴۵	۰/۰۳۵۴	۰/۰۱۰۰	۰/۰۴۰۷
۰/۰۰۵۰	۰/۰۳۳۵		
۰/۰۰۵۵	۰/۰۳۳۲		



شکل (۴-۱۱)- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب μ

۴-۱-۵-۴- ساختار شبکه عصبی مصنوعی بهینه شده

باتوجه به نتایج به دست آمده در بهینه‌سازی شبکه، مشخصات شبکه بهینه در جدول (۴-۱۲)

ارائه شده است.

جدول (۴-۱۲)- مشخصات شبکه بهینه

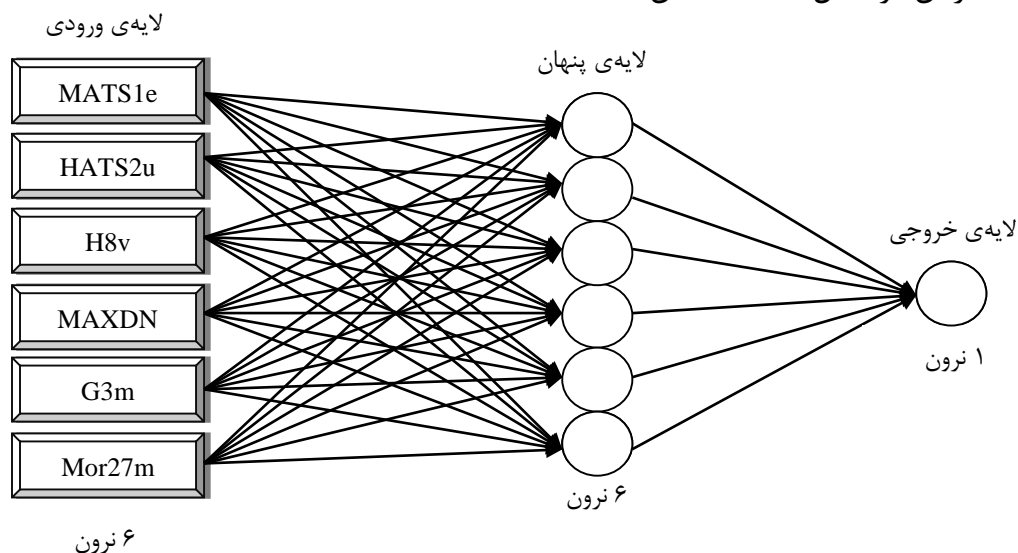
تابع آموزش	تنظیم بایزین (train br)
تابع انتقال لایه پنهان	تانژانت سیگموئید (tansig)
تابع انتقال لایه خروجی	purelin
تعداد نرون لایه پنهان (گره)	۶
تعداد متغیرهای ورودی (توصیف‌کننده‌ها)	۶
تعداد دوره‌های آموزش	۴
پارامتر μ	۰/۰۰۵۵

۴-۱-۵-۵- معماری شبکه عصبی بهینه شده

با توجه به نتایج به دست آمده در بهینه‌سازی شبکه، شبکه عصبی سه لایه پس انتشار با

الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی، به عنوان شبکه بهینه انتخاب شد

که ساختار آن در شکل (۴-۱۲) نشان داده شده است.



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

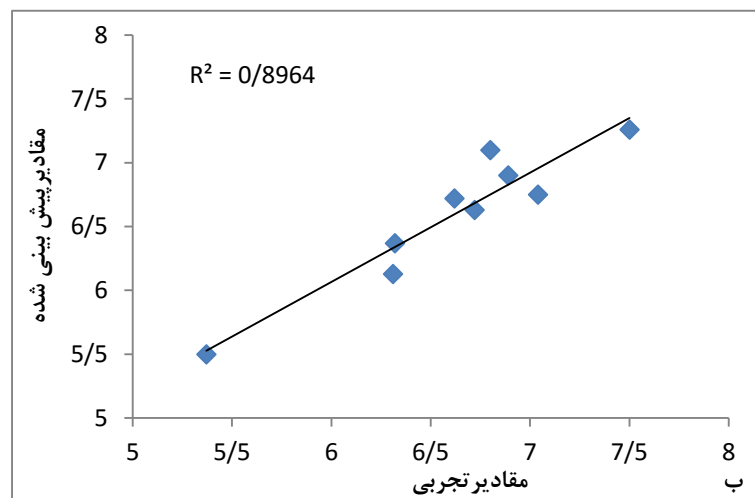
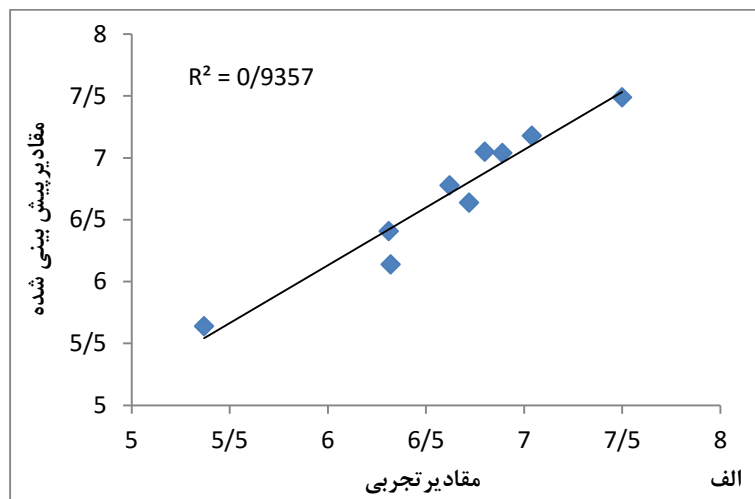
۴-۱-۶- ارزیابی مدل‌های SR-ANN و GA-ANN

۴-۱-۶-۱- ارزیابی مدل‌های SR-ANN و GA-ANN با استفاده از سری‌های ارزیابی و تست

به منظور مطالعه‌ی قدرت پیش‌بینی مدل‌های SR-ANN و GA-ANN ، سری‌های ارزیابی و تست مورد استفاده قرار گرفتند. بدین ترتیب که کل داده‌ها به سه قسمت با نسبت‌های ۲۰٪، ۲۰٪ و ۶۰٪ تحت عنوان سری‌های ارزیابی، تست و آموزش تقسیم شدند. در این راستا، با استفاده از مدل‌های بدست آمده مقادیر pIC_{50} این ترکیبات پیش‌بینی شدند. نتایج حاصل از این بررسی که مجذور ضریب همبستگی را برای سری ارزیابی و تست (شکل‌های (۴-۱۳) و (۴-۱۴)) توسط دو مدل SR-ANN و GA-ANN نشان می‌دهد، بیانگر توانایی بالاتر روش SR-ANN نسبت به روش GA-ANN در پیش‌بینی فعالیت دارویی این ترکیبات می‌باشد.

