

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده: شیمی

مطالعه ارتباط کمی ساختار- فعالیت مشتقات تیوکربامات‌ها به عنوان
دسته‌ی جدیدی از بازدارنده‌های غیر نوکلئوزیدی HIV

دانشجو: مطهره اشرفی

استاد راهنما:

دکتر منصور عرب چم‌جنگلی

استاد مشاور:

دکتر قدمعلی باقریان دهقی

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

دی ماه ۱۳۸۹

الهی،

در عرصه پهناور گیتی،

هر چه بیش می دانم ، بدانم که هیچ ندانم

مرا مدد کن تا این دانش اندکم نه نردبانی باشد

برای فزونی غرور و تکبر،

نه حلقه‌ای برای اسارت،

نه دستمایه‌ای برای تجارت،

بلکه گامی باشد برای تجلیل از تو و متعالی ساختن

زندگی خود و دیگران.

تقدیم به:

پدر و مادر عزیزم

و همسر صبور و دختر نازنینم فاطمه

آنان که وجودم برایشان همه رنج بود و وجودشان

برایم همه مهر

تشکر و قدردانی

اکنون که به یاری خداوند متعال این دوره از تحصیلاتم را به پایان رسانده‌ام، بر خود واجب می‌دانم، تلاشهای خستگی ناپذیر پدر و مادر دلسوزم، برادران مهربانم و همسر فداکارم که همواره ره‌گشای مشکلاتم در تمامی مراحل زندگی بوده، ارج نهاده و مراتب قدردانی و تشکر قلبی خویش را از الطاف و مهربانی‌های آنها ابراز دارم. همچنین لازم می‌دانم از زحمات فراوان استاد فرهیخته و توانمندم جناب آقای دکتر منصور عرب که راهنمایی‌ها و نظرات ارزنده، صبر و حوصله فراوان ایشان، نقش مهمی در به ثمر رساندن این پروژه داشت، صمیمانه تشکر کنم. از استاد فرزانه و بزرگوارم جناب آقای دکتر قدمعلی باقریان که به عنوان مشاور در این پروژه نقش داشتند نیز نهایت قدردانی را دارم. و سلامتی و موفقیت همیشگی این بزرگواران را از درگاه یزدان پاک خواستارم.

چکیده

بازدارنده‌های غیرنوکلئوزیدی نسخه‌بردار معکوس، یک طبقه جدید از داروهای ضد رتروویروسی هستند که برای درمان عفونت ویروس نقص ایمنی اکتسابی (HIV) استفاده می‌شوند. این بازدارنده‌ها از فعالیت آنزیم نسخه‌بردار معکوس که تکثیر ژنوم HIV را در کنترل دارد، جلوگیری می‌کنند. در بخش اول پروژه، مدل‌های QSAR برای پیش‌بینی فعالیت ضد HIV یکسری از آنالوگ‌های فنتیل تیازولیل تیواوره (PETT) که به عنوان بازدارنده‌های غیرنوکلئوزیدی قوی عمل می‌کنند، با استفاده از توصیفگرهای تئوری ساخته شدند. برای انتخاب مهم‌ترین توصیفگرها رگرسیون خطی مرحله‌ای مورد استفاده قرار گرفت. سپس توصیفگرهای مهم به عنوان ورودی برای ایجاد مدل‌های QSAR با استفاده از رگرسیون خطی چندگانه و شبکه عصبی، استفاده شد. سری داده‌ها به طور تصادفی به ۳۹ ترکیب آموزش و ۱۰ ترکیب تست تقسیم شد و ارزیابی تقاطعی بر روی سری آموزش برای انتخاب بهترین مدل MLR و ANN به کار رفت. بررسی ارزیابی مدل‌های به دست آمده، با استفاده از سری تست و رد مرحله‌ای تک‌تک (LOO) انجام گرفت. نتایج نشان داد که مجذور ضریب همبستگی برای پیش‌بینی خاصیت ضد HIV ترکیبات سری تست با مدل MLR و ANN به ترتیب ۰/۷۶۶ و ۰/۹۱۳ است. در بخش دوم تحقیق، روش‌های MLR و ANN برای مدل‌سازی و پیش‌بینی صحیح فعالیت ضد HIV یکسری از مشتقات ۲-۰-۲ (۲-هیدروکسی کربونیل بنزآمیدو) اتیل-N آریل تیوکربامات، استفاده شد. داده‌ها به طور تصادفی به دو سری آموزش و تست به ترتیب شامل ۲۵ و ۷ ترکیب تقسیم شدند. ارزیابی تقاطعی بر روی سری آموزش برای انتخاب بهترین مدل MLR و ANN به کار رفت. توانایی پیشگویی مدل با استفاده از سری تست و روش رد مرحله‌ای تک‌تک مورد ارزیابی قرار گرفت. مجذور ضریب همبستگی به دست آمده برای سری تست توسط روش MLR و ANN به ترتیب ۰/۹۲۰ و ۰/۹۷۹ بودند. نتایج به دست آمده، توانایی پیشگویی مناسب مدل‌های پیشنهادی را نشان می‌دهد.

کلمات کلیدی: QSAR، ضد HIV، شبکه عصبی مصنوعی، رگرسیون خطی چندگانه

نتایج حاصل از این پایان نامه در دو پوستر تحت عناوین:

“Use of Bayesian regularized Artificial Neural Network in a QSAR study of PETT analogs as non- nucleoside HIV-1 reverse transcriptase inhibitors”

“Linear and nonlinear QSAR study of O-[2-(2- hydroxy carbonyl benzamido) ethyl] -N-aryl thiocarbamates as novel Anti-HIV chemicals”

در هفدهمین سمینار شیمی تجزیه ایران در دانشگاه کا شان پذیرفته شد.

فهرست مطالب

فصل اول

- ۱-۱-۱- ایدز.....
- ۱-۱-۲- ساختمان ویروس HIV
- ۱-۱-۳- چرخه تکثیر ویروس HIV
- ۱-۱-۴- مراحل بیماری
- ۱-۱-۵- مرحله مقدماتی (مرحله حاد)
- ۱-۱-۶- مرحله بدون علامت (مزمن)
- ۱-۱-۷- مرحله ایدز(بحرانی)
- ۱-۱-۸- اهداف دارو درمانی بر ضد ویروس HIV
- ۱-۱-۹- درمان ضد رتروویروسی HIV
- ۱-۱-۱۰- انواع داروهای ضد رتروویروسی
- ۱-۱-۱۱- بازدارنده‌های ورودی
- ۱-۱-۱۲- بازدارنده‌های نوکلئوزیدی آنزیم نسخه‌بردار معکوس
- ۱-۱-۱۳- مهارکننده‌های غیرنوکلئوزیدی نسخه‌بردار معکوس
- ۱-۱-۱۴- بازدارنده‌های نوع پروتئاز
- ۱-۲- ضرورت تحقیق
- ۱-۳- مروری بر کارهای انجام شده

فصل دوم

۱۶.....	۱-۲- کمومتریکس
۱۶.....	۲-۲- ارتباط کمی ساختار- فعالیت
۱۹.....	۱-۲-۲- فراهم کردن سری داده‌ها
۱۹.....	۲-۲-۲- بهینه‌سازی ساختار مولکول‌ها
۲۱.....	۳-۲-۲- محاسبه توصیفگرها
۲۲.....	۴-۲-۲- انتخاب توصیفگرهای مناسب
۲۲.....	۵-۲-۲- ساختن مدل
۲۳.....	۶-۲-۲- رگرسیون خطی چندگانه
۲۴.....	۱-۶-۲-۲- روش ورود اجباری
۲۴.....	۲-۶-۲-۲- انتخاب پسرو
۲۴.....	۳-۶-۲-۲- انتخاب پیشرو
۲۵.....	۴-۶-۲-۲- انتخاب مرحله‌ای
۲۵.....	۷-۲-۲- مقدمه‌ای بر شبکه عصبی
۲۶.....	۱-۷-۲-۲- تاریخچه
۲۷.....	۲-۷-۲-۲- ساختمان و عملکرد نرون طبیعی
۲۸.....	۳-۷-۲-۲- ساختمان و عملکرد نرون مصنوعی
۳۱.....	۴-۷-۲-۲- انواع شبکه‌های عصبی از نظر برگشت‌پذیری

۳۲	۵-۷-۲-۲ معماری شبکه‌های پیشخور
۳۳	۶-۷-۲-۲ مراحل آموزش در شبکه‌های پیشخور
۳۳	۱-۶-۷-۲-۲ ایجاد شبکه
۳۳	۲-۶-۷-۲-۲ مقدار دهی آغازین به وزن‌ها
۳۳	۳-۶-۷-۲-۲ آموزش شبکه
۳۴	۴-۶-۷-۲-۲ روش‌های آموزش
۳۴	۷-۷-۲-۲ آموزش شبکه‌های پیشخور با تکنیک پس انتشار
۳۹	۸-۷-۲-۲ بهبود تعمیم
۴۱	۸-۲-۲ ارزیابی مدل
۴۶	۳-۲ نرم‌افزارهای مورد استفاده
۴۶	۱-۳-۲ بسته نرم‌افزاری Hyperchem
۴۷	۲-۳-۲ بسته نرم‌افزاری Dragon
۴۷	۳-۳-۲ بسته نرم‌افزاری SPSS
۴۸	۴-۳-۲ نرم‌افزار MATLAB

فصل سوم

۵۰	۱-۳ مدل‌سازی فعالیت دارویی مشتقات تیواوره به عنوان بازدارنده‌های غیرنوکلئوزیدی ویروس HIV
۵۰	۱-۳-۱-۱ انتخاب سری داده‌ها
۵۶	۲-۱-۳ رسم و بهینه‌سازی ساختار مولکول‌ها
۵۶	۳-۱-۳ محاسبه توصیفگرها
۵۷	۴-۱-۳ انتخاب توصیفگرهای مناسب
۵۸	۵-۱-۳ بررسی عدم وجود همخطی بین کل توصیفگرهای انتخاب شده
۵۹	۶-۱-۳ مدل‌سازی توسط روش رگرسیون خطی چندگانه
۶۱	۷-۱-۳ مدل‌سازی توسط شبکه عصبی مصنوعی
۶۲	۱-۷-۱-۳ انتخاب تعداد لایه‌های پنهان
۶۳	۲-۷-۱-۳ انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع آموزش و نوع تابع انتقال
۶۸	۳-۷-۱-۳ انتخاب تعداد دوره‌های آموزش
۷۰	۴-۷-۱-۳ انتخاب مقدار بهینه μ
۷۱	۵-۷-۱-۳ ساختار شبکه عصبی مصنوعی بهینه شده
۷۲	۸-۱-۳ ارزیابی مدل‌های خطی و غیرخطی
۷۲	۱-۸-۱-۳ ارزیابی مدل‌های خطی و غیرخطی با استفاده از سری تست
۷۳	۲-۸-۱-۳ ارزیابی مدل‌های خطی و غیرخطی توسط روش رد مرحله‌ای تک‌تک و رد مرحله‌ای گروهی
۸۱	۹-۱-۳ ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری
۸۱	۱۰-۱-۳ ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون Y-تصادفی
۸۲	۱۱-۱-۳ بررسی ارتباط توصیفگرهای منتخب با فعالیت ضد HIV
۸۷	۱۲-۱-۳ بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی
۸۸	۱۳-۱-۳ نتیجه‌گیری نهایی

فصل چهارم

۹۰	۱-۴ مدل‌سازی فعالیت دارویی مشتقات تیوکربامات به عنوان بازدارنده‌های غیر نوکلئوزیدی HIV
----	---

۹۰.....	۴-۱-۱- سری داده‌ها
۹۳.....	۴-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها
۹۳.....	۴-۱-۳- محاسبه و انتخاب توصیفگرهای مهم
۹۴.....	۴-۱-۴- کشف پدیده همخطی بین کل توصیفگرهای انتخاب شده
۹۵.....	۴-۱-۵- مدل سازی و انتخاب بهترین مدل با روش رگرسیون خطی (MLR)
۹۷.....	۴-۱-۶- مدل‌سازی به روش شبکه عصبی مصنوعی
۹۷.....	۴-۱-۶-۱- انتخاب تعداد لایه‌های پنهان
۹۸.....	۴-۱-۶-۲- انتخاب تعداد ورودی‌های شبکه، نوع تابع آموزش، نوع تابع انتقال و تعداد نرون‌های لایه پنهان
۱۰۳.....	۴-۱-۶-۳- انتخاب تعداد دوره‌های آموزش
۱۰۴.....	۴-۱-۶-۴- بهینه کردن مقدار μ
۱۰۵.....	۴-۱-۶-۵- معماری شبکه عصبی بهینه شده
۱۰۶.....	۴-۱-۷- ارزیابی مدل‌های خطی و غیرخطی
۱۰۶.....	۴-۱-۷-۱- ارزیابی مدل خطی و غیرخطی با استفاده از سری تست
۱۰۷.....	۴-۱-۷-۲- ارزیابی مدل‌های خطی و غیرخطی به روش رد مرحله‌ای تک‌تک و رد مرحله‌ای گروهی
۱۱۲.....	۴-۱-۸- ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری
۱۱۳.....	۴-۱-۹- آزمون χ^2 - تصادفی
۱۱۳.....	۴-۱-۱۰- بررسی ارتباط توصیفگرهای وارد شده در مدل با فعالیت بازدارندگی
۱۱۷.....	۴-۱-۱۱- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی
۱۱۸.....	۴-۱-۱۲- نتیجه‌گیری نهایی
۱۱۹.....	آینده نگری
۱۲۰.....	منابع

فهرست اشکال

- شکل (۱-۱) - ساختمان ویروس HIV ۳
- شکل (۲-۱) - چرخه تکثیر ویروس HIV ۵
- شکل (۳-۱) - نحوه ورود ویروس HIV به درون سلول میزبان ۸
- شکل (۴-۱) - نحوه عملکرد بازدارنده‌های نوکلئوزیدی آنزیم نسخه بردار معکوس ۸
- شکل (۵-۱) - نحوه عملکرد بازدارنده‌های غیرنوکلئوزیدی آنزیم نسخه بردار معکوس ۹
- شکل (۶-۱) - نحوه عملکرد بازدارنده‌های آنزیم پروتئاز ۱۰
- شکل (۱-۲) مراحل QSAR ۱۸
- شکل (۲-۲) - ساختار یک نرون زیستی ۲۸
- شکل (۳-۲) - ساختمان یک نرون محاسباتی ۲۸
- شکل (۴-۲) - نرون با یک بردار ورودی ۲۹
- شکل (۵-۲) - انواع توابع انتقال ۳۰
- شکل (۶-۲) - انواع شبکه‌های عصبی: الف: شبکه عصبی پیشخور ب: شبکه عصبی برگشتی ۳۱
- شکل (۷-۲) - شبکه پیشخور tansig/pureline ۳۲
- شکل (۸-۲) - شیوه آموزش با ناظر ۳۴
- شکل (۹-۲) - نحوه‌ی عملکرد تکنیک پس‌انتشار در یک شبکه عصبی پیشخور ۳۶
- شکل (۱۰-۲) - تغییرات خطای سری آموزش و سری ارزیابی ۳۹
- شکل (۱-۳) - ساختار $N - (۲-فنتیل) - N' - (۲-تيازولیل)$ تیواوره ۵۰
- شکل (۲-۳) - اسکلت اصلی ترکیبات مورد مطالعه ۵۱
- شکل (۳-۳) - تأثیر تعداد توصیفگرها بر مقادیر R^2_{adj} مدل‌های MLR ۶۰
- شکل (۴-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی ۶۴
- شکل (۵-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی ۶۵
- شکل (۶-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی ۶۶
- شکل (۷-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی ۶۷

- شکل (۳-۸) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در دوره‌های آموزش مختلف ۶۹
- شکل (۳-۹) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر μ ۷۰
- شکل (۳-۱۰) - ساختار شبکه عصبی بهینه شده ۷۱
- شکل (۳-۱۱) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pIC_{50} برای تست الف: با مدل MLR ب: با مدل ANN ۷۳
- شکل (۳-۱۲) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN ۷۶
- شکل (۳-۱۳) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای گروهی برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN ۷۹
- شکل (۳-۱۴) - نمودار باقیمانده‌ها بر حسب مقادیر تجربی pIC_{50} برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN ۸۰
- شکل (۳-۱۵) - مشارکت توصیف‌گرها در شبکه عصبی بهینه ۸۷
- شکل (۴-۱) - اسکلت اصلی O-۲-۲-هیدروکسی کربونیل بنزآمیدو اتیل -N- آریل تیوکربامات‌ها ۹۰
- شکل (۴-۲) - تأثیر تعداد توصیف‌گرها بر مقادیر R^2_{adj} مدل‌های MLR ۹۶
- شکل (۴-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال تانژانت سیگموئیدی ۹۹
- شکل (۴-۴) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-ماکوارت و تابع انتقال لگاریتم سیگموئیدی ۱۰۰
- شکل (۴-۵) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی ۱۰۱
- شکل (۴-۶) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی ۱۰۲
- شکل (۴-۷) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در دوره‌های آموزش متفاوت ۱۰۳
- شکل (۴-۸) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر μ ۱۰۴
- شکل (۴-۹) - ساختار شبکه عصبی بهینه شده ۱۰۵
- شکل (۴-۱۰) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pEC_{50} برای تست الف: با مدل MLR ب: با مدل ANN ۱۰۷
- شکل (۴-۱۱) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pEC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN ۱۰۹
- شکل (۴-۱۲) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pEC_{50} به روش رد مرحله‌ای گروهی برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN ۱۱۱

شکل (۴-۱۳) - نمودار باقیمانده‌ها بر حسب مقدار تجربی pEC_{50} برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN
۱۱۲.....