جدول (۴-۱۳) - نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری ارزیابی

شماره ترکیب	pIC ₅₀				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۲	۶/۷۲	۶/۶۴	۶/۶۳	-۱/۱۹	-۱/۳۴
۶	۶/۳۱	۶/۴۱	۶/۱۳	۱/۵۸	-۲/۸۵
۱۱	۶/۸۰	۷/۰۵	۷/۱۰	۳/۶۸	۴/۴۱
۱۶	۶/۳۲	۶/۱۴	۶/۳۷	-۲/۸۵	۰/۷۹
۲۰	۷/۰۴	۷/۱۸	۶/۷۵	۱/۹۹	-۴/۱۲
۲۴	۶/۶۲	۶/۷۸	۶/۷۲	۲/۴۲	۱/۵۱
۲۹	۷/۵۰	۷/۴۹	۷/۲۶	-۰/۱۳	-۳/۲۰
۳۴	۶/۸۹	۷/۰۴	۶/۹۰	۲/۱۸	۰/۱۵
۳۹	۵/۳۷	۵/۶۴	۵/۵۰	۵/۰۳	۲/۴۲

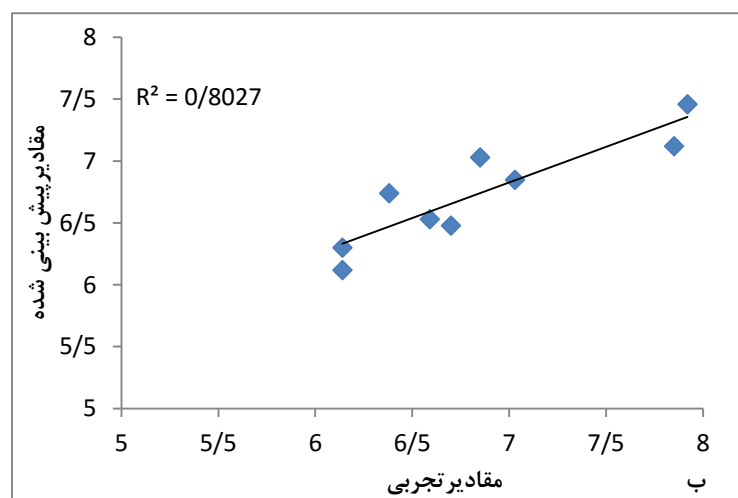
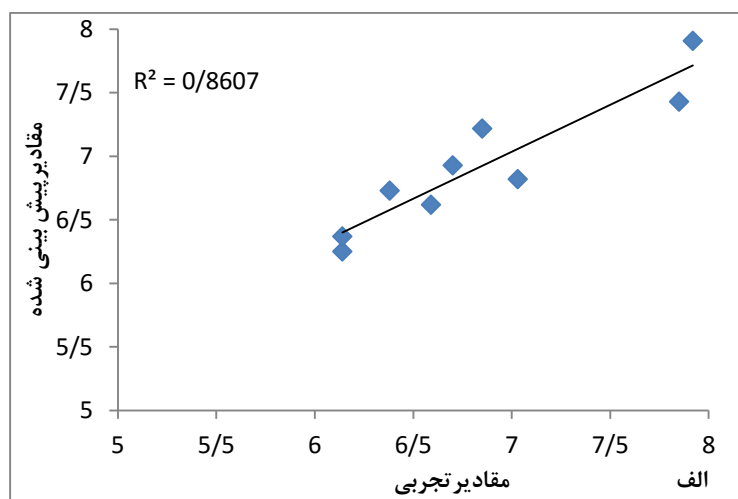


شکل (۴-۱۳) - نمودار مقادیر پیش‌بینی شده pIC₅₀ بر حسب مقادیر تجربی برای سری ارزیابی

الف: با مدل SR-ANN ب: با مدل GA-ANN

جدول (۴-۱۴) - نتایج حاصل از ارزیابی مدل‌های SR-ANN و GA-ANN برای سری تست

شماره ترکیب	مقدار تجربی	pIC ₅₀			
		مقدار پیش بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۳	۷/۰۳	۶/۸۲	۶/۸۵	-۲/۹۹	-۲/۵۶
۸	۶/۱۴	۶/۲۵	۶/۱۲	۱/۷۹	-۰/۳۳
۱۳	۶/۱۴	۶/۳۷	۶/۳۰	۳/۷۵	۲/۶۱
۱۸	۷/۹۲	۷/۹۱	۷/۴۶	-۰/۱۳	-۵/۸۱
۲۳	۶/۳۸	۶/۷۳	۶/۷۴	۵/۴۹	۵/۶۴
۲۸	۶/۵۹	۶/۶۲	۶/۵۳	۰/۴۶	-۰/۹۱
۳۲	۷/۸۵	۷/۴۳	۷/۱۲	-۵/۳۵	-۹/۳۰
۳۷	۶/۷۰	۶/۹۳	۶/۴۸	۳/۴۳	-۳/۲۸
۴۱	۶/۸۵	۷/۲۲	۷/۰۳	۵/۴۰	۲/۶۳



شکل (۴-۱۴) - نمودار مقادیر پیش بینی شده pIC₅₀ بر حسب مقادیر تجربی برای سری تست

الف: با مدل SR-ANN ب: با مدل GA-ANN

۴-۱-۶-۲- ارزیابی مدل‌های SR-ANN و GA-ANN توسط روش رد مرحله‌ای تک تک

از دیگر روش‌هایی که برای ارزیابی مدل‌های برتر ارائه شده توسط SR-ANN و GA-ANN به کار گرفته شد، تکنیک رد مرحله‌ای تک تک (برای کل داده‌ها) می‌باشد که نتایج حاصل از این روش در جدول (۴-۱۵) ارائه شده است. پس از رسم منحنی مقدار تجربی pIC_{50} بر حسب مقدار پیش‌بینی شده‌ی آن (شکل (۴-۱۵))، مقادیر R^2 برای مدل‌های SR-ANN و GA-ANN به ترتیب $0/8613$ و $0/8005$ بدست آمد، که این نشان‌دهنده‌ی برتری نسبی مدل SR-ANN به GA-ANN می‌باشد.

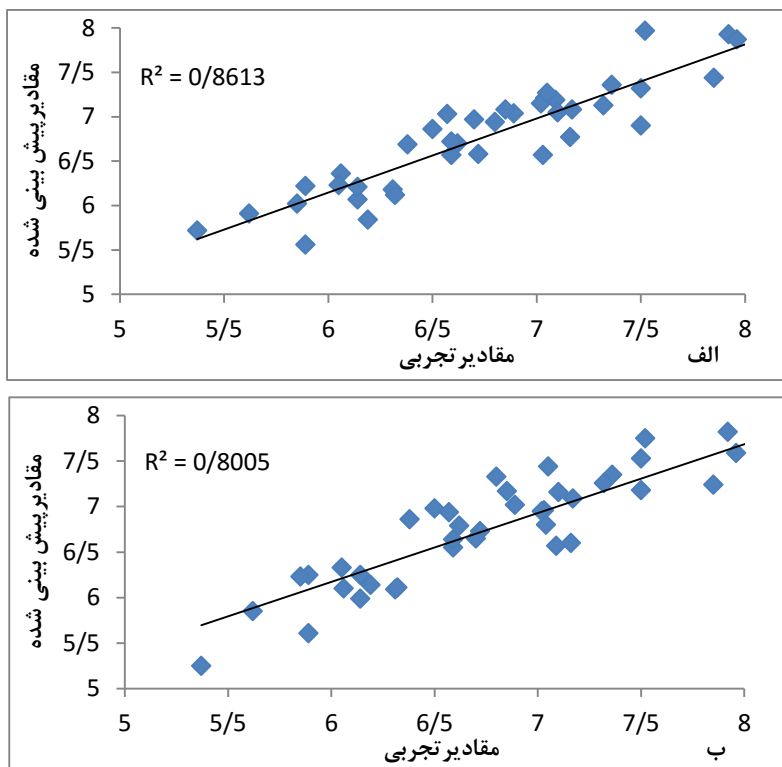
همچنین در شکل (۴-۱۶) مقدار باقیمانده‌ها بر حسب مقادیر تجربی pIC_{50} ترکیبات مورد مطالعه، ترسیم شده است. اگر باقیمانده‌ها به طور یکنواخت حول محور افقی پراکنده باشند، نشان می‌دهد که مدل مناسبی به دست آمده و هیچ خطای سیستماتیکی وجود ندارد و همانطور که از این شکل‌ها مشخص است تقارن پراکندگی نقاط در دو طرف محور افقی عدم وجود خطای سیستماتیک را نشان می‌دهد.

جدول (۴-۱۵) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط SR-ANN و GA-ANN با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها

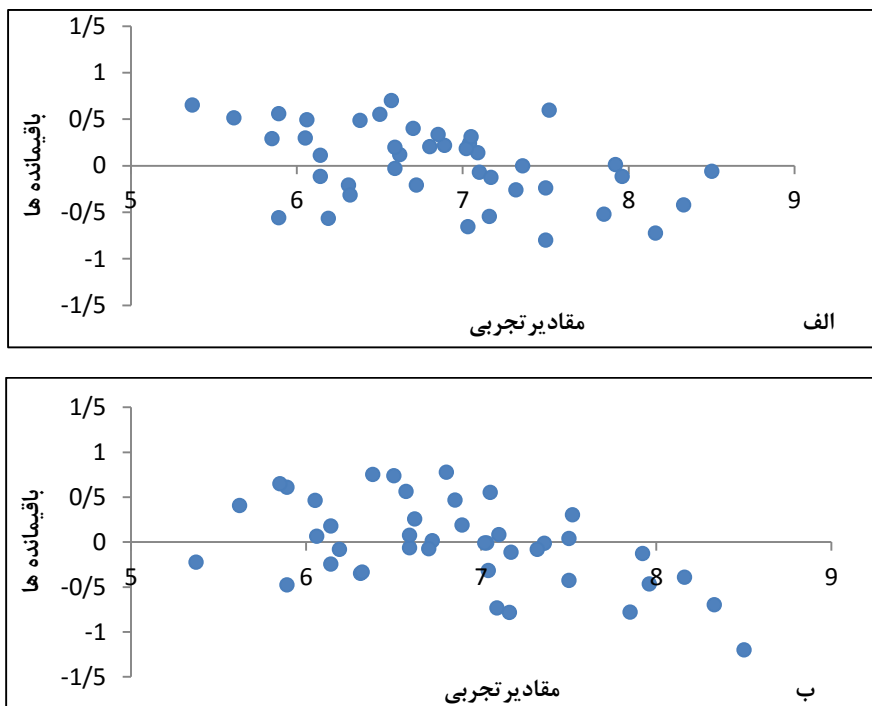
شماره ترکیب	pIC ₅₀				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۱	۸/۱۶	۷/۵۷	۷/۸۴	-۷/۲۳	-۳/۹۲
۲	۶/۷۲	۶/۵۸	۶/۷۳	-۲/۰۸	۰/۱۵
۳	۷/۰۳	۶/۵۷	۶/۹۶	-۶/۵۴	-۱/۰۰
۴	۶/۰۶	۶/۳۶	۶/۱۰	۴/۹۵	۰/۶۶
۵	۶/۰۵	۶/۲۳	۶/۳۳	۲/۹۸	۴/۶۳
۶	۶/۳۱	۶/۱۸	۶/۰۹	-۲/۰۶	-۳/۴۹
۷	۸/۵۰	۸/۴۵	۷/۴۸	-۰/۵۹	-۱۲/۰۰
۸	۶/۱۴	۶/۰۷	۵/۹۹	-۱/۱۴	-۲/۴۴
۹	۸/۳۳	۷/۹۸	۷/۷۵	-۴/۲۰	-۶/۹۶
۱۰	۵/۸۹	۶/۲۲	۵/۶۱	۵/۶۰	-۴/۷۵
۱۱	۶/۸۰	۶/۹۴	۷/۳۳	۲/۰۶	۷/۷۹
۱۲	۵/۸۹	۵/۵۶	۶/۲۵	-۵/۶۰	۶/۱۱
۱۳	۶/۱۴	۶/۲۱	۶/۲۵	۱/۱۴	۱/۷۹
۱۴	۶/۵۰	۶/۸۶	۶/۹۸	۵/۵۴	۷/۳۸
۱۵	۶/۱۹	۵/۸۴	۶/۱۴	-۵/۶۵	-۰/۸۱
۱۶	۶/۳۲	۶/۱۲	۶/۱۱	-۳/۱۶	-۳/۳۲
۱۷	۵/۸۵	۶/۰۲	۶/۲۳	۲/۹۱	۶/۵۰
۱۸	۷/۹۲	۷/۹۳	۷/۸۲	۰/۱۳	-۱/۲۶
۱۹	۷/۳۶	۷/۳۶	۷/۳۵	۰/۰۰	-۰/۱۴
۲۰	۷/۰۴	۷/۲۱	۶/۸۰	۲/۴۱	-۳/۴۱
۲۱	۷/۰۹	۷/۱۹	۶/۵۷	۱/۴۱	-۷/۳۳