شکل (۴-۱۴) - مشارکت توصیفگرها در شبکه عصبی بهینه ۱۱۸

فهرست جداول

جدول (۳-۱) - ساختار و مقادیر pIC_{50} ترکیبات مورد مطالعه	۵۲
جدول (۳-۲) - فهرست توصیفگرهای محاسبه شده توسط نرم افزار Dragon	۵۶
جدول (۳-۳) - کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله	۵۷
جدول (۳-۴) - ماتریس همبستگی کل توصیفگرها انتخاب شده توسط رگرسیون خطی چندگانه مرحله ای	۵۸
جدول (۳-۵) - مقادیر عامل افزایش واریانس کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه	۵۸
جدول (۳-۶) - پارامترهای آماری برای مدل‌های به دست آمده از روش MLR	۵۹
جدول (۳-۷) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت	۶۴
جدول (۳-۸) - مقادیر میانگین مربعات حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره	۶۵
جدول (۳-۹) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره	۶۶
جدول (۳-۱۰) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره	۶۷
جدول (۳-۱۱) - توصیفگرهای انتخاب شده توسط شبکه عصبی	۶۸
جدول (۳-۱۲) - مقادیر میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در دوره‌های آموزش متفاوت	۶۹
جدول (۳-۱۳) - میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ	۷۰
جدول (۳-۱۴) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی	۷۱
جدول (۳-۱۵) - نتایج حاصل از ارزیابی مدل‌های MLR و ANN با استفاده از سری تست	۷۲
جدول (۳-۱۶) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از رد مرحله‌ای تک تک برای کل داده‌ها	۷۴
جدول (۳-۱۷) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از روش رد مرحله‌ای گروهی برای کل داده‌ها	۷۷
جدول (۳-۱۸) - پارامترهای آماری برای مدل‌های برتر طراحی شده توسط MLR و ANN	۸۱

- جدول (۳-۱۹) - مقادیر R^2 سری تست با استفاده از آزمون Y - تصادفی..... ۸۲
- جدول (۳-۲۰) - مثال‌هایی از اثر توصیفگر nPhX بر فعالیت دارویی ۸۲
- جدول (۴-۱) - ساختار و مقادیر تجربی pEC_{50} مشتقات O-۲- (۲- هیدروکسی کربونیل بنزآمیدو) اتیل - N - آریل تیوکربامات ۹۱
- جدول (۴-۲) - کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله به مرحله و معانی آنها ۹۴
- جدول (۴-۳) - ماتریس همبستگی کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه ۹۴
- جدول (۴-۴) - مقادیر عامل افزایش واریانس کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه ۹۵
- جدول (۴-۵) - پارامترهای آماری برای مدل‌های به دست آمده از روش MLR ۹۶
- جدول (۴-۶) - توصیفگرهای مدل منتخب و ضرائب آنها ۹۶
- جدول (۴-۷) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره ۹۹
- جدول (۴-۸) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره ۹۹
- جدول (۴-۹) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد متفاوت توصیفگر ورودی و تعداد مختلف گره ۱۰۰
- جدول (۴-۱۰) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگر ورودی و تعداد مختلف گره ۱۰۲
- جدول (۴-۱۱) - مقادیر میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در دوره‌های آموزش مختلف ۱۰۳
- جدول (۴-۱۲) - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ ۱۰۴
- جدول (۴-۱۳) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی ۱۰۵
- جدول (۴-۱۴) - نتایج حاصل از ارزیابی مدل‌های برتر MLR و ANN با استفاده از سری تست ۱۰۶
- جدول (۴-۱۵) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با روش رد مرحله‌ای تک‌تک ۱۰۸
- جدول (۴-۱۶) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با روش رد مرحله‌ای گروهی ۱۱۰
- جدول (۴-۱۷) - پارامترهای آماری برای مدل‌های برتر طراحی شده توسط MLR و ANN ۱۱۳
- جدول (۴-۱۸) - مقادیر R^2 سری تست با استفاده از آزمون Y - تصادفی ۱۱۳
- جدول (۴-۱۹) - اثر توصیفگر MSD بر فعالیت دارویی ۱۱۴

فصل اول

مقدمه

۱-۱- ایدز

سندرم نقص ایمنی اکتسابی (AIDS)^۱ یک بیماری عفونی کشنده است که اولین بار در سال ۱۹۸۱ شناسایی شد. دو سال بعد یعنی در سال ۱۹۸۳ دو دانشمند به نامهای رابرت گالو^۲ و مونتگنیر^۳ ویروس نقص ایمنی انسانی (HIV)^۴ را به عنوان عامل این بیماری معرفی کردند [۱ و ۲]. عفونت با این ویروس منجر به تخریب پیشرونده سیستم ایمنی بدن به خصوص لنفوسیت‌ها می‌گردد و هنگامی که سیستم ایمنی آسیب ببیند نه تنها در برابر ویروس HIV (که در آغاز به آن صدمه زده) بلکه نسبت به سایر عفونت‌های فرصت طلب^۵ که در حالت عادی بدن به راحتی با آنها مقابله می‌کند، آسیب پذیر می‌شود. لذا با گذشت زمان افراد آلوده به HIV بیشتر و بیشتر بیمار می‌شوند و معمولاً سالها پس از آلودگی با این ویروس، به یکی از بیماری‌های خاص (بیماری‌های متأثر از نقص دستگاه ایمنی) مبتلا شده و در این زمان گفته می‌شود که آنها به ایدز مبتلا شده‌اند. بنابراین زمانی که فرد آلوده به ویروس HIV برای اولین بار به یک بیماری جدی مبتلا شود و یا وقتی که تعداد سلولهای ایمنی باقیمانده در بدن او از حد معینی کمتر شود (زیر ۲۰۰ عدد در هر میکرولیتر)، مبتلا به بیماری ایدز در نظر گرفته می‌شود [۳ و ۴].

HIV به خانواده رتروویروس‌ها تعلق دارد. رتروویروس‌ها، گروهی از ویروس‌ها هستند که ژنوم آنها از RNA تشکیل شده است. بنابراین برای تکثیر خود به آنزیمی به نام آنزیم نسخه‌بردار معکوس^۶ وابسته هستند که ژنوم RNA آنها را به DNA نسخه‌برداری کند تا بعد بتواند آن را به کمک آنزیم اینتگراز^۷ وارد ژنوم میزبان کند و به این ترتیب امکان تکثیر ویروس به وجود آید. کلمه‌ی رترو

^۱ - Acquired Immune Deficiency Syndrome

^۲ - Robert Gallo

^۳ - Montagnier

^۴ - Human Immunodeficiency Virus

^۵ - Infection opportunistic

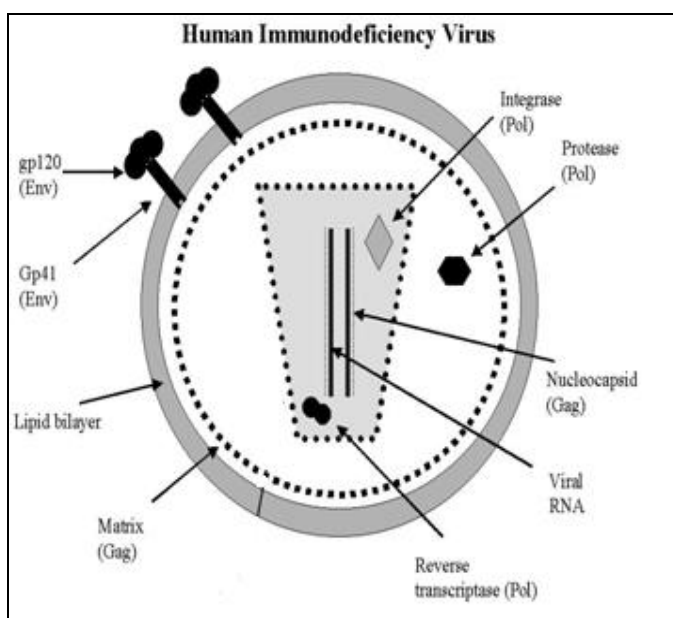
^۶ - Reverse transcriptase

^۷ - Integrase

(معکوس) در نام این ویروس‌ها به همین خاطر است چرا که معمولاً نسخه‌برداری از DNA به RNA انجام می‌شود اما در این دسته از ویروس‌ها عکس این عمل صورت می‌گیرد [۵ و ۶].

۱-۱-۱- ساختمان ویروس HIV

HIV یک ذره کاملاً کروی با قطر ۱۲۰-۱۰۰ نانومتر است که با غشای دو لایه لیپیدی پوشیده شده است. در سطح غشای لیپیدی دو نوع گلیکوپروتئین (gp120, gp41) وجود دارد که ویروس از طریق این گلیکوپروتئین‌ها قادر خواهد بود به گیرنده‌های CD_4^+ موجود در سطح لنفوسیت T کمک کننده^۱ بچسبد. سطح داخلی غشای لیپیدی توسط ماتریکس احاطه شده و اطلاعات ژنتیکی مورد نیاز برای تکثیر ویروسی را در خود جای داده است. این اطلاعات شامل دو کپی یکسان از ژنوم RNA ویروسی و سه آنزیم اینتگراز، پروتئاز و نسخه‌بردار معکوس می‌باشد [۷]. شکل (۱-۱) ساختمان ویروس HIV را نشان می‌دهد.



شکل (۱-۱)- ساختمان ویروس HIV

^۱ - T helper lymphocyte

۱-۲- چرخه تکثیر ویروس HIV

چرخه تولید و تکثیر ویروس شامل مراحل زیر است:

۱- چسبیدن ویروس به سلول‌های خونی و آلوده کردن سلول‌ها

۲- تولید DNA پروویروسی و اتصال به ژنوم میزبان

۳- رونویسی ژن‌های ویروسی

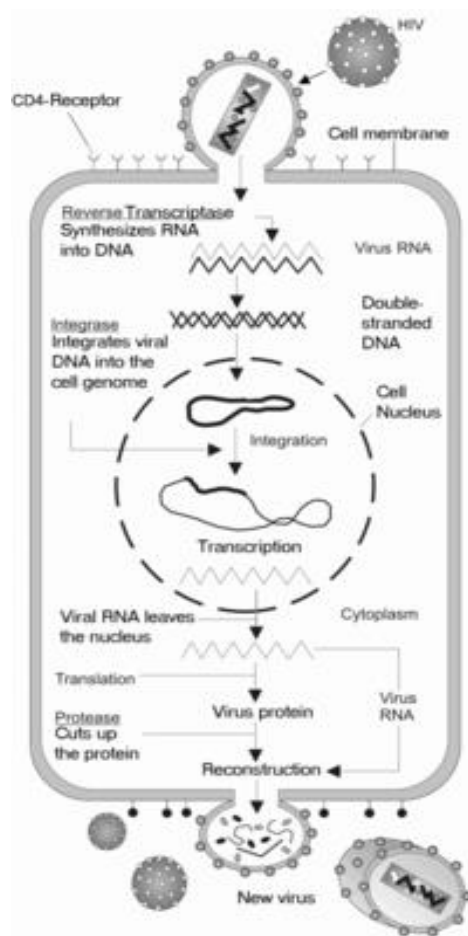
۴- تولید قطعات ویروسی و رهاسازی ویروس

سلول‌هایی که مورد تهاجم این ویروس قرار می‌گیرند، نوعی از سلول‌های خونی انسان به نام لنفوسیت‌های نوع T کمک کننده می‌باشند. این سلول‌ها دارای گیرنده‌ای به نام $CD4^+$ هستند که ویروس پس از اتصال به این گیرنده، توسط پروتئین GP41 (موجود در سطح ویروس HIV) به داخل غشای سلول نفوذ می‌کند و RNA ویروس و آنزیم‌های مختلف (از جمله نسخه‌بردار معکوس، اینتگراز و پروتئاز) به داخل سیتوپلاسم سلول تزریق می‌شود. مرحله بعد تولید DNA از روی RNA ویروس با کمک آنزیم نسخه بردار معکوس است. در صورت موفقیت این عمل، DNA ساخته شده که DNA پروویروسی نام دارد، با استفاده از آنزیم اینتگراز به DNA سلول میزبان ملحق می‌گردد. بعد از رونویسی^۱ در هسته سلول، RNA ناقل^۲ به سیتوپلاسم سلول میزبان منتقل می‌شود و از اجزای سلولی میزبان برای ساختن پروتئینهای ویروسی استفاده می‌کند. آنزیم پروتئاز فراهم شده به وسیله HIV وظیفه شکستن پلی پپتید ساخته شده به رشته‌های پروتئینی سازندهی HIV را بر عهده دارد. بدین ترتیب ویروس HIV می‌تواند یک ذره^۳ کامل بسازد که به غشای میزبان مهاجرت کرده و غشای لیپیدی مورد نیازش را از غشای سلول میزبان می‌گیرد و آماده‌ی آلوده کردن سلول‌های دیگر است [۷-۱۰]. شکل (۱-۲) این مراحل را نشان می‌دهد.

^۱ - Transcription

^۲ Messenger RNA

^۳ - Particle



شکل (۱-۲) - چرخه تکثیر ویروس HIV

۱-۳-۱-۱ - مراحل بیماری

۱-۳-۱-۱-۱ - مرحله مقدماتی (مرحله حاد)

علائم بالینی مرحله حاد عبارتند از: خستگی، احساس کسالت، سردرد، کاهش شدید وزن، تب و اسهال. در این مرحله ویروس در پلاسما مشخص بوده و تا ۱۰٪ لنفوسیت‌های $CD4^+$ خون آلوده هستند که سیستم ایمنی با تولید آنتی‌بادی شروع به پاسخ در برابر ویروس می‌کند. از هنگام ورود تا

مثبت شدن نتیجه آزمایشگاهی تست آنتی‌بادی HIV^۱ که نشانگر آلودگی فرد است حدود ۲ تا ۱۲ هفته و گاهی تا ۶ ماه طول می‌کشد.

۱-۳-۲- مرحله بدون علامت (مزمن)

بعد از مرحله حاد دوره بدون علامت وجود دارد که ۷ تا ۱۱ سال طول می‌کشد. این دوره که در گذشته به دوره نهفتگی معروف بود، در حقیقت عفونت خاموشی را نشان می‌دهد که طی آن تخریب مداوم و دائمی سیستم ایمنی انجام می‌گیرد. در این دوره تعداد لنفوسیت‌های TCD_4^+ آلوده افزایش می‌یابد و اثرات تخریبی سیستم ایمنی مشخص‌تر می‌گردد. پاسخ‌های سیستم ایمنی سلولی به طور فزاینده‌ای کاهش می‌یابد و فرد آلوده قادر نیست که نسبت به آنتی‌ژن‌های میکروبی پاسخ دهد. بایستی خاطر نشان کرد که ویروس در این مرحله و در طی مرحله سوم خصوصیات را که در شروع بیماری داشت، دارا نمی‌باشد. در این دوره ویروس HIV سریع‌تر تکثیر می‌یابد و محدوده سلولی آن فراتر است.

۱-۳-۳- مرحله ایدز (بحرانی)

پیشگویی آزمایشگاهی در مورد پیشرفت بیماری به سوی ایدز به وسیله شمارش لنفوسیت‌های TCD_4^+ انجام می‌گیرد. هنگامی که تعداد لنفوسیت‌ها خیلی کاهش یابد و به میزان کمتر از ۲۰۰ سلول در هر میکرولیتر برسد، فرد در طبقه مبتلایان به ایدز قرار می‌گیرد [۳].

۱-۴- اهداف دارو درمانی بر ضد ویروس HIV

- ❖ سرکوب ویروس‌ها و جلوگیری از تکثیر و ازدیاد آن‌ها
- ❖ تحکیم و بهبود عملکرد سیستم ایمنی که با شمارش سلول‌های TCD_4^+ ارزیابی می‌شود.
- ❖ بهبود در کیفیت زندگی بیماران با کاهش عوارض بیماری

^۱ - HIV Antibody Test

❖ بهبود امید به زندگی و افزایش طول عمر بیماران

❖ کاهش احتمال سرایت

۱-۵- درمان ضد رتروویروسی HIV

درمان ضد رتروویروسی نوع اصلی درمان ایدز می‌باشد که البته شفا بخش نیست اما زندگی مبتلایان به این بیماری را طولانی‌تر می‌کند. درمان شامل داروهایی است که بایستی روزانه و تا آخر عمر مصرف شوند [۵].

۱-۶- انواع داروهای ضد رتروویروسی

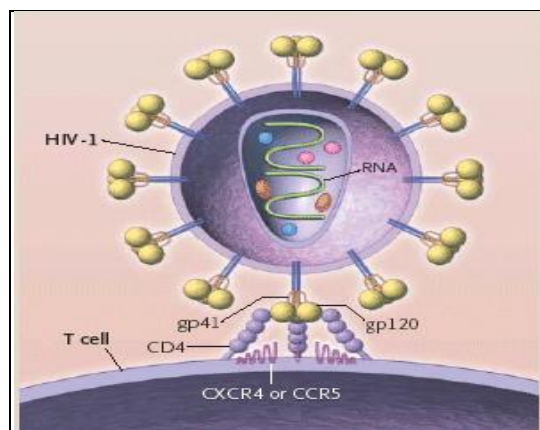
از زمان آغاز اپیدمی HIV مجموعه‌ای از داروها ساخته شده‌اند که از طریق مهار توانایی تکثیر این ویروس، عمر افراد مبتلا به این ویروس را به طور قابل ملاحظه‌ای افزایش می‌دهند. این داروها از سرعت نابودی $CD4^+$ بیماران می‌کاهد و می‌توانند بروز بیماری ایدز را به تأخیر بیندازد، اما آن را معالجه نمی‌کند. در مجموع چهار دسته دارو وجود دارد که در مقاطع مختلف آلودگی و رشد ویروس HIV عمل می‌کنند و با توجه به اینکه این ویروس از سه آنزیم نسخه‌بردار معکوس، پروتئاز و اینتگراز برای تکثیر خود و آلوده کردن سلول‌ها استفاده می‌کند، بنابراین داروهای سنتز شده بیشتر این نوع آنزیم‌ها را هدف قرار می‌دهند.

۱-۶-۱- بازدارنده‌های ورودی^۱

در سطح بیرونی ویروس HIV گلیکوپروتئین‌های gp120, gp41 وجود دارد (شکل (۱-۳)) که به وسیله آن‌ها ویروس به گیرنده‌های $CD4^+$ موجود در سطح لنفوسیت‌های T سلول میزبان متصل شده و وارد سلول می‌گردد. بازدارنده‌های ورودی به این پروتئین‌ها می‌چسبند و از پیوستن ویروس به

^۱ - Fusion or Entry inhibitors

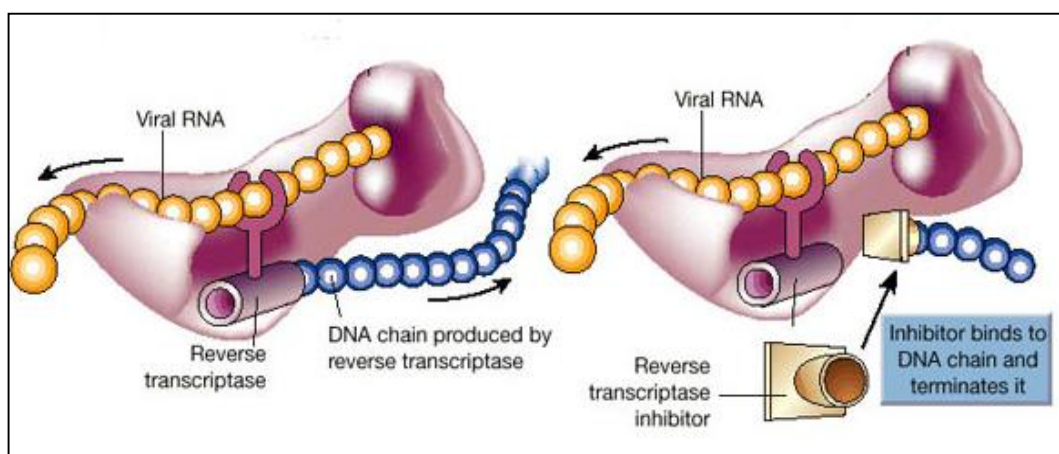
سلولهای $CD4^+$ جلوگیری می‌کند. بدین ترتیب با بلوکه کردن یکی از این پروتئین‌ها، تکثیر آهسته می‌شود. تاکنون یک نمونه از آن به نام فوزئون^۱ وارد بازار شده است [۱۱]. (فوزئون به پروتئین gp41 متصل می‌گردد)



شکل (۳-۱) - نحوه ورود ویروس HIV به دورن سلول میزبان

۱-۶-۲- بازدارنده‌های نوکلئوزیدی آنزیم نسخه‌بردار معکوس^۲

مهارکننده‌های نوکلئوزیدی به زنجیره نوکلئوزیدی DNA که توسط آنزیم نسخه‌بردار معکوس در حال تولید است، می‌چسبند و مانع از اتصال نوکلئوزیدهای بیشتر به DNA می‌شوند به این ترتیب یک DNA ناقص تولید می‌شود [۱۲]. شکل (۴-۱) نیز گویای این مطلب است.