ادامه جدول (۴-۱۵)

شماره ترکیب	pIC ₅₀				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		SR-ANN	GA-ANN	SR-ANN	GA-ANN
۲۲	۵/۶۲	۵/۹۱	۵/۸۵	۵/۱۶	۴/۰۹
۲۳	۶/۳۸	۶/۶۹	۶/۸۶	۴/۸۶	۷/۵۲
۲۴	۶/۶۲	۶/۷۰	۶/۷۹	۱/۲۱	۲/۵۷
۲۵	۷/۱۶	۶/۷۷	۶/۶۰	-۵/۴۵	-۷/۸۲
۲۶	۷/۳۲	۷/۱۳	۷/۲۶	-۲/۶۰	-۰/۸۲
۲۷	۷/۱۰	۷/۰۵	۷/۱۶	-۰/۷۰	۰/۸۵
۲۸	۶/۵۹	۶/۵۷	۶/۵۵	-۰/۳۰	-۰/۶۱
۲۹	۷/۵۰	۷/۳۲	۷/۵۳	-۲/۴۰	۰/۴۰
۳۰	۷/۵۰	۶/۹۰	۷/۱۸	-۸/۰۰	-۴/۲۷
۳۱	۷/۹۶	۷/۸۷	۷/۵۹	-۱/۱۳	-۴/۶۵
۳۲	۷/۸۵	۷/۴۴	۷/۲۴	-۵/۲۲	-۷/۷۷
۳۳	۷/۵۲	۷/۹۷	۷/۷۵	۵/۹۸	۳/۰۶
۳۴	۶/۸۹	۷/۰۴	۷/۰۲	۲/۱۸	۱/۸۹
۳۵	۷/۰۲	۷/۱۵	۶/۹۵	۱/۸۵	-۱/۰۰
۳۶	۷/۰۵	۷/۲۷	۷/۴۴	۳/۱۲	۵/۵۳
۳۷	۶/۷۰	۶/۹۷	۶/۶۵	۴/۰۳	-۰/۷۵
۳۸	۷/۱۷	۷/۰۸	۷/۰۹	-۱/۲۶	-۱/۱۲
۳۹	۵/۳۷	۵/۷۲	۵/۲۵	۶/۵۲	-۲/۲۳
۴۰	۶/۵۷	۷/۰۳	۶/۹۴	۷/۰۰	۵/۶۳
۴۱	۶/۸۵	۷/۰۸	۷/۱۷	۳/۳۶	۴/۶۷
۴۲	۶/۵۹	۶/۷۲	۶/۶۴	۱/۹۷	۰/۷۶



شکل (۴-۱۵) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای تک تک برای کل داده‌ها الف: با مدل SR-ANN ب: با مدل GA-ANN



شکل (۴-۱۶) - نمودار باقیمانده‌ها بر حسب مقدار تجربی pIC_{50} برای کل داده‌ها الف: با مدل SR-ANN ب: با مدل GA-ANN

۴-۱-۷- ارزیابی مدل‌های برتر SR-ANN و GA-ANN با استفاده از پارامترهای

آماري

در راستای بررسی کارایی مدل‌های طراحی شده توسط SR-ANN و GA-ANN چندین پارامتر آماری محاسبه گردید که نتایج آن در جدول (۴-۱۶) ارائه شده است. مقادیر این پارامترها نشان می‌دهد که مدل‌های ارائه شده از قدرت پیش‌بینی مناسبی برخوردار هستند. نتایج همچنین گویای برتری نسبی روش SR-ANN نسبت به GA-ANN در پیش‌بینی pIC_{50} سری داده‌ها می‌باشد.

جدول (۴-۱۶) - پارامترهای آماری برای مدل‌های SR-ANN و GA-ANN

پارامتر	سری تست (N=9)		سری ارزیابی (N=9)		کل داده‌ها (N=42)	
	SR-ANN	GA-ANN	SR-ANN	GA-ANN	SR-ANN	GA-ANN
MAE	۰/۲۱۷۸	۰/۲۶۳۳	۰/۱۴۸۹	۰/۱۵۴۴	۰/۲۲۸۳	۰/۲۵۴۸
MSE	۰/۰۶۶۵	۰/۱۱۳۰	۰/۰۲۸۰	۰/۰۳۳۵	۰/۰۷۵۵	۰/۱۱۰۳
PRESS	۰/۵۹۸۵	۱/۰۱۷۰	۰/۲۵۲۰	۰/۳۰۱۵	۳/۱۷۱۰	۴/۶۳۲۶
SEP	۰/۲۵۷۹	۰/۳۳۶۲	۰/۱۶۷۳	۰/۱۸۳۰	۰/۲۷۴۸	۰/۳۳۲۱
R ²	۰/۸۶۰۷	۰/۸۰۲۷	۰/۹۳۵۷	۰/۸۹۶۴	۰/۸۶۱۳	۰/۸۰۰۵
REP (%)	۳/۷۷۰۵	۴/۹۱۲۳	۲/۵۲۷۲	۲/۷۶۴۴	۴/۰۱۱۷	۴/۸۴۸۲
MRE	۳/۱۹۷۸	۳/۶۷۳۳	۲/۳۳۷۸	۲/۴۱۰۰	۳/۳۷۳۶	۳/۶۶۳۱

۴-۱-۸- ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون Y- تصادفی

برای اثبات قدرت مدل‌های ایجاد شده و اینکه نتایج به دست آمده تصادفی نبوده است، آزمون Y- تصادفی انجام گرفت. در این راستا pIC_{50} (متغیر وابسته) به صورت تصادفی ایجاد شد. شبکه عصبی بهینه شده با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر وابسته توسعه یافت. نتایج حاصل از چندین بار اجرای این آزمون در جداول (۴-۱۷) و (۴-۱۸) آورده شده که مقادیر ضریب تعیین کوچک، نشان‌دهنده‌ی عدم وابستگی در مدل‌های توسعه یافته توسط شبکه عصبی مصنوعی است.