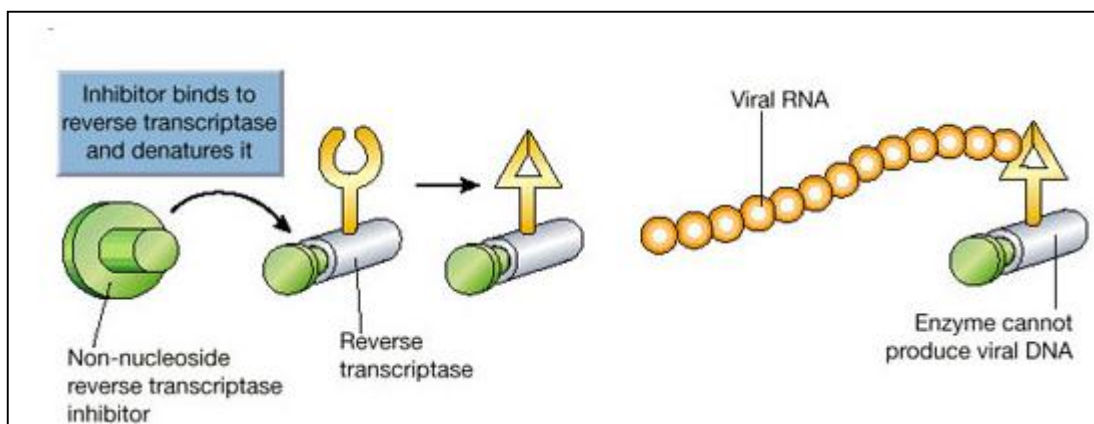
شکل (۴-۱) - نحوه عملکرد بازدارنده‌های نوکلئوزیدی آنزیم نسخه‌بردار معکوس

^۱ - Fuzeon

^۲ - Nucleoside Reverse transcriptase Inhibitors

۱-۱-۶-۳- مهارکننده‌های غیرنوکلئوزیدی نسخه‌بردار معکوس^۱

Nevirapin نخستین گروه از این مهارکننده‌ها بود که در سال ۱۹۹۶ توسط سازمان غذا و دارو مورد تایید قرار گرفت. بر اساس شکل (۱-۵) این نوع مهارکننده‌ها نیز فرایند رونویسی را مختل می‌کنند. آن‌ها با چسباندن خود به آنزیم نسخه‌بردار معکوس و محدود کردن تحرک آنزیم، مانع از نسخه‌برداری می‌شوند [۱۲]. مهمترین ویژگی این نوع مهارکننده‌ها اختصاصی بودن آن‌هاست.



شکل (۱-۵)- نحوه عملکرد بازدارنده‌های غیرنوکلئوزیدی آنزیم نسخه بردار معکوس

۱-۱-۶-۴- بازدارنده‌های نوع پروتئاز^۲

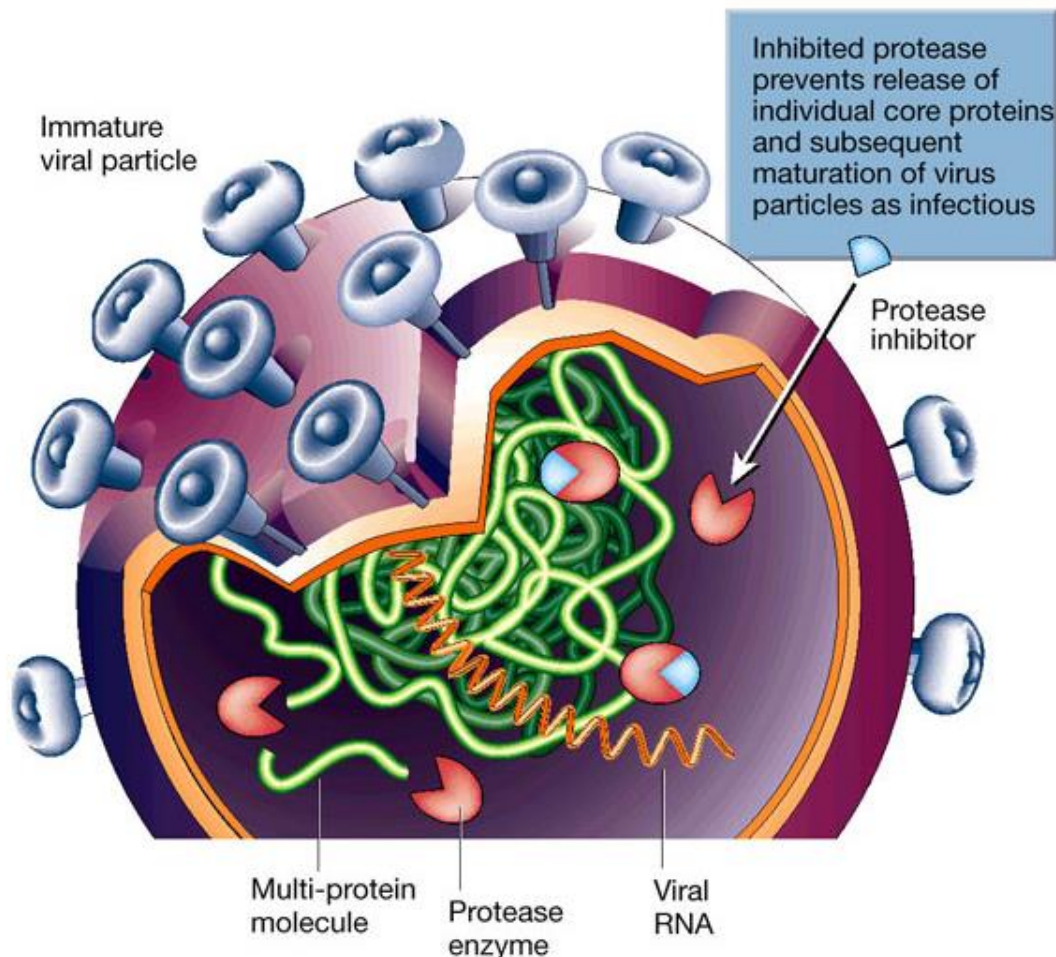
این داروها اولین بار در سال ۱۹۹۵ تایید شدند. پروتئاز یک آنزیم گوارشی است که پروتئین را تجزیه می‌کند و تقریباً در هر سلول زنده‌ای وجود دارد. همانطور که در چرخه زندگی ویروس اشاره شد، آنزیم پروتئاز در HIV به زنجیره طویل آنزیم‌ها و پروتئین‌ها، در سلول حمله کرده و آن‌ها را به قطعات کوچکتری تقسیم می‌کند تا از این طریق پروتئین‌های لازم برای تولید یک ویروس بالغ تولید شود. ولی طبق شکل (۱-۶) مهارکننده پروتئاز به مکان‌هایی از آنزیم که بریدن^۳ زنجیره پلی‌پپتیدی از آن مکان رخ می‌دهد، می‌چسبد و مانع از تولید پروتئین‌های لازم برای تولید ویروس جدید می‌شود. به

^۱ - Non Nucleoside Reverse transcriptase Inhibitors

^۲ Protease inhibitors(PI)

^۳ - Cut

عبارتی، مهارکننده پروتئاز قادر است فرایند تبدیل ویروس نابالغ غیرعفونی را به ویروس بالغ عفونی، آهسته کند [۱۳].



شکل (۱-۶) - نحوه عملکرد بازدارنده‌های آنزیم پروتئاز

۲-۱- ضرورت تحقیق

ایدز یکی از دلایل اصلی مرگ و میر در جهان است به طوری که در سال ۲۰۰۹ بیش از چهل میلیون نفر به این بیماری مبتلا شده‌اند [۱۴]. این آمار همچنان در حال گسترش است و اکثر مبتلایان به ایدز در کشورهای در حال توسعه زندگی می‌کنند. در نتیجه نیاز به درمان توسط مواد دارویی ارزان، مؤثر و سالم امری ضروری است و اگر چه جهش‌های^۱ مکرر ویروس HIV سبب افزایش مقاومت

^۱ - Mutations

دارویی^۱ و کاهش کارایی درمان توسط داروهای شیمیایی شده است ولی نتایج نشان می‌دهد، اگر دو یا چند داروی ضدویروسی همزمان مصرف شوند، مقاومت دارویی به میزان قابل توجهی کاهش می‌یابد. بنابراین درمان ترکیبی^۲ سبب گسترش تحقیقات جدید در جهت کشف داروهای گزینش‌پذیرتر و قویتر شده است و از آن جایی که بیشتر داروهای شیمیایی توسط آزمون و خطا سنتز می‌شوند که بسیار وقت‌گیر و هزینه‌بر است [۱۵ و ۲۴]؛ توسعه روشهای نظری^۳ به عنوان یک ابزار کارآمد برای رفع این محدودیت‌ها مورد توجه قرار گرفته است. به کارگیری این روش‌ها می‌تواند علاوه بر پیش‌گویی فعالیت بیولوژیکی مورد نظر، به هدفمندتر شدن انجام آزمایش‌های تجربی و همین‌طور توضیح پارامترهای مؤثر بر نتایج این آزمایش‌ها و مکانیسم‌های درگیر کمک کند.

از جمله روشهای نظری، رابطه کمی ساختار- فعالیت (QSAR)^۴ است که در آن می‌توان از روشهای مختلفی برای ایجاد روابط خطی و غیرخطی میان ساختار و خاصیت دارویی ترکیباتی که فعالیت دارویی آنها با استفاده از روشهای تجربی اندازه‌گیری شده‌اند، استفاده نمود. مبنای روش QSAR، به دست آوردن روابط کمی بهینه بین ساختار و فعالیت دارویی ترکیبات مورد مطالعه است و هنگامی که رابطه معتبری به دست آمد امکان پیش‌بینی فعالیت دارویی برای ساختارهای مشابه اندازه‌گیری شده یا ساختارهایی که فعالیت آنها هنوز اندازه‌گیری نشده است، وجود دارد. بر این اساس QSAR به راحتی می‌تواند برای کمک به بهینه‌سازی سنتز داروها مورد استفاده قرار گیرد.

۱-۳- مروری بر کارهای انجام شده

جلالی هروی و پرستار در سال ۲۰۰۰ در چارچوب مطالعات QSAR برای ایجاد مدل‌هایی که ارتباط دهنده‌ی فعالیت ضد HIV با ساختار ۱۰۷ مشتق (۱-۲- (هیدروکسی اتوکسی) متیل) -۶- (فنیل تیو) تیمین (HEPT) باشد، مطالعاتی انجام دادند. در این کار مدل‌سازی توسط شبکه عصبی

^۱ - Drug- Resistance

^۲ - Combination Therapy

^۳ - Theoretical approaches

^۴ - Quantitative Structure- Activity Relationship

مصنوعی (ANN)^۱ نسبت به روش رگرسیون خطی چندگانه (MLR)^۲ نتایج بهتری را ارائه داد به طوری که میانگین مربع خطا (MSE)^۳ برای سری تست با استفاده از روش خطی چندگانه ۰/۷۸۰ و با استفاده از شبکه عصبی ۰/۳۷۲ به دست آمد [۱۶].

بازوئی و همکارانش در سال ۲۰۰۲ با استفاده از مطالعات QSAR، فعالیت ضد ایدز را برای مشتقات HEPT بررسی نموده‌اند. در این کار هر دو مدل شبکه عصبی و مدل رگرسیون خطی به کار رفت. نتایج به دست آمده بیانگر این واقعیت است که فعالیت ترکیبات HEPT به طور مؤثر وابسته به خصوصیات آبگریزی^۴ و گروه‌های استخلافی می‌باشد [۱۷].

دوئلی^۵ و همکارانش در سال ۲۰۰۴ ارتباط کمی ساختار- فعالیت ۸۰ ترکیب از مشتقات HEPT که بازدارنده آنزیم نسخه بردار معکوس در ویروس HIV می‌باشند، را با استفاده از شبکه عصبی مورد بررسی قرار دادند و نتایج این مدل را با روش‌های رگرسیون خطی چندگانه (MLR) و حداقل مربعات جزئی (PLS)^۶ که توسط لوکو^۷ و فرتی^۸ ارائه شده بود، مقایسه کردند و از مقادیر مجذور ضریب همبستگی، قابلیت پیشگویی بالاتر مدل ارائه شده توسط ANN را نتیجه گرفتند [۱۸].

مطالعه ارتباط خطی بین ۴۲ مشتق ۵- فنیل-۱- فنیل آمینو-۱-هیدروژن-ایمیدازول و خاصیت دارویی آنها در سال ۲۰۰۵ توسط روی^۹ و لئوناردو^{۱۰} انجام گرفت [۱۹]. همچنین در سال ۲۰۰۷ عرب چم جنگلی و همکارانش ارتباط میان ساختار و فعالیت دارویی همین ترکیبات را با استفاده از شبکه مصنوعی به دست آوردند و نشان دادند که مدل ارائه شده توسط شبکه عصبی نسبت به مدل خطی ارائه شده توسط لئوناردو قدرت پیشگویی بهتری برای این دسته از ترکیبات دارد [۲۰].

¹ - Artificial Neural Network
² - Multiple Linear Regression
³ - Mean Square Error
⁴ - Hydrophobic
⁵ - Douali
⁶ - Partial Least Square
⁷ - Luco
⁸ - Ferreti
⁹ - Roy
¹⁰ - Leonardo

راکایک^۱ و همکارانش در سال ۲۰۰۷ ارتباط کمی ساختار- فعالیت ۹۵ ترکیب که به عنوان بازدارنده آنزیم نسخه‌بردار معکوس عمل می‌کنند، با استفاده از شبکه عصبی مصنوعی بررسی کردند و قابلیت پیش‌بینی مدل ساخته شده که شامل ۵ توصیفگر بود، توسط رد مرحله‌ای تک‌تک^۲ مورد ارزیابی قرار دادند. همچنین میزان مشارکت^۳ توصیفگرهای وارد شده در مدل را نیز مورد تجزیه و تحلیل قرار دادند و نشان دادند که توصیفگرهایی که به خاصیت آگریزی ترکیبات مربوط می‌شوند، دارای بیشترین اثر مشارکت هستند [۲۱].

در سال ۲۰۰۹ زارعی و عتباتی فعالیت ضد اچ- آی- وی ۲۳ ترکیب از مشتقات HEPT را بر اساس رگرسیون خطی چندگانه (MLR) بررسی کردند. ضریب همبستگی مدل خطی شامل ۱۳ توصیفگر ۰/۹۹۹۹ به دست آمد [۲۲].

به منظور دستیابی به ساختارهای مورد نیاز جهت بازدارنده آنزیم اینتگراز، مطالعه QSAR بر روی ۳۷ مشتق فتالیمید توسط دسالیو^۴ انجام گرفت. نتایج این بررسی نشان داد که فعالیت ضد رتروویروسی به اندازه، شکل و قطبیت استخلافها بستگی دارد و این بصیرت را ایجاد کرد که با اصلاح اندازه استخلاف و میزان قطبیت آن می‌توان برهمکنش آنزیم- بازدارنده را بهبود بخشید [۲۳].

عرب چم جنگلی در سال ۲۰۰۹ با استفاده از روش غیرخطی بر مبنای شبکه عصبی، مدلی مناسب برای فعالیت ضد HIV مشتقات پیرول آریل سولفون‌ها ارائه داد به طوری که ضریب همبستگی برای سری تست ۰/۹۸۹ بود [۲۴]. در همان سال مطالعه دیگر بر اساس شبکه عصبی توسط آل‌مریکو^۵ برای مطالعه ضد ویروسی تعدادی از آنالوگ‌های اسایکلوویر^۶ انجام شد [۲۵].

جلال هروی مدل‌سازی QSAR ۱- (۳ و ۳ دی فنیل پروپیل)- پایپیریدینیل آمیدها را که به عنوان اصلاح‌کننده CCR₅ در هنگام حمله ویروس HIV عمل می‌کنند، با استفاده از شبکه عصبی ژنتیک

¹ - Rakik

² - Leave One Out Cross Validation

³ - Contribution

⁴ - Dessalew

⁵ - Almerico

⁶ - Acyclovir

تنظیم بایزین (BRGNN) انجام داد. در این تحقیق هروی از الگوریتم ژنتیک برای انتخاب متغیرهای ورودی شبکه استفاده کرد و برتری این روش را نسبت به روش MLR اعلام کرد [۲۶].

در سال ۲۰۱۰ پاتل^۱ و همکارانش مطالعه QSAR را بر روی ۴۶ مشتق دی‌هیدرو-آلکوکسی-بنزیل-اکسوپیریمیدین (DABOs) که به عنوان بازدارنده‌های گیرنوکلتوزیدی ویروس HIV عمل می‌کنند، انجام دادند. مدل نهایی به دست آمده شامل ۴ توصیفگر بود و با توجه به علامت ضرائب این توصیفگرها، نتیجه گرفتند که استخلاف‌های آبرز و استخلاف‌های با شاخه بزرگتر، تأثیر مثبتی بر بازداری آنزیم نسخه‌بردار معکوس دارند [۲۷].

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

¹ - Patel

فصل دوم

کمو متریکس

۲-۱- کمومتریکس^۱

اصطلاح کمومتریکس برای اولین بار توسط اسوانت ولد^۲ دانشمند جوان سوئدی که در زمینه شیمی فیزیک آلی فعالیت داشت، مطرح گردید. همکاری ولد با بروس آر. کووالسکی^۳ که در دانشگاه واشنگتن بر روی الگوشناسی در شیمی تجزیه مطالعه می‌کرد، منجر به تأسیس انجمن بین‌المللی کمومتریکس (ICS)^۴ در سال ۱۹۷۴ گردید [۲۸]. بنا به تعریف ICS، کمومتریکس عبارت است از کاربرد روش‌های ریاضی، آماری برای برقراری ارتباط بین سنجش‌های انجام شده روی یک سیستم یا فرایند شیمیایی به منظور درک بهتر اطلاعات شیمیایی است [۲۹].

در حال حاضر با توجه به رشد سریع تجهیزات مورد استفاده در شیمی و حجم بسیار زیاد داده‌ها و اطلاعات به منظور جمع‌آوری، پردازش، تفسیر اطلاعات و استخراج نتایج مفید از آنها، نیاز به کامپیوتر امری اجتناب‌ناپذیر است. از طرف دیگر شیمیدانان، گاهی اوقات با موادی سر و کار دارند که بسیار گران سمی و خطرناک بوده و یا در مواردی به راحتی قابل دسترس نیستند. در این موارد، می‌توان از روش‌های ریاضی و آمار به طور مؤثری جهت توصیف و توجیه نتایج آزمایش‌های مختلف استفاده نمود [۳۰] و می‌توان گفت که موارد فوق مهم‌ترین دلایل افزایش کاربرد روش‌های کمومتریکس نزد شیمیدانان است.

۲-۲- ارتباط کمی ساختار - فعالیت

یکی از زمینه‌های مهم کمومتریکس، مطالعاتی است که فعالیت بیولوژیکی ترکیبات را به ویژگی‌های ساختاری آنها ارتباط می‌دهد. این نوع از مطالعات، به رابطه کمی ساختار- فعالیت (QSAR) معروف است [۱۵]. مولکول‌های مشابه با تغییر کوچکی در ساختارشان می‌توانند فعالیت

^۱ - Chemometrics

^۲ - Svante wold

^۳ - Bruce R. Kowaski

^۴ - International Chemometrics Society

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 [۳۱].

بررسی بیولوژیکی فراورده‌های طبیعی و برخی از مشتقات آن‌ها که خواص دارویی دارند، انگیزه‌ای جهت پیدا کردن وابستگی‌های احتمالی ساختار شیمیایی با فعالیت بیولوژیکی ایجاد نمود و مشخص شد که برخی از واحدهای ساختار مولکول‌هایی که از نظر بیولوژیکی فعال هستند، در مولکول‌های دیگری که دارای همان ویژگی بیولوژیکی می‌باشند، وجود دارد. این امر رهنمودی برای طراحی ساختار ترکیباتی با فعالیت بیشتر و در عین حال عملکردی اختصاصی‌تر بود.

در حال حاضر نیز تغییر و اصلاح ساختار یک ترکیب الگو، عمده‌ترین روش و روند در دستیابی به داروهای جدید می‌باشد. تغییرات ساختاری منجر به پیدایش ویژگی‌های جدید می‌شود که در نهایت موجب تغییر پخش دارو در سلول بافت، تغییر در کیفیت اثر متقابل با جایگاه‌های فعال آنزیم و گیرنده‌ها و در نهایت تغییر سرعت واکنش‌ها در جایگاه‌های مورد نظر می‌گردد [۳۲].