جدول (۴-۱۷) - مقادیر R^2 برای سری تست و ارزیابی مدل SR-ANN با استفاده از آزمون Y- تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2 سری تست	۰/۰۲۵	۰/۰۱۱	۰/۰۸۰	۰/۱۵۸	۰/۰۰۱	۰/۰۵۹	۰/۰۲۳	۰/۲۱۶	۰/۰۸۹	۰/۰۲۵
R^2 سری ارزیابی	۰/۱۳۰	۰/۰۸۷	۰/۰۳۵	۰/۰۱۸	۰/۰۴۱	۰/۱۸۷	۰/۰۵۴	۰/۰۳۲	۰/۰۰۵	۰/۰۴۰

جدول (۴-۱۸) - مقادیر R^2 برای سری تست و ارزیابی مدل GA-ANN با استفاده از آزمون Y- تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2 سری تست	۰/۱۴۲	۰/۱۰۰	۰/۲۰۱	۰/۰۲۷	۰/۲۰۹	۰/۰۱۴	۰/۰۷۲	۰/۰۲۳	۰/۰۴۱	۰/۰۳۷
R^2 سری ارزیابی	۰/۰۴۵	۰/۱۲۰	۰/۰۲۰	۰/۰۷۱	۰/۰۰۷	۰/۲۱۱	۰/۰۹۰	۰/۲۷۳	۰/۰۰۸	۰/۰۳۴

۴-۱-۹- بررسی ارتباط توصیف‌کننده‌های منتخب

در این بخش به بررسی ارتباط بین توصیف‌کننده‌های منتخب می‌پردازیم و سعی می‌کنیم تا حد امکان توضیحی برای ارتباط بین این توصیف‌کننده‌ها و فعالیت مورد نظر بیابیم. همچنین اثرات متوسط توصیف‌کننده‌های منتخب بدست آمده از روش رگرسیون مرحله‌ای در جدول (۴-۱۹) گزارش شده است.

جدول (۴-۱۹) - اثر متوسط توصیف‌کننده‌های بدست آمده توسط روش رگرسیون مرحله‌ای

توصیف‌کننده	MATS4e	H8v	MDDD	VAR	PW4	IDDE	G3m	GATS3p
اثر متوسط	-۶/۶۵۶	-۸/۲۵۰	-۰/۳۵۵	۰/۰۳۶	۵/۷۲۵	۱/۶۰۹	۱/۷۷۴	۲/۷۴۴

الف - توصیف‌کننده‌های 2D Autocorrelation

همانطور که در بخش (۳-۱-۱۰) اشاره شد، توصیف‌کننده‌های 2D Autocorrelation، در برگیرنده اطلاعات دو بعدی مولکول می‌باشند. Autocorrelation Geary یک نوع از این توصیف‌کننده‌هاست که بدین صورت محاسبه می‌شود:

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

w یک ویژگی اتم، \bar{w} میانگین مقدار آن ویژگی روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است. δ_{ij} نیز به تابع کرونگر^۱ معروف است و Δ هم مجموعه‌ی δ هاست. مقدار این توصیف‌کننده، از صفر تا بی‌نهایت متغیر می‌باشد. از این گروه توصیف‌کننده‌ها، MATS4e^۲ و GATS3p^۳ وارد مدل SR-ANN شده‌اند که به ترتیب دارای ضرایب منفی و مثبت می‌باشند و نقش خاصیت فیزیکیولوژی الکترونگاتیویته‌ی ساندرسون را بیان می‌کنند.

ب- توصیف‌کننده‌های GETAWAY

این توصیف‌کننده‌ها که بیان‌کننده ویژگی‌های توپولوژیکی و هندسی مولکول هستند، از ماتریس تأثیر مولکولی (MIM)^۴ محاسبه می‌شوند. این ماتریس به آسانی از مختصات فضایی اتم‌های مولکول در یک صورتبندی مشخص به دست می‌آید که با رابطه زیر بیان می‌گردد:

$$H = M (M^T M) \cdot M^T \quad (۲-۴)$$

که M ماتریس مختصات اتمی، بالانویس T به معنای ماتریس ترانپوز^۵ و H ماتریس تأثیر مولکولی است. عناصر قطری (h_{ii}) ماتریس H لوریج^۶ نام دارند که هر یک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. اولین سری از این توصیف‌کننده‌های 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:

که h_{ii} و h_{jj} لوریج (قدرت نفوذ) دو اتم i و j و r_{ij} فاصله آن دو اتم است [۷۰]. از این گروه

توصیف‌کننده‌ها، H8v وارد مدل SR-ANN شده است. این توصیف‌کننده‌ها از آن جهت مهم هستند

^۱ - Kronecker

^۲ - Moran autocorrelation – lag4

^۳ - Geary autocorrelation – lag3

^۴ - Molecular Influence Matrix

^۵ - Transpose matrix

^۶ - Leverage

که به ساختار سه‌بعدی مولکول حساس هستند و معمولاً برای مقایسه مولکول‌ها و حتی صورتبندی‌ها از نظر شکل، تقارن، اندازه اتم به کار می‌روند. این توصیف‌کننده دارای علامت منفی در مدل است و نشان می‌دهد با افزایش این مقدار pIC_{50} کاهش پیدا می‌کند.

پ- توصیف‌کننده‌های Topological

این دسته توصیف‌کننده‌ها که از گراف مولکولی تهی از هیدروژن محاسبه می‌شوند، از ساده‌ترین توصیف‌کننده‌های مولکولی بوده و به ساختار فضایی مولکول ارتباطی ندارند. این توصیف‌کننده‌ها اطلاعاتی راجع به ساختمان، اندازه، شکل، تقارن، شاخه‌دار شدن، نحوه اتصال اتم‌ها و نوع اتم‌های موجود در یک مولکول را در اختیار ما قرار می‌دهند [۷۱]. های IDDE^۱، VAR^۲ و PW₄^۳ دارای ضرایب مثبت و توصیف‌کننده‌ی MDDD^۴ دارای ضریب منفی می‌باشد. ضریب مثبت و ضریب منفی به ترتیب نشان‌دهنده‌ی ارتباط مستقیم و عکس این توصیف‌کننده‌ها با فعالیت ترکیبات می‌باشد.