برای اولین بار در سال ۱۸۶۸ کرام براون^۱ و فریزر^۲ با این گمان که ویژگی آمونیوم چهارتایی ممکن است، مسئول فلج عضلانی باشد، اثرات انسداد عصبی - عضلانی انواع نمک‌های آمونیوم ۴ تایی را در حیوانات آزمایش کردند. این پژوهشگران از این بررسی نتیجه گرفتند که عمل فیزیولوژیکی یک مولکول تابعی از ساختار شیمیایی آن است [۳۳ و ۳۴]. مدت کوتاهی پس از آن ریچاردسون^۳ متذکر شد که فعالیت خواب آوری الکل‌های آلیفاتیک تابعی از وزن مولکولی آن‌ها است. این مشاهدات، اصولی برای درک ارتباط ساختار شیمیایی و فعالیت بیولوژیکی بود. اولین کوشش برای ارتباط دادن پارامترهای فیزیکوشیمیایی به اثر دارویی توسط ریکت^۴ گزارش شد. او مشاهده کرد که عمل مخدری

^۱ - Crum-brown

^۲ - Frazer

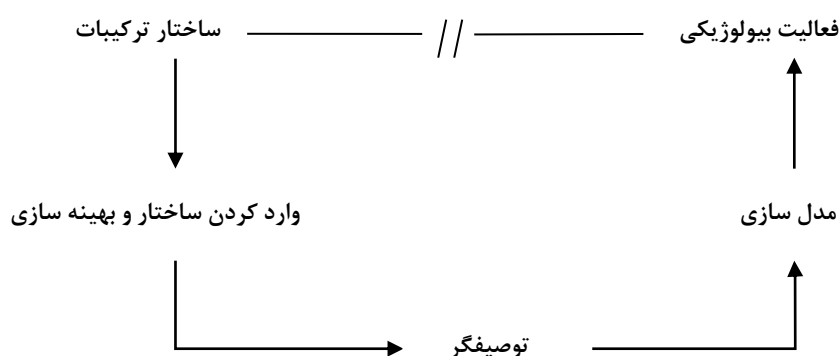
^۳ - Richardson

^۴ - Richet

گروهی از ترکیبات به طور معکوس با حلالیت آن‌ها متناسب است. در سال ۱۹۶۴ هانش^۱ و همکارانش بین فعالیت رشد گیاهی مشتقات فنوکسی استیک اسید و چگالی الکترونی در موقعیت اورتو ارتباطی مثبت پیدا کردند و برای اولین بار روش رگرسیون خطی چندگانه را برای به دست آوردن ارتباط کمی ساختار-فعالیت بکار بردند [۳۲]. او و دانشمندان پس از وی، از توصیفگرهای ساده‌ای برای توصیف ساختار ترکیبات استفاده کردند که توانایی توصیف ویژگی‌های الکترونی، فضایی و آب‌گریزی ترکیبات دارویی را داشتند [۳۵].

طبق شکل (۱-۲)، در مطالعات QSAR هرگز نمی‌توان به طور مستقیم ارتباط ساختار مولکول با فعالیت بیولوژیکی آن را بیان کرد، بلکه باید مراحل زیر به ترتیب دنبال شود [۳۶ و ۳۷]:

- ❖ فراهم کردن سری داده‌ها
- ❖ رسم ساختار و بهینه‌سازی
- ❖ محاسبه توصیف‌کننده‌ها
- ❖ انتخاب بهترین توصیف‌کننده‌ها
- ❖ ساختن مدل
- ❖ ارزیابی قدرت پیشگویی مدل



شکل (۱-۲) مراحل QSAR

^۱ - Hansch

۲-۲-۱- فراهم کردن سری داده‌ها

در مطالعات QSAR یافتن مجموعه‌ای از مولکول‌ها که مقادیر تجربی فعالیت بیولوژیکی مورد بررسی برای آنها در شرایط یکسانی اندازه‌گیری شده، در اولویت قرار دارد و زمانی مدل می‌تواند فعالیت مورد نظر را به درستی پیش‌بینی کند که مولکول‌های مورد استفاده تقریباً مشابه یا گروهی از مشتقات یک ترکیب باشند که این کار از طریق جستجو در مقالات صورت می‌گیرد و هر قدر سری داده‌ها بزرگتر و متنوع‌تر باشد، مدل حاصل از آن معتبرتر خواهد بود.

۲-۲-۲- بهینه‌سازی ساختار مولکول‌ها

ایجاد توصیفگرهای هندسی و هیبریدی بر مبنای ساختار و هندسه دقیق مولکولی استوار است و اگر ساختارها به صورت صورتبندی^۱ با حداقل انرژی نباشد، مقادیر غیر صحیحی برای این توصیفگرها ایجاد می‌شود [۲۴]. بنابراین باید ساختار هندسی ترکیب بهینه^۲ شود و مختصاتی از ساختار ترکیب به دست آید که در آن مختصات کارتیزین، انرژی مولکول حداقل باشد. این کار توسط شیمی محاسباتی^۳ انجام می‌شود. روش‌های مختلفی که در شیمی محاسباتی برای بهینه‌سازی ساختار مولکول مورد استفاده قرار می‌گیرند شامل [۳۸]:

۱- روش مکانیک مولکولی

۲- روش مکانیک کوانتومی

روش‌های مکانیک مولکولی بر مبنای روابط کلاسیک بنا شده است و به طور خیلی ساده ساختمان مولکولی به صورت گوی و فنر در نظر گرفته می‌شود. در این روش انرژی مولکول به صورت مجموعه‌ای از انرژی‌های کششی، خمشی، الکتروستاتیکی و ... بیان می‌شود. سپس با استفاده از برنامه‌های کامپیوتری مربوط به روش مکانیک مولکولی طول پیوندها، زوایای پیوندی و صورتبندی

^۱ - Conformation

^۲ - Optimize

^۳ - Computational Chemistry

تغییر می‌کند تا ساختاری که عبارت انرژی مکانیک مولکولی را مینیمم کند، پیدا شود. این روش بهینه‌سازی بیشتر برای ماکرومولکول‌ها به کار می‌رود [۳۹] ولی در مکانیک کوانتومی از معادله شرودینگر^۱ برای محاسبه انرژی مولکول استفاده می‌شود و نسبت به روش مکانیک مولکولی دارای صحت بیشتری است زیرا اثرات الکترونی اعمال شده روی مولکول را نیز در محاسبات وارد می‌کند در حالی که مکانیک مولکولی الکترون‌ها را در نظر نمی‌گرفت. از طرف دیگر این محاسبات که پیچیده می‌باشند، وقت بیشتری را طلب می‌کند. از جمله روش‌های کوانتومی می‌توان به روش‌های آغازین^۲ و روش‌های نیمه تجربی^۳ اشاره کرد [۳۸].

در روش‌های آغازین از هیچ داده تجربی استفاده نشده است بلکه ثابت‌هایی مانند جرم الکترون، ثابت پلانک، ثابت سرعت نور و بار الکترون به کار گرفته می‌شوند و محاسبه انرژی مولکول از حل معادله شرودینگر با استفاده از یکسری تقریب به دست می‌آید. در این روش چون کل الکترون‌ها را در محاسبات وارد می‌کند، بیشتر منحصر به مولکول‌های کوچک است و برای مولکول‌های بزرگتر محاسبات پیچیده‌تر و نیازمند ابر رایانه است [۳۸ و ۴۰].

روش‌های نیمه تجربی روش‌هایی هستند که در برنامه‌هایی نظیر MOPAC و Hyperchem وارد شده‌اند. در این روش‌ها برخلاف روش‌های آغازین، فقط الکترون‌های لایه ظرفیت را در محاسبات وارد می‌کنند. در نتیجه زمان محاسبات در این روش‌ها کوتاهتر از روش‌های آغازین است و می‌تواند برای مولکول‌های بزرگتر به کار رود. روش‌های نیمه تجربی مختلفی مانند چشم‌پوشی کامل از انتگرال‌های همپوشانی دیفرانسیلی (CNDO)^۴، چشم‌پوشی از برخی از انتگرال‌های همپوشانی دیفرانسیلی (INDO)^۵، اصلاح شده صرف نظر کردن دو اتمی از انتگرال همپوشانی (MNDO)^۶ و AM1^۷

^۱ - Schrodinger Equation

^۲ - Ab Initio

^۳ - Semi empirical

^۴ - Complete Neglect of Differential Overlap

^۵ - Intermediate Neglect of Differential Overlap

^۶ - Modified neglected of diatomic overlap

^۷ - Austian Method 1

وجود دارد [۴۱]. روش AM1 جزء دقیق‌ترین روش‌های نیمه تجربی است که در این پروژه از آن استفاده شده است [۳۸].

۲-۲-۳- محاسبه توصیفگرها

توصیفگرها مقادیر عددی هستند که ویژگی‌های مختلف ساختاری و الکترونی مولکول‌ها را به طور کمی نشان می‌دهند. به عبارت دیگر هر توصیفگر بیانگر خصوصیت ویژه‌ای از مولکول است که ممکن است بر خاصیت مورد نظر مؤثر باشد. برخی از ویژگی‌های یک توصیفگر مناسب عبارت است از [۴۲]:

❖ ساده بودن

❖ توانایی تفسیر ساختار مولکول

❖ عدم همبستگی با سایر توصیفگرها

❖ قابلیت تمایز بین ایزومرهای مختلف مولکول

❖ قابل کاربرد برای دامنه وسیعی از ساختارهای مولکولی

توصیفگرهای مولکولی را به روش‌های مختلفی طبقه‌بندی می‌کنند. ساده‌ترین طبقه‌بندی بر اساس ماهیت (تجربی یا تئوری) است. توصیفگرهای تجربی بر اساس اندازه‌گیرهای تجربی به دست می‌آیند و برای تعداد محدودی از ترکیبات شیمیایی قابل کاربرد هستند در حالی که توصیفگرهای تئوری با استفاده از ساختار مولکول و بدون نیاز به داده‌های تجربی محاسبه می‌شوند که باعث صرفه‌جویی در وقت، هزینه، مواد و تجهیزات خواهند شد و برای هر مولکول واقعی و یا فرضی که هنوز سنتز نشده‌اند، در دسترس می‌باشند [۲۴ و ۴۳]. به همین جهت در این تحقیق از توصیفگرهای تئوری استفاده شده است. تقسیم بندی دیگری برای توصیفگرها، بر اساس ابعاد توصیفگر (صفر بعدی، یک بعدی، دوبعدی و سه بعدی) می‌باشد.

۲-۲-۴- انتخاب توصیفگرهای مناسب

مرحله بعدی در مطالعات QSAR گزینش بهترین توصیفگرهای محاسبه شده، می باشد که بتوانند به خوبی فعالیت بیولوژیکی را به ساختار مولکول مرتبط کنند. یکی از روش های انتخاب توصیفگرهای مناسب، روش برازش مرحله ای^۱ است که از طریق محاسبه ضریب همبستگی^۲ بین توصیفگر و فعالیت بیولوژیکی انجام می شود. اگر توصیفگری دارای ضریب همبستگی برابر یک باشد، بدین معنی است که می تواند خواص مورد نظر را به درستی توصیف کند و ضریب همبستگی صفر بدین معنی است که هیچ ارتباطی بین توصیفگر و خاصیت دارویی وجود ندارد. پس توصیفگرهایی با ضرائب همبستگی بزرگ در برازش منحنی، برای ساخت مدل استفاده می شود. از دیگر روش های انتخاب توصیفگرها می توان به الگوریتم ژنتیک^۳ و روش جایگزینی^۴ اشاره کرد.

۲-۲-۵- ساختن مدل

مدل، در واقع یک رابطه ریاضی است که بیان کننده ی رابطه بین متغیر مستقل و متغیر وابسته است و به کمک آن می توان با داشتن مقادیر متغیر مستقل، متغیر وابسته را تخمین زد. مرحله اصلی در مطالعات QSAR ساختن مدل می باشد. مدل های QSAR را می توان بر اساس پارامترهایی که به صورت تجربی بدست آمده و یا از طریق تئوری محاسبه شده اند، به زبان ریاضی بیان نمود و اگر نتایج این نوع مطالعات همبستگی قابل قبولی را ارائه کند، علاوه بر شفاف سازی نحوه ی ارتباط بین خواص مولکول ها و ویژگی های ساختمانی آنها، به پژوهشگران در پیش بینی رفتار مولکول های جدید بر اساس رفتار مولکول های مشابه کمک می کند [۴۴]. چندین روش متفاوت برای ساختن مدل QSAR وجود دارد که می توان به رگرسیون خطی چندگانه (MLR)، آنالیز اجزای اصلی (PCA)^۵ و حداقل مربعات

¹ - Stepwise

² - Correlation coefficient

³ - Genetic Algorithm

⁴ - Replacement Method

⁵ - Principle Component Analysis

جزئی (PLS) به عنوان روش‌های خطی و شبکه عصبی مصنوعی (ANN) به عنوان یک روش غیرخطی اشاره کرد.

۲-۲-۶- رگرسیون خطی چندگانه

رگرسیون به مفهوم ایجاد یک رابطه بین یک متغیر از یک طرف و یک یا چند متغیر از طرف دیگر می باشد. اگر تنها یک متغیر مستقل داشته باشیم، رگرسیون را ساده و در غیر این صورت، رگرسیون را چندگانه می گویند [۴۵]. در بسیاری موارد نمی توان تغییرات یک متغیر را فقط به مقادیر یک متغیر دیگر مربوط کرد. به عبارت دیگر برای پیش بینی مقادیر یک متغیر، دانستن مقادیر دو یا چند متغیر دیگر لازم است در این حالت از روش‌های رگرسیون چندگانه استفاده می شود [۴۶].

ساده ترین و متداول ترین مدلی که برای ارتباط چند متغیر به کار می رود، مدل خطی چندگانه است. این مدل با رابطه خطی زیر تعریف می شود:

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

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

غالباً روش حداقل مربعات برای ساخت معادله خطی به کار می رود و مدلی که با این روش بدست می آید، دارای بهترین معادله برای مرتبط کردن متغیرهای مستقل و وابسته است، هر چند ممکن است همه‌ی مشاهدات تجربی بر روی این معادله واقع نشوند. ویژگی چنین خطی این است که مجموع مربعات انحراف عمودی نقاط مشاهده شده (y_i) از این خط، همواره کمتر از مجموع مربعات انحراف عمودی نقاط مشاهده شده (y) از هر خط مفروض دیگر است

در روش MLR چندین روش رگرسیون مختلف برای ساختن مدل وجود دارد که عبارتند از:

۲-۶-۱-۲-۲-۱ روش ورود اجباری^۱

در این روش همه‌ی متغیرهای مستقل (توصیفگرها) در مرحله مدل‌سازی به کار گرفته می‌شوند و چون اثرات مثبت و منفی متغیرهای وارد شده در مدل و میزان همبستگی آن‌ها با خاصیت مورد مطالعه (متغیر وابسته) نادیده گرفته می‌شود، برای ساخت مدل استفاده از این روش چندان مناسب نمی‌باشد.

۲-۶-۲-۲-۲ انتخاب پسرو^۲

در این روش ابتدا کلیه متغیرهای مستقل به طور همزمان در رگرسیون (ساخت مدل) به کار گرفته می‌شوند. در مرحله بعد، اثر حذف تک تک متغیرها از مدل مورد بررسی قرار می‌گیرد. به این ترتیب که اگر با حذف یک متغیر مستقل آماره F و R به مقدار قابل توجهی تغییر نکرد، آن متغیر از مدل حذف می‌شود و در غیر اینصورت به عنوان یک پارامتر مؤثر در ساخت مدل به کار می‌رود. این کار برای همه متغیرها تکرار می‌شود تا در نهایت متغیرهای مستقلی که بیشترین ارتباط با متغیر وابسته داشته باشند، به دست آیند.

۲-۶-۳-۲-۲ انتخاب پیشرو^۳

این روش عکس روش پسرو است. در اینجا ابتدا قویترین متغیرهای مستقل شناسایی می‌شوند یعنی متغیری که ضریب همبستگی آن با متغیر وابسته از بقیه متغیرها بیشتر باشد. سپس با توجه به ضریب همبستگی، متغیر بعدی وارد مدل می‌شود و در هر مرحله اثر ورود هر متغیر مورد بررسی قرار می‌گیرد یعنی اگر با ورود متغیر جدید تغییر در آماره F و R از نظر آماری معنادار باشد، متغیر مورد نظر در مدل نگه داشته می‌شود و در غیر این صورت آن متغیر از مدل حذف می‌شود.

^۱ - Enter

^۲ - Back ward selection

^۳ - Forward selection

۲-۶-۴- انتخاب مرحله‌ای

این روش ترکیبی از روش‌های پیشرو و پسرو است. بدین صورت که نخستین متغیر مستقل به روش انتخاب پیشرو وارد مدل می‌گردد (یعنی متغیری که بیشترین همبستگی با متغیر وابسته داشته باشد) در قدم دوم کلیه مدل‌های ممکن با دو متغیر انتخاب می‌شود، به طوری که یکی از متغیرها همان X_1 و متغیر دیگر از سایر متغیرهای خارج از مدل باشد. مقدار R برای هریک از مدل‌های دو متغیره محاسبه می‌شود و مدلی که دارای بزرگترین ضریب همبستگی است، انتخاب می‌شود. در این هنگام جهت بررسی معنادار بودن ضریب رگرسیون دو متغیر انتخاب شده، آزمون F انجام می‌شود و متغیری که ضریب رگرسیون آن از نظر آماری بی‌معنا باشد، از مدل حذف می‌شود. در صورت معنادار بودن هر دو ضریب، آن مدل دو متغیره انتخاب می‌شود. در قدم بعدی کلیه مدل‌های ممکن با سه متغیر انتخاب می‌شوند و کلیه مراحل قدم دوم تکرار می‌شود. در این روش با توجه فرایندی که رخ می‌دهد، هیچ تضمینی وجود ندارد که متغیر وارد شده در مدل تا مدل نهایی باقی بماند [۴۵].

۲-۷- مقدمه‌ای بر شبکه عصبی

شبکه عصبی یک برنامه نرم افزاری است که می‌تواند همانند مغز انسان عمل نماید. در واقع یک شبکه عصبی مصنوعی ایده‌ای است برای پردازش اطلاعات که از سیستم عصبی زیستی الهام گرفته شده و مانند مغز انسان به پردازش اطلاعات می‌پردازد.

این سیستم‌ها از تعداد زیادی عنصر پردازش به نام نرون^۱ تشکیل شده‌اند که برای حل یک مسئله به صورت هماهنگ با هم عمل می‌کنند. شبکه‌های عصبی مصنوعی نظیر مغز انسان‌ها، با مثال یاد می‌گیرند و با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه

^۱ - Neuron

منتقل می‌کنند. به همین خاطر به این سیستم‌ها، هوشمند^۱ گفته می‌شود زیرا شبکه‌ها بر اساس محاسبات روی داده‌های عددی یا مثال‌ها، قوانین کلی را فرا می‌گیرند.

۲-۷-۱- تاریخچه

پیشرفت عمده شبکه‌های عصبی از سال ۱۹۴۳ با فعالیت‌های وارن مک کلوث^۲ و والتر پیتز^۳ آغاز شد. آن‌ها نشان دادند که شبکه عصبی می‌تواند هر تابع حسابی یا منطقی را محاسبه کند [۴۷]. به دنبال آن دونالد هب^۴ مکانیسمی برای آموزش شبکه ارائه داد. اولین کاربرد عملی شبکه‌های مصنوعی در سال ۱۹۵۸ بود که در آن از شبکه‌ای که روزنبلات^۵ پیشنهاد داده بود، استفاده شد [۴۸]. روزنبلات یک شبکه چند لایه را ارائه کرد و توانایی آن را در انجام شناسایی الگو اثبات نمود. این موفقیت زود هنگام باعث پیدایش زمینه وسیعی در تحقیقات شبکه عصبی شد. تقریباً در همان زمان برنارد ویدرو^۶ الگوریتم آموزشی جدیدی ارائه داد که از لحاظ ساختار به شبکه روزنبلات شبیه بود. قاعده آموزش ویدروهاف امروزه نیز در شبکه‌های عصبی استفاده می‌شود. اما شبکه ویدروهاف و روزنبلات دارای محدودیت‌هایی بود که اولین بار توسط ماروین مینسکی^۷ مطرح شد. محققان تحت تاثیر نظریه مینسکی بر این عقیده بودند که تحقیق در مورد شبکه‌های عصبی پایان یافته است و از آن جایی که کامپیوترهای دیجیتال قوی در دسترس نبود، نزدیک به یک دهه تحقیق در مورد شبکه‌های عصبی معلق ماند. اما در سال ۱۹۸۶ با ظهور کامپیوترهایی با توان محاسباتی سریع، مجدد تحقیق در مورد شبکه‌های عصبی آغاز شد. در این سال الگوریتم آموزشی پس انتشار توسط جیمز مک کلاند^۸ ارائه شد [۴۹ و ۵۰] که نقطه عطفی در تحقیقات شبکه‌های عصبی به شمار می‌آید. از سال ۱۹۸۶ تاکنون

^۱ - Intelligent

^۲ - Warren McCulloch

^۳ - Walter Pitts

^۴ - Donald Hebb

^۵ - Rosenblatt

^۶ - Bernard Widrow

^۷ - Marvin Minsky

^۸ - James McClelland

محققان زیادی در مورد شبکه‌های عصبی تحقیق نموده‌اند که نتیجه آن صدها مقاله در این زمینه است.

امروزه از شبکه‌های عصبی به عنوان یک ابزار کارآمد در زمینه‌های مختلف علمی از جمله صنایع الکترونیک، پزشکی، اکتشاف نفت و گاز، رباتیک، شیمی و داروسازی استفاده می‌شود [۵۰].

۲-۷-۲-۲- ساختمان و عملکرد نرون طبیعی

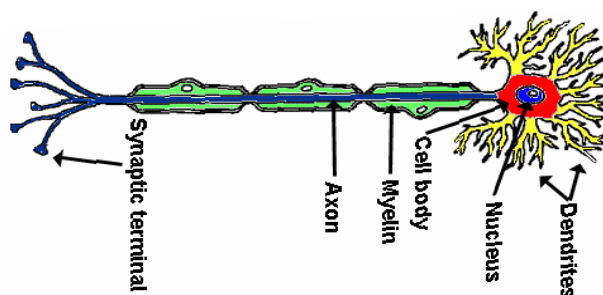
با بررسی‌های انجام شده روی مغز انسان مشخص شده است که مغز انسان از میلیاردها نرون عصبی منحصر به فرد تشکیل شده است. نرون اصلی‌ترین عنصر پردازش^۱ اطلاعات بشمار می‌آید که متشکل از دندريت^۲، بدنه‌ی سلول^۳ (شامل هسته و قسمت‌های حفاظتی دیگر) و اکسون^۴ می‌باشد. دندريت‌ها به عنوان مناطق دریافت سیگنال‌ها، شبکه‌هایی متشکل از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بی‌شمار می‌باشند. به همین علت آن‌ها را شبکه‌های دریافتی "درخت گونه" می‌گویند. دندريت‌ها پس از دریافت اطلاعات به شکل سیگنال آن را به هسته‌ی سلول که در بدنه‌ی سلول قرار دارد، هدایت می‌کنند. بدنه‌ی سلول ضمن فراهم کردن انرژی لازم برای فعالیت نرون، بر روی سیگنال‌های دریافتی عملیات لازم را انجام داده که این عملیات شبیه یک عمل جمع ساده بر روی کل سیگنال‌ها و مقایسه با یک سطح آستانه می‌باشد. اکسون که برخلاف دندريت‌ها از سطحی هموارتر و تعداد شاخه‌های کمتر ولی طول بیشتر برخوردار می‌باشد، سیگنال‌های الکتروشیمیایی دریافتی از بدنه‌ی سلول را توسط سیناپس‌ها که ارتباط دهنده نرون‌ها هستند، به نرون‌های دیگر منتقل می‌کند و بدین ترتیب فعالیت‌های مغزی انجام می‌شود [۵۱ و ۵]. شکل (۲-۲) ساختار یک نرون زیستی را نشان می‌دهد.

^۱ - Processing Element

^۲ - Dendrite

^۳ - Cell body(soma)

^۴ - Axon

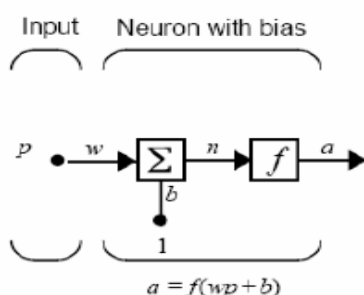


شکل (۲-۲) - ساختار یک نرون زیستی

۲-۲-۷-۳ - ساختمان و عملکرد نرون مصنوعی

با بحث مختصری که در مورد نرون‌های طبیعی انجام شد، برای ایجاد یک شبکه عصبی مصنوعی باید سیستمی طراحی شود که دارای تعدادی ورودی باشد و با توجه به اهمیت هر یک، آن‌ها را با یکدیگر جمع جبری کند و توسط یک تابع موسوم به تابع انتقال^۱، آن‌ها را به نرون‌های دیگر ارسال نماید.

شکل (۳-۲) الگویی از یک نرون محاسباتی با توجه به نحوه‌ی عملکرد نرون طبیعی ارائه می‌دهد. با نگاهی به ساختار می‌توان ورودی‌ها را به دندریت، مجموعه جمع‌کننده و تابع محرک را به بدنه‌ی سلول، وزن‌ها^۲ را به شدت سیناپس‌ها و خروجی را به سیگنال گذرنده از اکسون تشبیه نمود [۵۰].



شکل (۳-۲) - ساختمان یک نرون محاسباتی

^۱ - Transfer function

^۲ - Weight

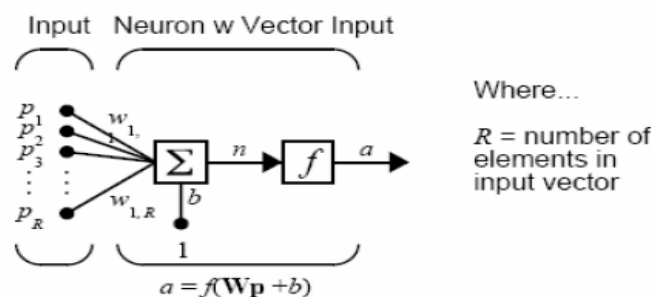
در شکل فوق کمیت‌های p و a به ترتیب ورودی و خروجی نرون می‌باشد. میزان تاثیر ورودی p روی خروجی a به وسیله پارامتر وزن (w) تعیین می‌شود. ورودی دیگر یک مقدار ثابت 1 است که در جمله بایاس^۱ ضرب شده و سپس با wp جمع می‌شود. این حاصل جمع، ورودی خالص n برای تابع محرک f را تشکیل می‌دهد.

$$n = wp + b \quad (2-2)$$

سپس خروجی نرون به کمک معادله زیر به دست می‌آید:

$$a = f(wp + b) \quad (3-2)$$

و در حالتی که طبق شکل (2-4) ورودی نرون به صورت یک بردار باشد، خروجی نرون از طریق روابط (2-4) و (2-5) به دست می‌آید [51 و 52].



شکل (2-4) - نرون با یک بردار ورودی

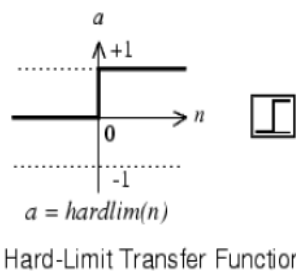
$$n = w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b \quad (4-2)$$

$$a = f(w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b) \quad (5-2)$$

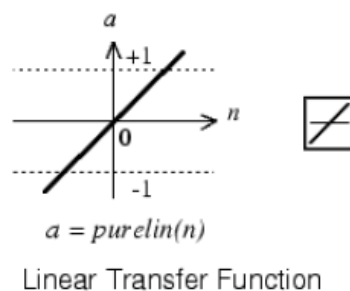
یکی از اجزای اصلی شبکه‌های عصبی تابع انتقال مورد استفاده می‌باشد که بر حسب کاربرد، می‌تواند توابع متفاوتی باشد. در برخی کاربردها، پاسخ مسئله از نوع دودویی است یعنی مسئله به گونه‌ای است که خروجی شبکه باید چیزی مانند آری یا نه باشد، در چنین مواردی نمی‌توان صرفاً بر جمع جبری سیگنال‌های ورودی تکیه کرد، بلکه ویژگی‌های خواسته شده در تابع انتقال گنجانده می‌شود. مثلاً اگر

¹ - Bias

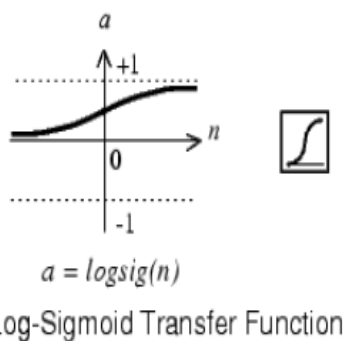
قرار باشد خروجی تابع فقط یکی از مقادیر صفر یا یک باشد، از تابع آستانه‌ای حدی^۱ استفاده می‌شود. بر اساس شکل (۲-۵ الف) در این تابع اگر آرگومان ورودی n (ورودی تابع انتقال) کوچکتر از صفر باشد، مقدار خروجی (a) این تابع صفر می‌شود و اگر آرگومان ورودی n بزرگتر یا مساوی از صفر باشد، خروجی برابر یک می‌شود. از توابع انتقال مشتق پذیری که مخصوصاً در شبکه‌های عصبی پس انتشار به کار می‌رود و در این تحقیق نیز مورد استفاده قرار گرفته، می‌توان به توابع انتقال خطی، لگاریتم سیگموئیدی و تانژانت سیگموئیدی اشاره کرد. تابع انتقال خطی (شکل (۲-۵ ب)) همان مقدار ورودی را به عنوان خروجی برمی‌گرداند ولی تابع لگاریتم سیگموئیدی ($\log \text{sig}$) مقادیر ورودی را در محدوده‌ی $-\infty$ تا $+\infty$ دریافت کرده و خروجی بین صفر و یک را تولید می‌کند و بر اساس شکل (۲-۵ د) تابع تانژانت سیگموئیدی ($\tan \text{sig}$) خروجی بین ۱ و -۱ تولید می‌کند [۵۰-۵۲].



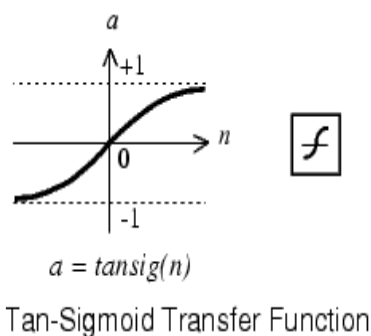
الف



ب



ج



د

شکل (۲-۵) - انواع توابع انتقال

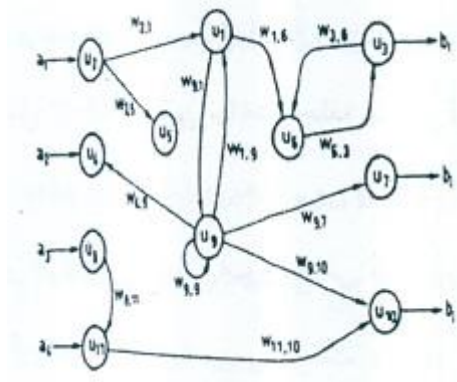
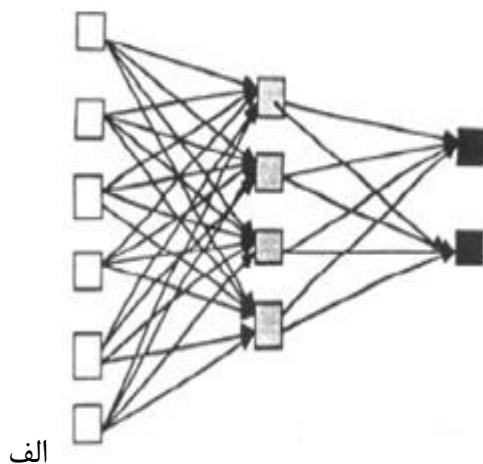
^۱ - Hard limit

۲-۷-۴- انواع شبکه‌های عصبی از نظر برگشت‌پذیری

❖ شبکه‌های پیشخور^۱

❖ شبکه‌های برگشتی^۲

شبکه‌های پیشخور، شبکه‌هایی هستند که مسیر پاسخ در آن‌ها همواره رو به جلو پردازش می‌شود و به نرون‌های لایه قبل باز نمی‌گردد. در این نوع شبکه‌ها به سیگنال اجازه می‌دهد که در مسیر یکطرفه عبور کند یعنی از ورودی تا خروجی. بنابراین بازخوردی^۳ وجود ندارد بدین معنی که خروجی هر لایه تأثیری بر همان لایه و همچنین لایه‌های قبلی ندارد. شکل (۲-۶ الف) نمونه‌ای از شبکه پیشخور را ارائه می‌دهد.



شکل (۲-۶)- انواع شبکه‌های عصبی: الف: شبکه عصبی پیشخور ب: شبکه عصبی برگشتی

ولی طبق شکل (۲-۶ ب) در شبکه‌های برگشتی حداقل یک سیگنال برگشتی از یک نرون به همان نرون یا نرون‌های دیگر همان لایه و یا نرون‌های لایه قبل وجود دارد. به عبارتی در این نوع شبکه‌ها بازخورد وجود دارد [۵۰ و ۵۳].

^۱ - Feed forward

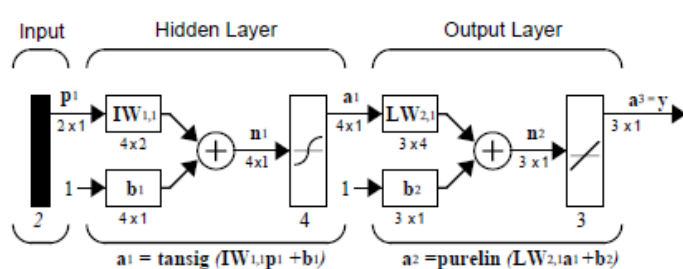
^۲ - Recurrent

^۳ - Feedback

۲-۷-۵- معماری شبکه‌های پیشخور

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

طبق شکل (۷-۲)، شبکه‌های پیشخور اغلب از یک یا چند لایه مخفی از نرون‌های سیگموئیدی و از یک لایه پایانی خطی استفاده می‌کنند. شبکه چند لایه از نرون‌ها با یک تابع انتقال غیرخطی به شبکه اجازه می‌دهد که توانایی یادگیری رابطه خطی و غیرخطی را بین ورودی‌ها و خروجی‌ها داشته باشد. لایه خروجی خطی به شبکه این امکان را می‌دهد که خروجی در هر محدوده دلخواهی از $-\infty$ تا $+\infty$ باشد. البته اگر بخواهیم خروجی در دامنه محدودی قرار گیرد، می‌توان از توابع محدودسازی مانند توابع سیگموئیدی در لایه خروجی استفاده کرد [۵۲].



P : بردار ورودی‌ها

w : ماتریس وزن‌ها

f : تابع انتقال

b : بایاس

n : ورودی تابع انتقال

a : خروجی تابع انتقال

شکل (۷-۲) - شبکه پیشخور tansig/pureline

¹ - Input layer

² - Output layer

³ - Hidden layer

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

به طور کلی مراحل که در فرایند آموزش دنبال می‌شوند، شامل ایجاد شبکه و تعیین مجموعه داده‌های آموزش، مقدار دهی آغازین به وزن‌ها، آموزش شبکه و شبیه‌سازی^۱ پاسخ شبکه می‌باشد [۵۲].

۲-۲-۷-۶-۱- ایجاد شبکه

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

۲-۲-۷-۶-۲- مقدار دهی آغازین به وزن‌ها

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

۲-۲-۷-۶-۳- آموزش شبکه

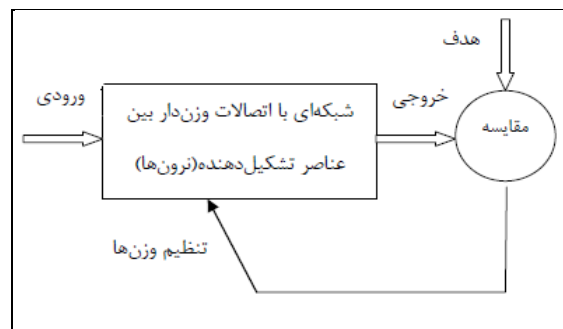
پس از مقدار دهی به وزن‌ها و بایاس‌ها نوبت به آموزش شبکه می‌رسد. فرایند آموزش به یک سری مثال‌ها از رفتار مورد انتظار شبکه نیاز دارد که شامل ورودی شبکه و مقادیر هدف آن ورودی می‌باشد. در طول آموزش وزن‌ها و بایاس‌ها تنظیم می‌شوند تا تابع کارایی حداقل شود. تابع کارایی پیش‌فرض برای شبکه‌های پیشخور میانگین مربع خطا (MSE) می‌باشد.

¹ - Simulation

² - Initialization function

۲-۲-۷-۶-۴- روش‌های آموزش

برای آموزش شبکه‌های عصبی دو روش آموزش با ناظر^۱ و آموزش بدون ناظر^۲ مورد استفاده قرار می‌گیرند. در شیوه آموزش با ناظر، مجموعه‌ای از زوج داده‌ها که به داده‌های آموزش معروفند، در اختیار شبکه گذاشته می‌شود. هر داده آموزشی شامل ورودی و خروجی متناظر با آن ورودی می‌باشد. در این شیوه یادگیری، پس از اعمال ورودی‌ها به شبکه، خروجی شبکه با خروجی مطلوب مقایسه می‌گردد (شکل (۲-۸)) و پس از محاسبه‌ی خطای یادگیری، از آن برای اصلاح پارامترهای شبکه (وزن‌ها و بایاس‌ها) استفاده می‌گردد، به گونه‌ای که اگر دفعه بعد همین ورودی به شبکه اعمال گردد، خروجی شبکه به خروجی مطلوب نزدیک‌تر باشد [۴۴ و ۵۳]. در حالی که در آموزش بدون ناظر، تنها سیگنالی که از محیط به شبکه اعمال می‌گردد، بردار ورودی شبکه است، بنابراین ناظری وجود نخواهد داشت تا خروجی مطلوب را برای آموزش به شبکه اعمال نماید. شبکه‌هایی را که از این نوع یادگیری بهره می‌برند، اصطلاحاً شبکه‌های خود سازمان‌ده^۳ می‌گویند [۵۳].



شکل (۲-۸) - شیوه آموزش با ناظر

۲-۲-۷-۷- آموزش شبکه‌های پیشخور با تکنیک پس انتشار^۴

پس انتشار خطا یک روش متداول آموزش با ناظر برای شبکه‌های پیشخور است یعنی برای به دست آوردن ارتباط بین متغیرهای ورودی و خروجی در یادگیری به الگوی آموزشی نیاز است.

^۱ - Supervised training

^۲ - Unsupervised training

^۳ - Self organized

^۴ - Back propagation

بنابراین، همانطور که در شکل (۲-۹) مشاهده می‌شود، باید داده‌های آموزشی (زوج ورودی-هدف) به شبکه ارائه شود. دو روش مختلف برای ارائه الگوهای آموزشی و پیاده‌سازی الگوریتم وجود دارد:

❖ آموزش گام به گام^۱

❖ آموزش دسته‌ای^۲

در آموزش گام به گام وزن‌ها و بایاس‌ها بعد از اعمال هر ورودی به روز^۳ می‌شوند در حالی که در روش دسته‌ای پس از اعمال تمام ورودی‌ها (اعضای مجموعه آموزشی) عملیات به روزرسانی وزن‌ها انجام می‌شود. بدین صورت که شیب‌های محاسبه شده برای هر ورودی با هم جمع می‌شوند تا در نهایت وزن‌ها و بایاس‌ها از طریق آن به روز شود. در این تحقیق شیوه آموزش دسته‌ای مورد استفاده قرار گرفته است [۵۲].

به طور کلی آموزش به کمک تکنیک پس‌انتشار بر طبق مراحل زیر انجام می‌شود [۴۴]:

- ۱- انتشار ورودی‌ها از نرون‌های ورودی به سمت نرون‌های خروجی
 - ۲- اختصاص ماتریس وزن‌های تصادفی به هریک از اتصالات
 - ۳- مقایسه خروجی‌های شبکه با مقادیر واقعی (مقادیر هدف) و محاسبه‌ی خطای شبکه
 - ۴- پس‌انتشار خطا از نرون‌های خروجی به سمت نرون‌های ورودی و اصلاح وزن‌ها
 - ۵- ارزیابی عملکرد شبکه با توجه به تابع کارایی تعیین شده
- مراحل فوق تا زمانی تکرار می‌شود که به حداکثر تکرار^۴ مجاز رسیده باشد یا مقدار تابع کارایی از مقداری که تعیین شده کمتر باشد.