ت- توصیف‌کننده‌های WHIM^۵

این شاخص از مختصات کارتیزین ساختار سه‌بعدی مولکول، با استفاده از صورتبندی با حداقل انرژی محاسبه می‌شود و شامل اطلاعاتی درباره‌ی اندازه، شکل، تقارن و توزیع اتمی ساختار سه‌بعدی مولکول می‌باشد. این توصیف‌کننده از رابطه (۳-۴) به دست می‌آید:

$$S_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_j) (q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad (3-4)$$

¹ - Mean information on the distance degree equality

² - Variation

³ - Path/walk4- Randic shape index

⁴ - Mean Distance Degree Deviation

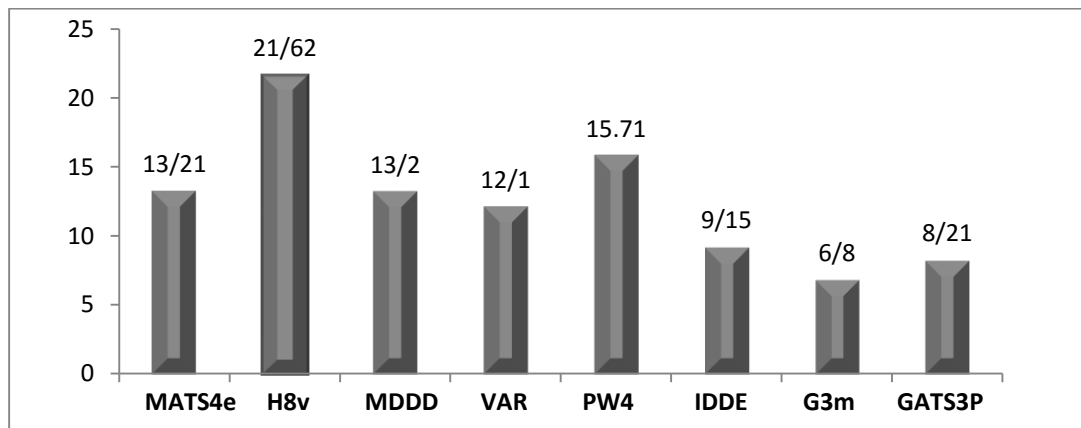
⁵ - Weighted Holistic Invariant Molecular descriptors

که S_{jk} کوواریانس وزن دار شده بین کئوردینه زام و k ام، A تعداد اتم‌ها، w_i وزن i امین اتم، q_{ij} و q_{ik} به ترتیب زامین و k امین کئوردینه‌های اتم i ام، و \bar{q} مقدار میانگین متناظر است. شش طرح وزن دار شدن پیشنهاد شده است که عبارتند از:

- ۱- حالت بدون وزن (u) - وزن دار شده با جرم اتمی ۳- وزن دار شده با حجم و اندروالس ۴-
- وزن دار شده با الکترون‌گاتیویته ساندرسون ۵- وزن دار شده با حالت الکتروتوپولوژیکی کیپر و هال ۶-
- وزن دار شده با قطبش‌پذیری [۷۲]. $G3m$ از این گروه توصیف‌کننده‌هاست که دارای ضریب مثبت می باشد.

۴-۱-۱۰- بررسی میزان مشارکت توصیف‌کننده‌های منتخب در شبکه عصبی

بر اساس روشی که در بخش (۳-۱-۱۱) توضیح داده شد، میزان مشارکت توصیف‌کننده‌های منتخب در پیش‌بینی فعالیت دارویی موردنظر برآورد شد که نتایج آن در شکل (۴-۱۷) ارائه شده است.



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

همان‌طور که مشاهده می‌شود، توصیف‌کننده $H8v$ بیشترین اثر مشارکت را در مدل ارائه شده دارد. $H8v$ در برگزیده اطلاعات سه بعدی مولکول است که این اطلاعات به صورتبندی دارو ارتباط

¹ - Kier-hall

دارد. بنابراین می‌توان گفت که فعالیت بازدارنده این ترکیبات وابستگی قابل ملاحظه‌ای به تغییرات صورتبندی دارو دارد.

۴-۱-۱۱- نتیجه‌گیری نهایی

در این تحقیق برای اولین بار از روش‌های SR-ANN و GA-ANN برای مطالعه‌ی ارتباط کمی ساختار- فعالیت مشتقات پیریدوپیریدازین به عنوان بازدارنده‌های p38 MAP کیناز استفاده شد. ارزیابی مدل نشان‌دهنده صحت و اعتبار مدل‌های پیشنهادی است، در واقع این روش‌ها به طراحان دارو این امکان را می‌دهد که قبل از اینکه بخواهند یک ترکیب دارویی را سنتز کنند، ابتدا با استفاده از مدل معتبر فعالیت دارویی آن را پیش‌بینی کنند و در صورتی که دارای فعالیت دارویی مناسبی باشند، ترکیب سنتز شود و بدین ترتیب با صرف هزینه و وقت کمتر در مسیر تهیه داروهایی با اثر بخشی بهتر گام بردارند.

آینده‌نگری

- ❖ پژوهشگران می‌توانند نتایج حاصل از این مطالعات را در طراحی و سنتز داروهای جدید با فعالیت دارویی بهتر به کار ببرند.
- ❖ می‌توان از روش‌هایی همچون بهینه‌سازی اجتماع مورچگان^۱ و طرح‌ریزی متوالی^۲ برای انتخاب توصیف‌کننده‌های مهم و از روش‌هایی مانند SVM^۳ و LS-SVM^۴ به جای ANN استفاده کرد.
- ❖ بکار بردن روش‌های سه بعدی مانند CoMFA و CoMSIA جهت پیش‌بینی pIC₅₀ این ترکیبات بجای روش مرسوم دو بعدی QSAR.