به دلیل فرایندی که در طی این تکنیک رخ می‌دهد، این روش انتشار به عقب نامیده می‌شود زیرا خطای خروجی برای اصلاح وزن‌ها به شبکه برگشت داده می‌شود. و ابتدا وزن‌های ارتباطی بین لایه

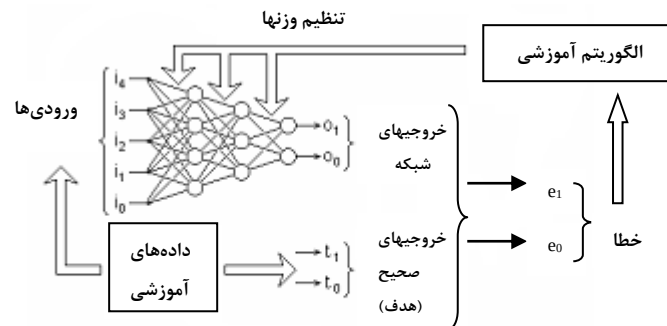
^۱ - Incremental training

^۲ - Batch training

^۳ - Update

^۴ - Epoch

خروجی و لایه پنهان و سپس وزن‌های ارتباطی بین لایه پنهان و لایه ورودی اصلاح می‌گردند. شکل (۹-۲) نیز گویای این مطلب است.



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

متداول‌ترین الگوریتمی که در الگوی آموزشی پس‌انتشار به کار می‌رود، الگوریتم آموزشی نزول گرادیانی^۱ است. در این الگوریتم آموزشی پارامترهای شبکه از قبیل وزن‌ها و بایاس‌ها در جهتی که تابع کارایی حداقل شود، تغییر می‌کند. به روز رسانی وزن‌ها در گام آموزشی $(n+1)$ به صورت زیر است:

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

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

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

^۱ - Gradient descent

نزول گرادیانی با وجود متداول بودنش دارای دو ضعف عمده و اساسی است. اول سرعت همگرایی پایین [۵۴] و دوم ساکن شدن و متوقف شدن پارامترهای شبکه به دلیل به دام افتادن آنها در نقاط مینیمم محلی^۱ [۵۵ و ۵۶] قبل از پایان یادگیری کامل شبکه می‌باشد.

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

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

$$g(x) = \nabla F(x) \approx 2J^T(x)E(W) \quad (۹-۲)$$

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

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

طبق این الگوریتم، تغییر پارامترها در هر گام آموزشی برابر است با [۵۴ و ۵۷]:

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

^۱ - Local minimum

^۲ - Levenberg-Marquardt

این الگوریتم حالت بینابین روش نزول گرادیانی و روش گوس- نیوتن است که در آن I ماتریس یکانی و μ_K ضریب یادگیری مارکوات و عددی غیر منفی می باشد. هر قدر μ_K بزرگ تر شود، نحوه آموزش به نزول گرادیانی با سرعت یادگیری کوچک ($\frac{1}{2\mu_k}$) نزدیک تر می شود:

$$x_{k+1} = x_k - \frac{1}{\mu_k} J^T(x_k) E(x_k) = x_k - \frac{1}{2\mu_k} g(x_k) \quad (12-2)$$

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

$$x_{k+1} = x_k - [J^T(x_k) J(x_k)]^{-1} J^T(x_k) E(x_k) \quad (13-2)$$

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

۱- ارائه ورودی به شبکه و تعیین مقدار اولیه وزن ها و μ

۲- محاسبه خروجی شبکه و خطای مربوط به آن

۳- محاسبه ماتریس ژاکوبین

۴- به روز کردن وزن ها بر طبق رابطه (2-11)

۵- محاسبه مجدد خطای شبکه

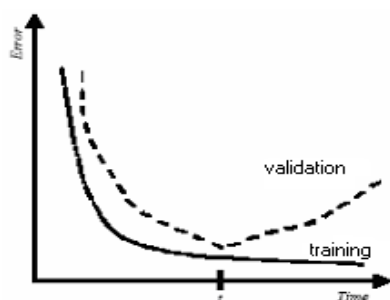
اگر خطای محاسبه شده در این حالت کمتر از حالت اول باشد، مقدار μ توسط μ_{dec} کاهش داده می شود و سپس با توجه به مقدار جدید μ و مقادیر جدید وزن ها به مرحله اول برمی گردد و چنانچه خطا افزایش یافته باشد، نتایج مربوط به آن دور نادیده گرفته می شود (تغییرات مربوط به وزن ها اعمال نمی شود) و μ در فاکتوری بزرگتر (μ_{inc}) ضرب می شود و به مرحله ۴ برمی گردد.

تکرار مراحل این الگوریتم تا زمانی ادامه دارد که تابع کارایی به کمترین مقدار خود برسد و یا به حداکثر تعداد چرخه های آموزش رسیده باشد.

۲-۷-۸- بهبود تعمیم^۱

یکی از مشکلات در آموزش شبکه‌های عصبی، بیش برآزش^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 ولی با ارائه داده‌های جدید به عنوان ورودی، خطای شبکه بسیار بالا می‌باشد [۵۸ و ۲۶]. 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^۳، تنظیم^۴ وجود دارد [۲۶].

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 (۲-۱۰) خطای مجموعه ارزیابی شروع به افزایش می‌کند که در این هنگام باید آموزش متوقف شود [۵۲].



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

^۱ - Improving generalization

^۲ - Over fitting

^۳ - Early stoping

^۴ - Regularization

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.

$$F = \beta E_D + \alpha E_W \quad (14-2)$$

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 (15-2) و (16-2) تعریف می‌شوند.

$$E_D = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - t_i)^2 \quad (15-2)$$

$$E_W = \frac{1}{N} \sum_{j=1}^N (w_j)^2 \quad (16-2)$$

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 [59] 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 (14-2))

۳- محاسبه تعداد مؤثر پارامترهای تابع کارایی (γ)

$$\gamma = N - 2\alpha \text{tr}(H)^{-1} \quad (17-2)$$

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 (۱۸-۲) و (۱۹-۲)

$$\alpha = \frac{\gamma}{2E_w} \quad (18-2)$$

$$\beta = \frac{n - \gamma}{2E_D} \quad (19-2)$$

که در رابطه (۱۹-۲) 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.

۲-۸-۲-۲- ارزیابی مدل

برای اطمینان از اینکه مدل به دست آمده، مدل مناسبی است که توانایی پیشگویی نمونه‌های مختلفی از یک جمعیت را داراست، باید مدل را ارزیابی کرد. این ارزیابی از طریق شاخص‌های کمی است که به وسیله آن‌ها صحت نتایج ارائه شده توسط مدل مورد سنجش قرار می‌گیرند. برخی از این شاخص‌های کمی عبارتند از:

ضریب همبستگی^۱: ساده‌ترین راه برای بررسی میزان همبستگی دو یا چند متغیر، محاسبه آماره‌ی ضریب همبستگی آنهاست. ضریب همبستگی دو متغیر x, y با رابطه (۲۰-۲) تعریف می‌شود. مقدار این آماره بین ۱ تا ۱- متغیر است. مقدار بزرگ‌تر آن نشان دهنده‌ی این است که ارتباط خطی بیشتری میان متغیر وابسته و متغیرهای مستقل وجود دارد.

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (20-2)$$

^۱ - Correlation coefficient

ضریب تعیین^۱: به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و نشان‌دهنده‌ی نسبت تغییرات متغیر وابسته توضیح داده شده توسط متغیر مستقل است. به عنوان مثال R^2 برابر با ۰/۹۲۴۳ نشان می‌دهد که ۹۲/۴۳ درصد تغییرات در متغیر وابسته می‌تواند توسط متغیر مستقل توضیح داده شود. رابطه ریاضی مربوط به ضریب تعیین به صورت زیر است:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (21-2)$$

که SSR ^۲ طبق رابطه (۲۲-۲) بیانگر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (22-2)$$

SST ^۳ طبق رابطه (۲۳-۲) نشانگر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (23-2)$$

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

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (24-2)$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (25-2)$$

^۱ - Determination coefficient

^۲ - Sum Square Regression

^۳ - Sum Square Total

^۴ - Sum Square Error

طبق رابطه (۲-۲۵) اگر تمام مشاهدات بر روی خط برازش شده قرار گرفته باشند، یعنی به ازای تمام نقاط $y_i = \hat{y}_i$ باشد، مقدار R^2 برابر یک می‌شود و هرگونه انحرافی از این حالت باعث می‌شود که مقدار R^2 از یک کوچکتر شود.

ضریب تعیین تصحیح شده^۱: ضریب تعیین تصحیح شده یکی از پارامترهایی است که در ارزیابی و مقایسه مدل‌های مختلف با تعداد متفاوت متغیرهای پیشگو استفاده می‌شود و از آنجایی که R^2 (ضریب تعیین)، تعداد پارامترهای موجود در مدل را به حساب نمی‌آورد، بدین منظور از ضریب تعیین تصحیح شده طبق رابطه زیر برای انتخاب مدل برتر استفاده می‌شود:

$$R_{adj}^2 = 1 - \frac{n-1}{n-p-1} \cdot \frac{SSE}{SST} = 1 - (1 - R^2) \frac{n-1}{n-p-1} \quad (۲-۲۶)$$

که در این رابطه p تعداد متغیرهای مستقل و n تعداد ترکیبات مورد بررسی می‌باشد. **آماره F:** آزمون F یا آزمون فیشر در واقع آزمون معنی‌دار بودن آماری در تحلیل رگرسیون ساده و چند متغیره است و برابر با نسبت میانگین مربعات رگرسیون (MSR) به میانگین مربعات باقیمانده‌ها (MSE) است. بیان ریاضی آن به صورت زیر می‌باشد:

$$F = \frac{MSR}{MSE} = \frac{SSR/df_m}{SSE/df_{res}} \quad (۲-۲۷)$$

$$df_m = p, \quad df_{res} = n - p - 1$$

df_m درجه آزادی مدل، p تعداد متغیرهای مستقل مدل، n تعداد کل ترکیبات مربوط به مدل، df_{res} درجه آزادی باقیمانده‌ها است. و SSR ، SSE در روابط قبل توضیح داده شد.

^۱ - Adjusted determination coefficient

مجموع مربع باقیمانده ها^۱ (PRESS): برابر مجموع مربعات تفاوت بین مقدار کمیت مشاهده شده (y_i) و مقدار تخمین زده شده $(\hat{y}_i)$ است.

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (28-2)$$

خطای استاندارد پیش‌بینی^۲ (SEP)

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (29-2)$$

خطای مطلق میانگین^۳ (MAE)

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (30-2)$$

خطای نسبی پیش‌بینی^۴ (REP)

$$REP(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (31-2)$$

میانگین مربع خطاها^۵ (MSE)

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (32-2)$$

میانگین خطای نسبی^۶ (MRE)

$$MRE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \times 100 \quad (33-2)$$

¹ - Predictive Residual Sum of Squares

² - Standard Error of Prediction

³ - Mean Absolute Error

⁴ - Relative Error of Prediction

⁵ - Mean Square Error

⁶ - Mean Relative Error

تولرانس^۱: همانطور که می‌دانیم وجود یا عدم وجود همخطی^۲ در بین متغیرهای پیشگو مسئله مهمی است که باید بررسی شود. پدیده همخطی می‌تواند به یکی از دو دلیل زیر باشد:

❖ حداقل دو تا از متغیرهای مستقل از لحاظ خطی خیلی به هم وابسته باشند. یعنی قدر مطلق ضریب همبستگی بین آن دو خیلی به یک نزدیک باشد.

❖ حداقل یکی از متغیرهای مستقل وابستگی شدیدی به مجموعه سایر متغیرهای مستقل داشته باشد که معمولاً این مورد مهمتر است.

لذا برای کشف پدیده همخطی برای هر متغیر مستقل مقداری محاسبه می‌شود که به آن تولرانس گفته می‌شود. تولرانس برای هر متغیر x_i از رابطه زیر به دست می‌آید [۴۵]:

$$T_i = 1 - R_i^2 \quad i = 1, 2, 3 \dots p \quad (۲-۳۴)$$

که در آن p تعداد متغیرهای پیشگو و R_i^2 مجذور ضریب همبستگی چندگانه است که از رگرسیون x_i بر تمام متغیرهای پیشگوی دیگر به دست می‌آید. پیشنهاد شده است که اگر برای هر یک از متغیرهای مستقل، T_i از ۰/۱ بیشتر باشد، پدیده همخطی وجود ندارد.

عامل افزایش واریانس^۳: در برخی مواقع از عامل افزایش واریانس به جای تولرانس استفاده می‌شود که رابطه آن با تولرانس به صورت زیر است:

$$VIF = \frac{1}{T_i} = \frac{1}{1 - R_i^2} \quad i = 1, 2, 3, \dots p \quad (۲-۳۵)$$

واضح است که اگر x_i رابطه خطی با سایر متغیرها داشته باشد، آنگاه R_i^2 نزدیک به یک است و عامل افزایش واریانس بزرگ می‌شود. مقادیر VIF_i بیشتر از ۱۰ اغلب به عنوان علامتی از اینکه داده‌ها مشکل همخطی دارند، تلقی می‌شود. اگر هیچ‌گونه رابطه خطی بین متغیرهای پیشگو نباشد، آنگاه R_i^2

^۱ - Tolerance

^۲ - Multicollinearity

^۳ - Variance Inflation Factor (VIF)

صفر و VIF_i برابر یک است. انحراف از یک، نشانه گرایش به همخطی در بین متغیرهای مستقل است [۶۰ و ۶۱].

FIT: به عنوان یک پارامتر آماری برای مقایسه مدل‌های مختلف با تعداد متفاوت توصیفگر به کار می‌رود که هر چه مقدار این پارامتر برای یک مدل بیشتر باشد، آن مدل مناسب‌تر می‌باشد. مقدار آن از رابطه زیر به دست می‌آید [۶۲ و ۷۲]

$$FIT = \frac{R^2(n - p - 1)}{(n + p^2)(1 - R^2)} \quad (۳۶-۲)$$

که R^2 مجذور ضریب همبستگی مدل، p تعداد توصیفگرهای مدل و n تعداد ترکیبات مربوط به مدل است.

۳-۲- نرم‌افزارهای مورد استفاده

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

۳-۲-۱- بسته نرم‌افزاری Hyperchem

از بسته نرم‌افزاری Hyperchem [۳۸] برای رسم شکل مولکول‌ها و بهینه‌سازی ساختار با استفاده از روش‌های کوانتومی و مکانیکی، استفاده می‌شود. به کمک این برنامه می‌توان طول پیوند، زاویه پیوندی و زوایای پیچشی را در مولکول تعیین کرد. داده‌های حاصل از این نرم‌افزار را می‌توان به عنوان ورودی به سایر نرم‌افزارها معرفی نمود. به عنوان مثال ساختار سه بعدی مولکول‌ها که با استفاده از روش مکانیک کوانتومی در Hyperchem بهینه شده است، را جهت محاسبه توصیفگرهای مختلف به

وسیله نرم‌افزار Dragon، به کار برد. همچنین به کمک این نرم‌افزار می‌توان تعدادی از توصیفگرها از جمله حجم مولی و قطبش‌پذیری را محاسبه کرد.

۲-۳-۲- بسته نرم‌افزاری Dragon

نرم‌افزار Dragon که توسط گروه تحقیقاتی کمومتریکس میلانو ارائه شده [۶۳]، امکان محاسبه ۱۴۸۱ توصیفگر مختلف را فراهم می‌کند. جهت محاسبه توصیفگر به کمک این نرم‌افزار لازم است ساختار هندسی بهینه مولکول مورد استفاده قرار گیرد. برای این منظور می‌توان ساختار بهینه مولکول‌ها را به صورت فایل‌هایی با فرمت hin, solf, mol ... به عنوان اطلاعات ورودی به کار برد.

۳-۳-۲- بسته نرم‌افزاری SPSS^۱

SPSS [۶۴] که نخستین نسخه آن در سال ۱۹۷۰ توسط جمعی از فارغ التحصیلان دانشگاه استنفورد آمریکا ارائه شده، امکان تجزیه و تحلیل آماری داده‌ها را فراهم می‌آورد. برخی از قابلیت‌های این بسته نرم‌افزاری عبارتند از:

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

^۱ - Statistical Package for the Social Science

۲-۳-۴- نرم افزار MATLAB

نرم افزار MATLAB [۶۵] یکی از جامع ترین و کارآمد ترین نرم افزارهای علمی و محاسباتی است که طی چند سال گذشته تهیه و به بازار عرضه شد و در سال های اخیر با تدوین نسخه های جدیدتر و کامل تر روز به روز بر غنای آن افزوده می شود.

MATLAB^۱ به معنای آزمایشگاه ماتریس است. ورودی ها اساساً به صورت ماتریس در نظر گرفته می شوند و هیچ نیازی به مشخص کردن ابعاد ماتریس نمی باشد. در MATLAB حتی اعداد اسکالر، ماتریس های (x) به حساب می آیند و بردارها، حالت خاصی از ماتریس های سطری یا ستونی در نظر گرفته می شوند. از جمله کاربردهای جالب آن که مورد توجه شیمیدانان قرار گرفته است شبکه های عصبی مصنوعی است که در این پژوهش از آن استفاده شده است. اطلاعات شیمیایی به عنوان ورودی و به صورت یک ماتریس به نرم افزار داده می شود. در محیط این برنامه با استفاده از آرایه ها و فرامین موجود امکان مدل سازی غیرخطی فعالیت بیولوژیکی با ساختار ترکیبات فراهم می گردد.

¹ - Matrix Laboratory

فصل سوم

مطالعه ارتباط کمی ساختار - فعالیت مشتقات

تیواوره با استفاده از رگرسیون خطی چندگانه و

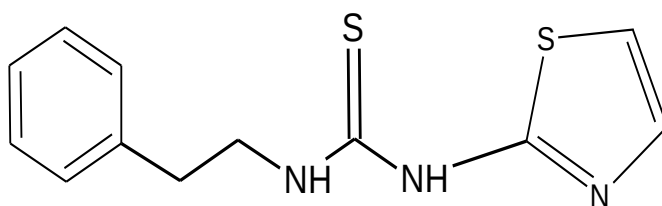
شبکه عصبی مصنوعی

۳-۱- مدل سازی فعالیت دارویی مشتقات تیواوره به عنوان بازدارنده های غیرنوکلئوزیدی ویروس HIV

به منظور مطالعه ارتباط کمی ساختار- فعالیت از شبکه عصبی مصنوعی و رگرسیون خطی چندگانه به عنوان روش های مدل سازی برای پیش بینی فعالیت ضد HIV یک طبقه ی جدید از مشتقات ترکیب N-(۲-فنتیل)-N'-(۲-تiazولیل) تیواوره^۱، استفاده شد و توانایی مدل های به دست آمده از هر دو روش در پیش بینی قدرت بازدارندگی این مشتقات، مورد ارزیابی قرار گرفت. به طور کلی بخش تجربی این تحقیق شامل انتخاب سری داده ها، رسم و بهینه سازی ساختار مولکول ها، محاسبه توصیفگرهای مولکولی، انتخاب توصیفگرهای مولکولی مناسب، مدل سازی و ارزیابی مدل برتر است.

۳-۱-۱- انتخاب سری داده ها

سری داده ها شامل فعالیت دارویی ۴۹ ترکیب از مشتقات N-(۲-فنتیل)-N'-(۲-تiazولیل) تیواوره^۱ می باشد که توسط فرانک^۲ و همکارانش گزارش شده است [۶۶].



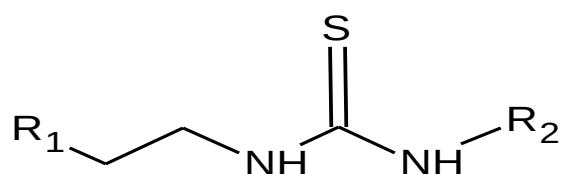
شکل (۳-۱)- ساختار N-(۲-فنتیل)-N'-(۲-تiazولیل) تیواوره

^۱ - N-(2-phenethyl) - N'-(2-thiazolyl) thiourea

^۲ - Frank W. Bell

اسکلت اصلی ترکیبات مورد بررسی در شکل (۲-۳) ارائه شده است. در جدول (۱-۳) نیز جزئیات استخلاف‌ها و مقادیر عددی pIC_{50} بر حسب مولار برای هر ترکیب آورده شده است. IC_{50}^1 غلظتی از دارو است که ۵۰٪ اثر بازدارندگی داشته باشد. به عبارت دیگر به غلظتی از ترکیب (دارو) که از فعالیت آنزیم نسخه بردار معکوس^۲ در فرایند نسخه برداری از ژنوم RNA ویروس HIV تا حدود ۵۰٪ جلوگیری کند، IC_{50} گفته می‌شود. ذکر این نکته ضروری است که هرچه دارو قویتر باشد، غلظت کمتری از آن برای داشتن ۵۰٪ اثر بازدارندگی، نیاز است.

داده‌های جدول نشان می‌دهد که نوع استخلاف و نحوه آرایش آنها 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، بدون اینکه نیازی به سنتز و آزمایش سلولی ترکیب مورد نظر باشد.

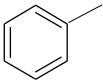
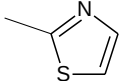
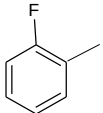
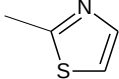
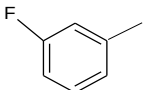
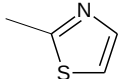
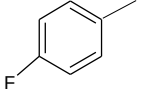
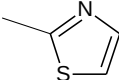
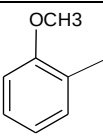
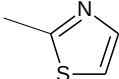
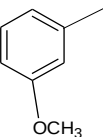
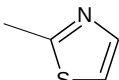
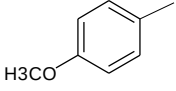
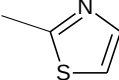
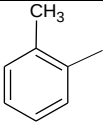
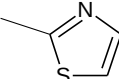
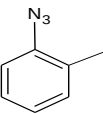
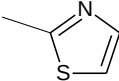
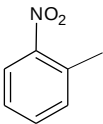
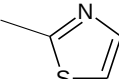
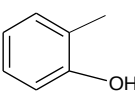
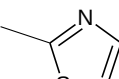
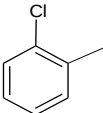
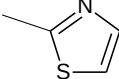


شکل (۲-۳) - اسکلت اصلی ترکیبات مورد مطالعه

¹ - Half maximal inhibitory concentration

² - Reverse transcriptase enzyme

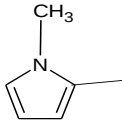
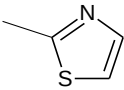
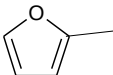
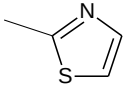
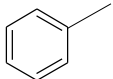
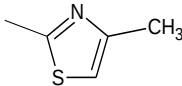
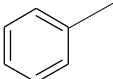
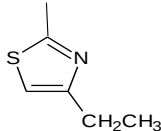
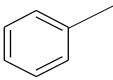
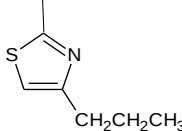
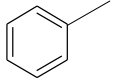
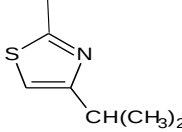
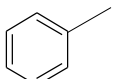
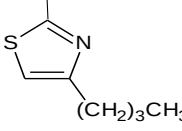
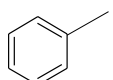
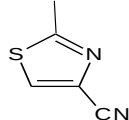
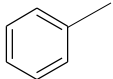
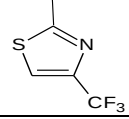
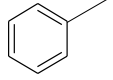
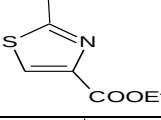
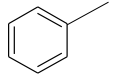
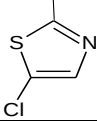
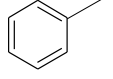
جدول (۳-۱) - ساختار و مقادیر pIC_{50} ترکیبات مورد مطالعه

شماره ترکیب	R_1	R_2	pIC_{50}
۱			۶/۰۴
۲			۷/۲۲
۳			۶/۸۲
۴			۶/۰۰
۵			۷/۰۴
۶			۶/۸۲
۷			۶/۴۶
۸			۷/۱۰
۹			۷/۵۲
۱۰			۶/۸۲
۱۱			۵/۹۶
۱۲			۶/۲۲

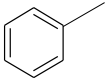
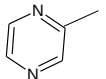
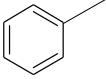
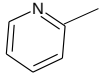
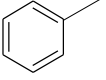
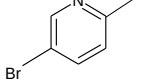
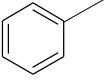
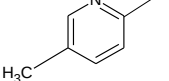
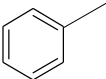
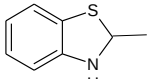
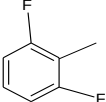
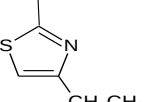
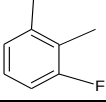
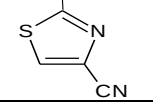
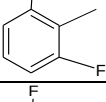
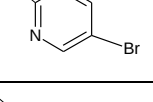
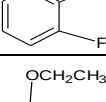
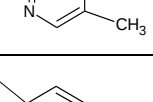
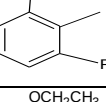
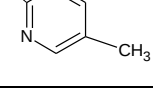
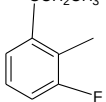
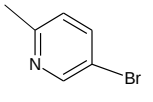
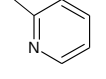
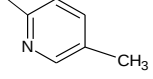
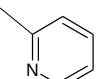
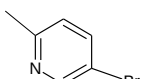
ادامه جدول (۳-۱) - ساختار و مقادیر pIC_{50} ترکیبات مورد مطالعه

شماره ترکیب	R_1	R_2	pIC_{50}
۱۳			۷/۲۲
۱۴			۶/۷۰
۱۵			۶/۴۰
۱۶			۵/۹۶
۱۷			۷/۰۵
۱۸			۶/۷۰
۱۹			۷/۵۲
۲۰			۸/۰۰
۲۱			۸/۰۰
۲۲			۸/۰۰
۲۳			۸/۲۲
۲۴			۶/۷۰

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

شماره ترکیب	R_1	R_2	pIC_{50}
۲۵			۵/۷۲
۲۶			۶/۱۹
۲۷			۷/۰۰
۲۸			۶/۲۲
۲۹			۶/۴۶
۳۰			۶/۷۰
۳۱			۵/۶۰
۳۲			۶/۷۰
۳۳			۶/۲۶
۳۴			۶/۷۰
۳۵			۵/۶۲
۳۶			۵/۷۲

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

شماره ترکیب	R_1	R_2	pIC_{50}
۳۷			۵/۴۱
۳۸			۷/۷۰
۳۹			۷/۸۲
۴۰			۷/۵۲
۴۱			۶/۷۰
۴۲			۸/۲۶
۴۳			۸/۲۲
۴۴			۹/۰۰
۴۵			۸/۵۲
۴۶			۸/۳۵
۴۷			۸/۲۲
۴۸			۷/۳۰
۴۹			۷/۸۲

۳-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها

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

۳-۱-۳- محاسبه توصیفگرها

برای ایجاد ارتباط ریاضی بین ساختار-فعالیت، به مقادیر عددی نیاز است پس باید ساختار مولکول‌ها به مقادیر کمی تبدیل شود. بدین منظور ساختار بهینه شده ترکیبات به نرم‌افزار Dragon منتقل شد و برای هریک از مولکول‌ها ۱۴۸۱ توصیفگر در ۱۸ طبقه مختلف محاسبه گردید. جدول (۲-۳) فهرست این ۱۸ طبقه را نشان می‌دهد.

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

ردیف	نام توصیفگر	ردیف	نام توصیفگر
۱	توصیفگرهای زیر ساختاری	۱۰	توصیفگرهای هندسی
۲	توصیفگرهای توپولوژیکی	۱۱	توصیفگرهای RDF
۳	شمارنده‌های مولکولی	۱۲	توصیفگرهای سه بعدی موریس
۴	توصیفگرهای BCUT	۱۳	توصیفگرهای WHIM
۵	شاخص‌های بار توپولوژیکی Galvez	۱۴	توصیفگرهای GETAWAY
۶	خود ارتباطی‌های دوبعدی	۱۵	گروه‌های عاملی
۷	توصیفگرهای بار	۱۶	اجزای میان اتمی
۸	شاخص‌های آروماتیسیته	۱۷	توصیفگرهای تجربی
۹	پروفاایل‌های مولکولی راندیک	۱۸	خصوصیات مولکولی

^۱ - Root Mean Square Gradient

^۲ - Restricted Hartee-Fock

۳-۱-۴- انتخاب توصیفگرهای مناسب

مهمترین مرحله در مطالعات QSAR انتخاب توصیفگرهای مناسب است که بتوانند فعالیت دارویی را به ساختار مولکول ارتباط دهند. با عنایت به این موضوع و با توجه به اینکه تعداد زیاد توصیفگر باعث پیچیدگی محاسبات می‌شود و همچنین تعدادی از توصیفگرها حاوی اطلاعات یکسان و در نتیجه اثر یکسان بر روی فعالیت مهارکنندگی هستند، لذا باید روش کاهش تعداد متغیرها به کار گرفته شود. به همین منظور، ابتدا توصیفگرهایی که دارای مقادیر یکسانی برای کل مولکول‌ها بودند از فهرست توصیفگرها حذف شدند. سپس با بررسی همبستگی بین متغیرها، از میان هر دو توصیفگری که ضریب همبستگی بزرگتر از ۰/۹ داشتند، آنکه که همبستگی کمتری با متغیر وابسته داشت، حذف شد. سپس رگرسیون مرحله به مرحله توسط نرم افزار SPSS مورد استفاده قرار گرفت. در روش رگرسیون مرحله‌ای، متغیرها یکی پس از دیگری وارد مدل می‌شوند و ابتدا متغیری وارد مدل می‌شود که بالاترین میزان همبستگی را با متغیر وابسته دارد. با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آنها سطح معناداری خود را از دست داده باشند، قبل از ورود متغیر جدید از مدل خارج می‌شوند. به این ترتیب داده‌های pIC_{50} به عنوان متغیر وابسته و توصیفگرها به عنوان متغیر مستقل در نظر گرفته شدند و تکنیک رگرسیون مرحله‌ای انجام شد که در این مرحله فقط ۱۱ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط را با فعالیت دارویی دارند، شناخته شدند. این ۱۱ توصیفگر به همراه طبقه آنها در جدول (۳-۳) ارائه شده است.

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

No	Symbol	Class	Meaning
۱	nPhX	Functional group	Number of X-C on aromatic ring
۲	BEHp8	BCUT	Highest eigenvalue n.8 of Burden matrix
۳	DECC	Topological	Eccentric
۴	RDF105m	RDF	Radial Distribution Function -105
۵	Mor 26 p	3D MORSE	3D -MORSE-signal 26
۶	CIC2	Topological	Complementary information content
۷	R5u_A	GETAWAY	R autocorrelation of lag 5
۸	MATS 5p	2D autocorrelation	Moran auto correlation -lag 5
۹	Mor 15p	3D MORSE	3D -MORSE-signal 15
۱۰	Mor 28u	3D MORSE	3D -MORSE-signal 28
۱۱	ATS4m	2D autocorrelation	Broto- moreau auto correlation - lag4

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

همبستگی بین توصیفگرها را نشان می‌دهد.

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

	nPhX	BEHp8	DECC	RDF105m	Mor26p	CIC2	R5u_A	MATS5p	Mor15p	Mor28u	ATS4m
nPhX	۱										
BEHp8	-۰/۰۳۶	۱									
DECC	-۰/۱۲۶	۰/۶۳۳	۱								
RDF105m	۰/۱۵۴	۰/۴۳۲	۰/۴۳۲	۱							
Mor26p	-۰/۳۵۹	-۰/۳۰۱	-۰/۳۴۸	-۰/۴۴۰	۱						
CIC2	-۰/۳۸۹	۰/۵۵۶	۰/۵۰۶	۰/۳۵۵	-۰/۲۱۲	۱					
R5u_A	-۰/۱۰۴	-۰/۱۹۱	-۰/۴۹۵	-۰/۱۶۷	۰/۱۴۹	-۰/۱۰۶	۱				
MATS5p	-۰/۰۶۷	-۰/۴۲۶	۰/۰۲۲	۰/۲۴۸	۰/۰۴۰	۰/۱۴۰	۰/۰۰۳	۱			
Mor15p	۰/۲۴۵	۰/۵۱۱	۰/۰۴۷	۰/۲۲۲	-۰/۱۶۱	۰/۲۵۹	-۰/۰۰۶	-۰/۲۳۷	۱		
Mor28u	۰/۳۲۸	-۰/۵۶۰	-۰/۳۶۲	-۰/۲۷۲	۰/۰۶۴	-۰/۶۵۲	-۰/۱۱۳	۰/۴۰۲	-۰/۳۳۴	۱	
ATS4m	۰/۴۵۶	-۰/۳۶۳	-۰/۳۰۲	-۰/۱۵۹	۰/۲۲۸	۰/۶۲۹	-۰/۰۷۰	۰/۰۳۶	-۰/۰۳۷	۰/۶۹۹	۱

۳-۱-۵- بررسی عدم وجود همخطی بین کل توصیفگرهای انتخاب شده

به منظور بررسی وجود یا عدم وجود همخطی در بین توصیفگرهای انتخاب شده از کمیتی به نام

عامل افزایش واریانس (VIF) استفاده شد که مقدار این کمیت برای توصیفگرها در جدول (۳-۵)

گزارش شده است. با توجه به اینکه مقدار این کمیت برای همه توصیفگرها خیلی از ده کوچکتر است،

عدم وجود همخطی در بین توصیفگرها تایید می‌شود.

جدول (۳-۵) - مقادیر عامل افزایش واریانس کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه

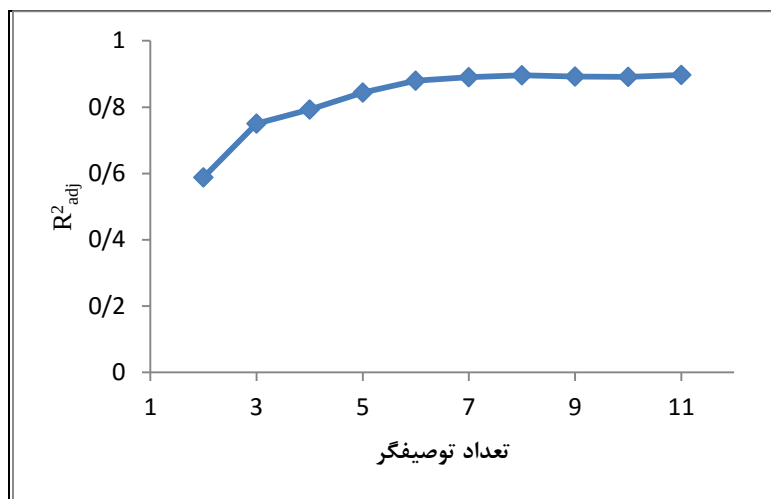
توصیفگر	nphx	BEHp8	DECC	RDF105m	Mor26p	CIC2	R5u_A	MATS5p	Mor15p	Mor28u	ATS4m
VIF	۲/۲۰	۴/۸۳	۵/۲۱	۱/۶۱	۲/۰۹	۲/۶۹	۱/۸۳	۲/۵۲	۲/۴۹	۵/۰۵	۳/۹۷

۳-۱-۶- مدل سازی توسط روش رگرسیون خطی چندگانه

برای ساختن مدل خطی مناسب که بیانگر ارتباط ساختاری ترکیبات مورد مطالعه با فعالیت بازدارندگی آنها باشد، مدل های مختلفی که توسط رگرسیون خطی چندگانه مرحله به مرحله به دست آمد، مورد بررسی قرار گرفتند. برای این کار داده ها به طور تصادفی به دو سری آموزش و تست تقسیم شدند. سری تست ۱۰ ترکیب و سری آموزش ۳۹ ترکیب را شامل می شوند. سپس توسط سری آموزش و روش ارزیابی تقاطعی قدرت پیشگویی مدل های مختلف مورد سنجش قرار گرفت. بدین صورت که ابتدا در مدل شامل دو توصیفگر، یک ترکیب از سری داده های آموزش کنار گذاشته شد، آنگاه با بقیه داده های سری آموزش و استفاده از روش Enter ضرائب توصیفگرهای وارد شده در مدل استخراج گردید و سپس با استفاده از این ضرائب مقدار فعالیت دارویی برای ترکیب کنار گذاشته شده، محاسبه شد. این فرایند برای همه داده های سری آموزش تکرار شد. سپس ضریب تعیین تصحیح شده (R^2_{adj}) و FIT برای کل داده های سری آموزش محاسبه گردید. در مراحل بعدی با مدل های ایجاد شده توسط دیگر توصیفگرها نیز سری آموزش مورد ارزیابی تقاطعی قرار گرفت. در بین مدل های ایجاد شده، مدلی که تعداد توصیفگر کمتر، FIT و ضریب تعیین تصحیح شده (R^2_{adj}) بیشتر دارد، به عنوان مدل برتر انتخاب می شود. مقادیر این پارامترها برای مدل های ۲ تا ۱۱ در جدول (۳-۶) ذکر شده است. روابط ریاضی مربوط به هر یک از این پارامترهای آماری در بخش (۲-۸) بیان شده است. همچنین در شکل (۳-۳) تأثیر تعداد توصیفگر بر مقدار R^2_{adj} نشان داده شده است.

جدول (۳-۶) - پارامترهای آماری برای مدل های به دست آمده از روش MLR

مدل	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
FIT	۱/۳۱	۲/۴۴	۲/۷۲	۳/۳	۳/۷۶	۳/۵۲	۳/۲۶	۲/۷۰	۲/۳۲	۲/۱۴
R^2_{adj}	۰/۵۸۸	۰/۷۵۰	۰/۷۹۳	۰/۸۴۴	۰/۸۸۰	۰/۸۹۰	۰/۸۹۶	۰/۸۹۲	۰/۸۹۱	۰/۸۹۷
R^2	۰/۶۱۰	۰/۷۷۰	۰/۸۱۵	۰/۸۶۵	۰/۸۹۸	۰/۹۰۹	۰/۹۱۸	۰/۹۱۸	۰/۹۲۰	۰/۹۲۷



شکل (۳-۳)-تأثیر تعداد توصیفگرها بر مقادیر R^2_{adj} مدل‌های MLR

بدین ترتیب مدل ۶ به عنوان بهترین مدل برگزیده شد که معادله آن به صورت زیر می‌باشد:

$$pIC_{50} = ۴/۴۱۷ + ۰/۵۳۴ \text{ nPhX} + ۳/۰۰۹ \text{ BEHp8} - ۲/۵۳ \text{ DECC} - ۰/۲۰۹ \text{ RDF105m} - ۳/۹۵۶$$

$$\text{Mor26p} - ۱/۰۷۵ \text{ CIC2} \quad \text{رابطه (۱-۳)}$$

ولی با توجه به اینکه واحدهای اندازه‌گیری متغیرهای مستقل (ضرائب رگرسیون) یکسان نیست، هرگز نمی‌توان از روی ضرائب رگرسیون غیر استاندارد، به میزان اهمیت و تأثیر یک متغیر مستقل بر روی متغیر وابسته پی برد. برای حل این مشکل می‌توان از ضریب استاندارد شده که بر طبق رابطه زیر محاسبه می‌شود [۴۵ و ۶۷]، استفاده نمود.

$$\beta'_k = \left(\frac{S_K}{S_Y}\right) \beta_K \quad \text{رابطه (۲-۳)}$$

در این رابطه β'_k ضریب رگرسیون استاندارد شده توصیفگر k ام، β_K ضریب رگرسیون غیر استاندارد همان توصیفگر، S_K انحراف استاندارد متغیر مستقل مورد نظر (توصیفگر k ام) و S_Y انحراف استاندارد متغیر وابسته (فعالیت دارویی) می‌باشد که از روابط زیر به دست می‌آیند:

$$S_K = \sqrt{\frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{n - 1}} \quad \text{رابطه (۳-۳)}$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad \text{رابطه (۴-۳)}$$

X_{ik} مقدار توصیفگر k ام برای ترکیب i ام و $\bar{X}_k$ مقدار میانگین متغیر مستقل (توصیفگر)، n تعداد ترکیبات، Y_i مقدار متغیر وابسته برای ترکیب i ام و $\bar{Y}$ مقدار میانگین متغیر وابسته است [۴۵].