¹ - Ant Colony Optimization

² - Successive Projection Algorithm

³ - Supported Vector Machine

⁴ - Least Square Supported Vector Machine

- [1]- <http://en.wikipedia.org/wiki/trypanosoma>.
- [2]- <http://www.who.int/mediacentre/factsheets>.
- [3]- Croft S.L., Barrett M.P., Urbina J.A., (2005), "*Chemotherapy of trypanosomiasis and leishmaniasis*", Trends in Parasitology, 21, pp 508–512.
- [4]- Hamilton P.B., Stevens J.R., Gaunt M.W., Gidley J., Gibson W.C., (2004), "*Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA*", Journal of Parasitology, 34, pp 1393–1404.
- [5]- Lee J.C., Kumar S., Griswold D.E., Underwood D.C., Votta B.J., Adams J.L., (2000), "*Inhibition of p38 MAP kinase as a therapeutic strategy*" Immunopharmacology, 47, pp 185–201.
- [6]- Goldstein D., Gabriel T., (2005), "*Pathway to the clinic: inhibition of P38 MAP kinase. A review of ten chemotypes selected for development*", Journal of Current Topics in Medicinal Chemistry, 5, pp 1017–1029.
- [7]- Jalali-Heravi M., Asadollahi-Baboli M., (2008), "*QSAR analysis of platelet-derived growth inhibitors using GA-ANN and shuffling crossvalidation*", Journal of QSAR Combustion Science and Technology, 6, pp 750-757.
- [8]- Cerecetto H., Aguirre G., Boiani L., Boiani M., (2005), "*New potent 5-substituted benzofuroxans as inhibitor of Trypanosoma cruzi growth: quantitative structure–activity relationship studies*", Journal of Bioorganic & Medicinal Chemistry, 13, pp 6336–6346.
- [9]- Schmidt T.J., Nour A.M., Khalid S.A., Kaiser M., (2009), "*Quantitative structure-antiprotozoal activity relationships of sesquiterpene lactones*", Molecule, 14, pp 2062-2076.
- [10]- Noolvi M.N., Patel H.M., (2010), "*3D QSAR studies on a series of 1,3,4-thiadiazole-2- arylhydrazones derivatives as antitrypanosomal agents the k-nearest neighbor molecular field analysis approach*", Journal of Nonomaterials and Biostructures, 5, pp 519-533.
- [11]- Barbosa E.G., Pasqualoto K.F., Ferreira E.I., (2012), "*Receptor Dependent (RD) LQTA-QSAR applied to a set of phenothiazine derivatives as inhibitors of t. cruzi trypanothione reductase*", Journal of Computer-Aided Molecular Design, 26, pp 1055-1065.
- [12]- Lozano B., Honorio M., (2012), "*Molecular features for antitrypanosomal activity of thiosemicarbazones revealed by OPS-PLS QSAR studies*", Journal of Medicinal Chemistry, 12, pp 1045-1056.

- [13]- Nayana M.R.S., Sekhar Y.N., (2007), "Comparative Molecular Field Analysis (CoMFA) for p38 Inhibitors", Journal of Molecular Design, 6, PP 385-395.
- [14]- Sarma R., Sinha S., Ravikumar M., Kumar K., Mahmood S., (2008), "3D-QSAR (Quantitative Structure-Activity Relationships) Studies on Urea Derivatives as Inhibitors of p38 MAP Kinase", Journal of Molecular Design, 7, PP 38-46.
- [15]- Achaiah G., Ravindra G.K., Sastry G.N., (2008), "Molecular modeling studies of phenoxypyrimidinyl imidazoles as p38 kinase inhibitors using QSAR and docking", Journal of Medicinal Chemistry, 43, PP 830-838.
- [16]- Kulkarni V.M., Shah U.A., Deokar H.S., Kadam S.S., (2010), "3D-QSAR and pharmacophore identification of novelimidazolyl benzimidazoles and imidazo[4,5-b]pyridines as potent p38 α mitogen activated protein kinase inhibitors", Journal of Chemical Technology Research, PP 194-204.
- [17]- Brereton R.G., (1990), "Chemometrics", Ellis Horwood Chichester, pp 308.
- [18]- منصور عرب چم جنگلی (۱۳۸۶)، "پیش‌بینی فعالیت دارویی ضد ایدز (سیتوتوکسیستی) مشتقات ۵-فنیل -۱- فنیل آمینو (H-ایمیدازول بوسیله شبکه عصبی مصنوعی"، دانشگاه صنعتی شاهرود، گزارش طرح پژوهشی.
- [19]- <http://www.qsar.pe.kr/descriptors>.
- [20]- Wold S., (1995), "Chemometrics; what do we mean with it and what do we want from it", Chemometrics and Intelligent Laboratory Systems, 30, pp 109-115.
- [21]- Jurs P.C., (2005), "Assessing the reliability of a QSAR models predictions", Journal of Molecular Graphics and Modelling, 23, pp 503-523.
- [22]- Guha R., (2005), "Methods to improve the reliability, validity and interpretability of QSAR models", PhD thesis, Pennsylvania state university.
- [23]- Arab Chamjangali M., (2009), "Modeling of cytotoxicity data (CC₅₀) of anti-HIV 1-[5-chlorophenyl) sulfonyl]-1H-pyrrole derivatives using calculated molecular descriptors and Levenberg- Marquardt artificial neural network", Chemical and Biology and Drug Design, 73, pp 456-465.
- [24]- [http:// www.hyper.com](http://www.hyper.com).
- [25]- فرشادفرع، (۱۳۸۰)، "اصول و روشهای پیشرفته آماری (تجزیه رگرسیون)"، چاپ دوم، انتشارات طاق بستان.
- [26]- مرکز انفورماتیک و توسعه، (۱۳۷۳)، "راهنمای نرم‌افزار SPSS/PC+"، چاپ اول، انتشارات راهبرد.
- [27]- نیروئی م، (۱۳۷۵)، پایان‌نامه ارشد، "طراحی یک مدل هیبرید و شبکه‌ی عصبی برای تشخیص سرطان سینه در ماموگرافی"، دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران.
- [28]- باوری ا، صالحی م، (۱۳۸۷)، "الگوریتم‌های ژنتیک و بهینه‌سازی سازه‌های مرکب"، انتشارات عابد، تهران.
- [29]- www.docstoc.com/docs/132602251/genetic.

[30]- رضائی ع، رنجبران س، (۱۳۸۶) "آموزش کاربردی الگوریتم ژنتیک در نرم افزار MATLAB"، انتشارات آذر، تهران.

[31]- www.en.wikipedia.org/wiki/Genetic_Algorithm.

[32]- مقاله الگوریتم ژنتیک/ www.prozhe.com.

[33]- فردوسی م، (۱۳۸۹)، پایان نامه ارشد: "پیش بینی ثابت های هنری بعضی از ترکیبات آلی با استفاده از روش های خطی و غیر خطی QSPR"، دانشگاه صنعتی شاهرود.

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

[35]- Nanakorn P., Meesomklin K., (2001), "An adaptive penalty function in Genetic Algorithms for structural design optimization", Computers and Structures, 79. pp 2527-2539.

[36]- Imam H., Al-Shihri M. A., (2000), "A primitive crossover for improving the reliability of Genetic Algorithms for structural optimization, in computational engineering using metaphors from nature", Civil-Comp Press, Edit. B.H.V. Topping, pp 91-97.

[37]- معرفی الگوریتم ژنتیک/ <http://chemometrics.blogfa.com/category/8/>

[38]- Leardi R., Boggia R., Terrile M., (1992), "Genetic Algorithms as a strategy for feature selection", Chemometrics, 6, pp 267-281.

[39]- Fisz J. J., (2006), "Combined Genetic Algorithm and Multiple Linear Regression (GA-MLR) optimizer: application to multiexponential fluorescence decay surface", Journal of Physical Chemistry, 48, pp 12977-12985.

[40]- Bazarraa M.S., Sherali H.D., Shetty C.M., (2006), "Nonlinear programming theory and algorithms", Third ed. John-Wiley., New York, pp 853.

[41]- Mcculloch W.S., Pitts W.H., (1943), "A logical calculus of the ideas immanent in neural nets", Bulletin of Mathematical Biophysics, 5, pp 115-133.

[42]- 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, 110, pp 13-22.

[43]- <http://www.emc.maricopa.edu/faculty/farabee/biobk/biobooknerv>.

[44]- بن کوپین، داورپناه ه، میرزایی ر، (۱۳۸۵)، "هوش مصنوعی"، انتشارات سخن گستر، مشهد.

[45]- اشرفی م، (۱۳۸۹)، پایان نامه ارشد: "مطالعه ارتباط کمی ساختار- فعالیت مشتقات تیوکربامات ها به عنوان دسته ی جدیدی از بازدارنده های غیر نوکلئوزیدی HIV"، دانشگاه صنعتی شاهرود.

[46]- 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", Journal of Chem. Inf. Model., 46, pp 1255-1268.

[47]- منهای م، سیفی پور ن، (۱۳۷۷) "کاربرد هوش محاسباتی در کنترل"، مرکز نشر دانشگاه صنعتی امیرکبیر، تهران.

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

[49]- Hornik K., Stinchcombe M., White H., (1989) "*Multilayer feed forward networks are universal approximators*", Neural Networks, 2, pp 359-366.

[50]- Rumelhart D.E., Hinton G.E., Williams R.J., (1986), "*Learning internal representations by error propagation*", Parallel Data Processing, Vol.1. Chapter 8, The M.I.T. Press, Cambridge, MA, pp 318-362.

[51]- <http://foram.takdownload.ir/threads/25029>.

[52]- منهای م، (۱۳۸۷) "مبانی شبکه‌های عصبی (هوش محاسباتی)" جلد اول، چاپ پنجم، مرکز نشر دانشگاه صنعتی امیرکبیر، تهران.

[53]- Hagan M.T., Demuth H.B., Beale M.H., (1992), "*Neural network design*", Chapter 11-12, ISBN 0-9717321-0-8 (available from John Stovall, john.stovall@colorado.edu, 303.492.3648).

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

[55]- Hagan M.T., Menhaj M.B., (1994), "*Training feed forward networks with the marquardt algorithm*", IEE Trans. Neural network., pp 989-993.

[56]- <http://www.disat.unimib.it>.

[57]- <http://www.spss.com>.

[58]- MATLAB (2010), the Math Work, Inc., Natick, MA, USA.

[59]- Fotie J., Kaiser M., Dawn A., (2010), "*Antitrypanosomal Activity of 1,2-Dihydroquinolin-6-ols and Their Ester Derivatives*", Journal of Medicinal Chemistry, 3, pp 966-982.

[60]- <http://en.wikipedia.org/wiki/PLS>.

[61]- Wold S., (1993), "*Discussion PLS in chemical practice*", Technometrics, 35, pp 136-139.

[62]- Haupt R. L., Haupt S. E., (1998), "*Practical Genetic Algorithm*", John Wiley and Sons, Inc., New York.

[63]- Ghasemi J., Niazi A., Leardi R., (2003), "*Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: application on copper and zinc mixture*", Talanta, 59, pp 311-317.

- [64]- Colt W., Smith D., Smith E., (1999), "Using a Neural Network as a function evaluator during GA search for reliability optimization", industrial engineering university of Pittsburg.
- [65]- Karelson M., Lobanov V.S., Katritzky A.R., (1996), "Quantum-chemical descriptors in QSAR/QSPR studies", Chemical Reviews, 96, pp 1027–1043.
- [66]- Todeschini R., Consonni V., (2000), "Handbook of molecular descriptors", Wiley-VCH, Weinheim, Germany.
- [67]- Julio Caballero, Miguel Garrigab and Michael Fernandez., (2006), "2D Autocorrelation modeling of the negative inotropic activity of calcium entry blockers using bayesian-regularized genetic neural networks", Bioorganic & Medicinal Chemistry, 14, pp 3330-3340.
- [68]- Douali L., Villemin D., Cherqaoui D., (2003), "Neural networks: Accurate nonlinear QSAR model for HEPT derivatives", Journal of Chemical Information and Computer Science, 43, pp 1200-1207.
- [69]- Robert M., Swaminathan R., Edward A., (2011), "Synthesis and biological activity of pyridopyridazin-6-one p38 MAP kinase inhibitors", Bioorganic & Medicinal Chemistry, 21, pp 411–416.
- [70]- Consonni V., Todeschini R., Pavan M., (2002), "Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors.1. Theory of the Novel 3D Molecular Descriptors", Journal of Chemical Information and Computer Science, 42, pp 682-692.
- [71]- Todeschini R., Consonni V., (2000), "Handbook of molecular descriptors", WILEY-VCH, pp 492-497.
- [72]- Todeschini R., Gromatica P., (1997), "3D- modeling and prediction by WHIM descriptors. Part5. Theory development and chemical meaning of WHIM descriptors", Quantitative Structure Activity Relationships, 16, pp 113-119.

Abstract

In the first section of this project, Quantitative Structure-Activity Relationship (QSAR) has been developed for modeling and prediction of anti-trypanosomal activities of some dihydroquinolin compounds. The stepwise regression (SR) and genetic algorithm (GA) were used as variable selection methods. The selected variable by these methods was used as input of artificial neural network (ANN) to construct of model for prediction of anti-trypanosome activities of these compounds. The obtained models were validated by different techniques such as using the external test set, leave one out (LOO) and Y-randomization. The obtained results show that both models of SR-ANN and GA-ANN have a good prediction ability. The determination coefficient of the test set for SR-ANN and GA-ANN were obtained 0.9042, 0.8402 respectively.

In the second part of this project, QSAR modeling for 42 derivatives of pyridopyridazine as P38 MAP kinase inhibitors was proposed. The descriptors for building of model based on ANN were selected by two methods of stepwise regression (SR) and genetic algorithm (GA). The determination coefficient of the test set for SR-ANN and GA-ANN were obtained 0.8607, 0.8027 respectively.

Keyword: Anti trypanosomal activity, P38 MAP inhibitors, QSAR, SR-ANN, GA-ANN.



Shahrood University of Technology

Faculty of Chemistry

**Prediction activity of some anti trypanosomal drug
compounds using linear and non linear QSAR
methods**

Hasan Rameh

Supervisor:

Dr. N. Goudarzi

Advisor:

Dr. M. Arab Chamjangali

Date: **Feb** – 2013