بدین ترتیب در این تحقیق نیز بر اساس روابط فوق، ضرائب رگرسیون استاندارد شده برای مدل برتر محاسبه شد که در معادله زیر ارائه شده است.

$$pIC_{50} = 0.462 \text{ nPhX} + 0.801 \text{ BEHp8} - 0.448 \text{ DECC} - 0.280 \text{ RDF105m} - 0.351 \text{ Mor26p} - 0.290 \text{ CIC2} \quad \text{رابطه (۵-۳)}$$

۳-۱-۷- مدل سازی توسط شبکه عصبی مصنوعی

با توجه به پیچیدگی داده‌ها و احتمال وجود روابط غیرخطی بین متغیرهای مستقل (توصیفگرها) و متغیر وابسته (فعالیت دارویی ترکیبات) علاوه بر مدل‌سازی به روش رگرسیون خطی چندگانه، روشهای غیرخطی نیز برای مدل‌سازی فعالیت دارویی ترکیبات ضد HIV مورد استفاده قرار گرفت. از میان روشهای غیرخطی، شبکه‌ی عصبی به عنوان یک تکنیک برتر برای مطالعات QSAR ترکیبات جدید ضد HIV به کار می‌رود. شبکه عصبی انتخاب شده در این مطالعه یک شبکه پیشخور با الگوریتم آموزشی پس‌انتشار می‌باشد که الگوریتم آن با استفاده از نرم افزار مطلب ۷/۸ نوشته شد. مقادیر توصیفگرهای وارد شده در مدل خطی به عنوان ورودی به شبکه عصبی داده شد. مقدار تجربی فعالیت ضد HIV نیز به عنوان هدف به دنبال ورودی‌ها وارد شبکه شدند تا پاسخ شبکه با آنها سنجیده شود. برای ایجاد یک مدل ANN پیشگو با درصد خطای پیش بینی پایین، باید ابتدا پارامترهای مؤثر بر شبکه بهینه شوند که این پارامترها عبارتند از:

تعداد متغیرهای ورودی (تعداد توصیفگرها)

تعداد گره‌های لایه پنهان

نوع تابع انتقال

نوع تابع آموزش

تعداد دوره‌های آموزش

پارامتر μ (مقدار اولیه برای μ)

در فرایند بهینه کردن پارامترهای شبکه، سری داده‌ها به طور تصادفی به سری تست (شامل ۱۰ ترکیب) و سری آموزش (شامل ۳۹ ترکیب) تقسیم شدند. سری آموزش برای محاسبه‌ی گرادیان و به روز کردن اوزان و بایاس‌های شبکه با روش ارزیابی تقاطعی استفاده شد. در این روش، یک ترکیب از سری آموزش حذف شد، شبکه با ۳۸ ترکیب باقیمانده آموزش داده شد و سپس برای پیش‌بینی ترکیب کنار گذاشته شده، استفاده گردید. این فرایند برای هر ترکیب در سری آموزش تکرار شد به طوری که در طی فرایند آموزش، میانگین مربع خطا بر طبق رابطه (۳-۶) محاسبه گردید و معیار بهینه سازی، به حداقل رساندن مقدار E در نظر گرفته شد.

$$E = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad \text{رابطه (۳-۶)}$$

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

۳-۱-۷-۱- انتخاب تعداد لایه‌های پنهان

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 و ساختار شبکه‌های عصبی، اغلب به روش سعی و خطا ایجاد می‌شود. با این وجود در بیشتر موارد به نظر می‌رسد که یک لایه پنهان مناسب باشد [۲۱ و ۶۸]. بنابراین در این تحقیق نیز یک لایه پنهان مورد استفاده قرار گرفت.

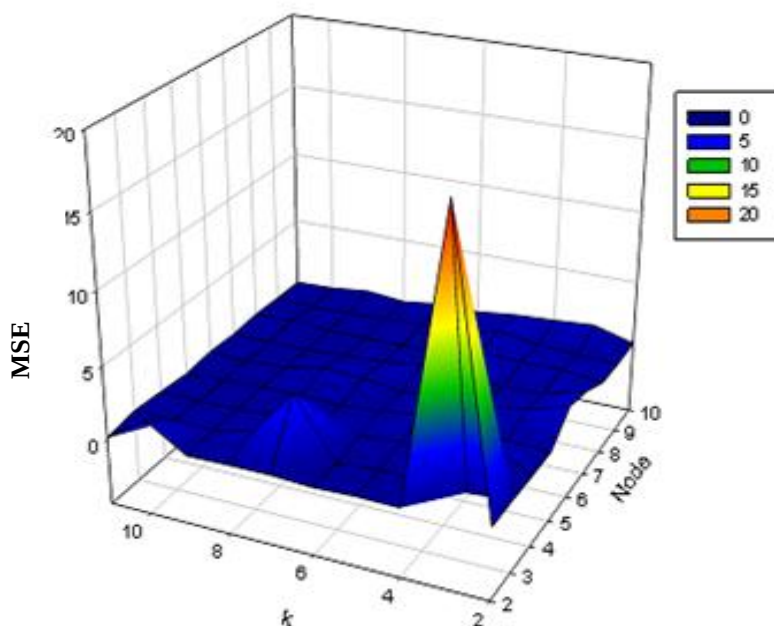
۳-۱-۷-۲- انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع

آموزش و نوع تابع انتقال

پس از طراحی اولیه شبکه، برای اینکه شبکه قدرت پیشگویی مناسبی داشته باشد باید به طور مناسب آموزش دیده باشد و با توجه به اینکه فرایند آموزش وقتگیر است، انتخاب الگوریتم آموزشی سریعتر امری مهم است. از میان الگوریتم‌های آموزشی متفاوت که در جعبه ابزار مطلب وجود دارد، الگوریتم لونبرگ - مارکوارت و تنظیم بایزین دارای سرعت مناسب، قابلیت پیشگویی و تعمیم پذیری بهتری نسبت به سایر روش‌های معمول مانند روش نزول گرادیانی می‌باشند. بر این اساس شبکه با ورودی‌های از ۲ تا ۱۱ توسط دو الگوریتم آموزشی لونبرگ - مارکوارت و تنظیم بایزین، با تعداد متفاوت گره در لایه پنهان از ۲ تا ۱۰ و همچنین توابع لگاریتم سیگموئیدی (logsig) و تانژانت سیگموئیدی (tansig)، به عنوان توابع انتقال لایه پنهان، آموزش داده شد. در تمامی این شبکه‌ها تعداد دوره‌های آموزش ثابت و برابر ۲۰ در نظر گرفته شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده شد. بدین ترتیب در مجموع ۳۶۰ شبکه سه لایه آموزش داده شد و همانطور که در بخش قبل گفته شد بهینه سازی از طریق ارزیابی تقاطعی بر روی سری آموزش انجام گرفت. معیار نیز به حداقل رساندن میانگین مربع خطا (MSE) در نظر گرفته شد که نتایج این محاسبات در جداول (۷-۳)، (۸-۳)، (۹-۳) و (۱۰-۳) گرد آوری شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۴-۳)، (۵-۳)، (۶-۳) و (۷-۳) ارائه شده است.

جدول (۷-۳) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

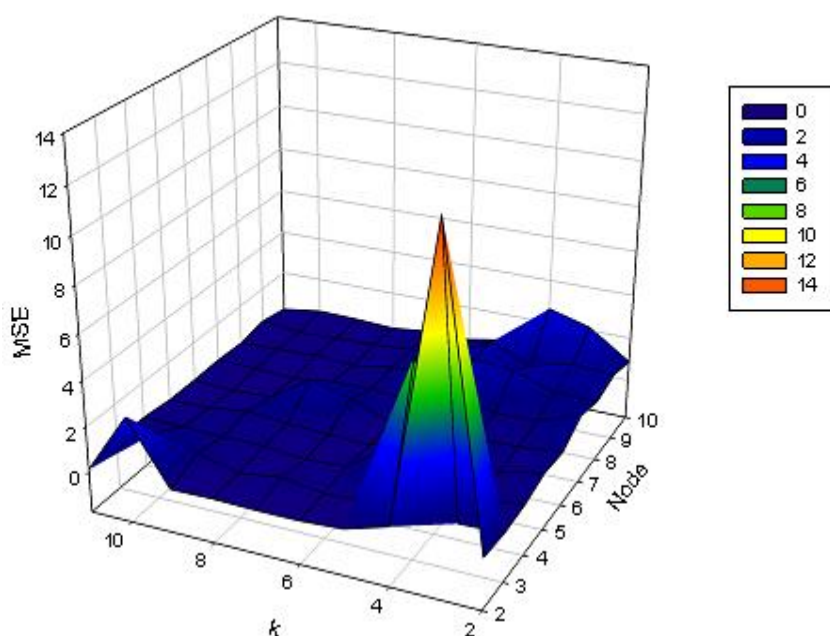
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره‌های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	۲	۰/۴۱۹۱	۱۹/۷۱۴	۰/۲۲۵۹	۰/۲۷۱	۰/۱۷۲	۰/۳۱۸۰	۰/۳۰۴۲	۰/۲۴۲۴	۲/۰۱۲	۰/۱۴۴۸
	۳	۰/۴۳۲۸	۰/۲۱۹۷	۰/۲۵۷۵	۰/۳۵۰۴	۰/۱۲۳۸	۴/۱۹۱۲	۰/۹۹۹۸	۰/۳۷۴۱	۰/۴۷۵۷	۰/۴۶۹۰
	۴	۰/۳۵۰۲	۱/۹۵۴۶	۰/۶۵۷۹	۰/۶۳۵۱	۰/۳۸۱۵	۰/۴۹۸۱	۰/۵۳۶۹	۰/۳۳۴۴	۰/۲۲۷۵	۰/۳۱۴۹
	۵	۰/۳۳۶	۰/۴۴۴۴	۰/۷۲۸۶	۰/۹۸۷۴	۰/۵۸۷۳	۰/۳۱۱۰	۰/۴۳۸۲	۰/۳۶۸۷	۰/۲۴۸۵	۰/۱۱۶۳
	۶	۱/۹۰۸۴	۱/۰۰۲۱	۰/۷۲۲۶	۰/۷۰۹۰	۰/۴۱۰۸	۰/۷۵۳۷	۰/۹۵۸۲	۰/۳۴۹۶	۰/۱۸۷۰	۰/۳۱۲۲
	۷	۱/۴۴۵۴	۰/۶۷۰۴	۰/۴۵۵۶	۰/۹۱۷۰	۰/۸۹۷۵	۰/۵۰۷۱	۰/۳۹۴۰	۰/۱۹۲۷	۰/۱۷۸۳	۰/۲۰۵۷
	۸	۰/۵۵۶۹	۰/۹۱۰۷	۰/۸۸۰۹	۰/۹۲۸۱	۰/۲۶۰۸	۰/۳۲۶۵	۰/۲۳۸۰	۰/۲۳۷۷	۰/۴۴۷۱	۰/۳۳۷۳
	۹	۰/۴۶۹۳	۰/۴۹۷۳	۰/۷۱۰۱	۱/۱۴۸۷	۰/۳۷۴۸	۰/۴۵۸۵	۰/۳۲۹۹	۰/۲۳۳۱	۰/۱۲۳۳	۰/۲۴۵۳
	۱۰	۰/۵۴۲۸	۱/۳۲۱۲	۱/۱۱۹۰	۰/۹۵۸۵	۰/۶۴۲۶	۰/۶۶۵۸	۰/۳۰۹۱	۰/۵۷۷۹	۰/۲۸۹۱	۰/۱۹۵۲



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

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

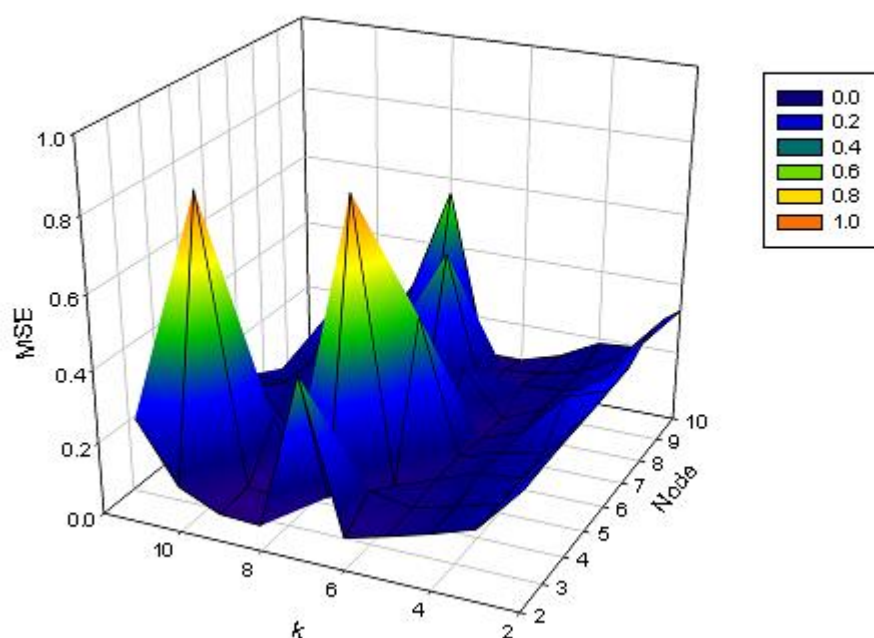
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	۲	۰/۴۵۴	۱۳/۳۵۰۷	۰/۸۰۴۵	۰/۲۳۲	۰/۱۲۰۲	۰/۱۱۵۷	۰/۱۵۴۹	۰/۱۱۷۱	۲/۸۵۴۵	۰/۱۲۷۸
	۳	۰/۳۳۸۲	۰/۲۶۹۱	۶/۷۷۶۱	۰/۲۸۲۷	۰/۱۹۸۴	۰/۴۶۶۸	۰/۲۲۱۸	۰/۶۷۳۳	۱/۱۷۸۶	۰/۳۴۳۳
	۴	۰/۴۳۴۵	۰/۲۴۸۱	۰/۳۹۸۷	۱/۶۰۶۴	۰/۳۴۲۱	۰/۱۹۱۶	۰/۳۸۸	۰/۳۸۴۵	۰/۴۱۲۸	۰/۴۳۹۳
	۵	۰/۷۲۱۴	۰/۳۲۷۱	۱/۳۰۰۸	۱/۱۶۴۳	۰/۲۴۷۸	۰/۳۱۶۶	۰/۳۹۹۲	۰/۵۵۹۷	۰/۳۰۸۵	۰/۲۸۰۲
	۶	۰/۶۰۹۹	۰/۴۲۹۱	۰/۳۹۰۲	۰/۶۳۴۲	۰/۲۷۶	۰/۹۶۸۸	۱/۴۴۶۳	۰/۲۶۸۸	۰/۱۳۳۸	۰/۲۶۵۷
	۷	۱/۲۳۵۵	۰/۵۸۳۲	۰/۵۵۰۵	۰/۹۳۳۸	۰/۲۶۳۶	۰/۴۵۸۹	۰/۴۰۷۵	۰/۱۶۴۱	۰/۲۵۱۶	۰/۲۹۴۳
	۸	۰/۹۲۶۴	۰/۹۰۱۲	۱/۵۸۹۱	۱/۶۸۱۳	۰/۶۲۶۴	۰/۵۵۵۵	۰/۲۳۸۴	۰/۳۶۱۷	۰/۲۹۰۲	۰/۲۱۵
	۹	۱/۱۹۲۴	۱/۳۱۱۴	۰/۸۳۰۹	۱/۵۹۸۳	۰/۵۶۳۷	۰/۴۹۱۴	۰/۳۵۱۹	۰/۲۷۵۴	۰/۲۱۵۰	۰/۴۲۳۳
	۱۰	۰/۸۸۴۲	۲/۱۴۷۶	۲/۶۶۳۳	۰/۸۰۳۸	۰/۴۶۱۵	۰/۴۷۷۴	۰/۲۷۷۱	۰/۳۷۰۲	۰/۴۱۲۶	۰/۱۵۶۰



شکل (۳-۵) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لوببرگ-مارکوارت و تابع انتقال تانزانت سیگموئیدی

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

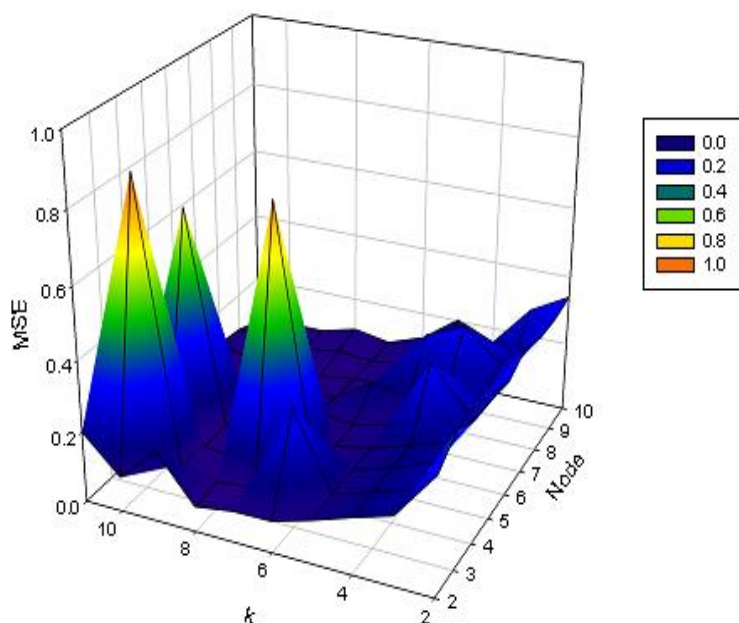
تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	۲	۰/۳۲۳۲	۰/۱۹۴۷	۰/۱۵۵۵	۰/۱۲۱۸	۰/۰۹۰۲	۰/۴۹۴۸	۰/۰۷۱	۰/۰۷۶۶	۰/۱۲۷۲	۰/۲۹۲۲
	۳	۰/۳۲۷۳	۰/۱۹۳۲	۰/۱۵۸۴	۰/۱۲۷۳	۰/۱۵۱۰	۰/۱۰۶۲	۰/۰۷۲۵	۰/۰۸۲۷	۰/۸۵۳۲	۰/۱۱۷۶
	۴	۰/۳۲۸۳	۰/۱۹۳۰	۰/۱۸۰۹	۰/۱۳۸۲	۰/۰۹۵۷	۰/۸۵۱۶	۰/۰۷۳۳	۰/۲۱۱۹	۰/۳۷۲۶	۰/۱۱۱۸
	۵	۰/۳۲۶۹	۰/۱۹۵۰	۰/۱۷۱۰	۰/۱۳۶۹	۰/۴۹۹۸	۰/۱۳۳۴	۰/۱۵۲۳	۰/۰۷۵۸	۰/۲۱۸۶	۰/۱۰۴۰
	۶	۰/۳۲۸۸	۰/۲۸۷۴	۰/۱۶۴۳	۰/۱۲۱۴	۰/۰۹۵۷	۰/۱۰۰۸	۰/۵۷۷۰	۰/۰۷۶۱	۰/۱۷۴۶	۰/۱۷۲۲
	۷	۰/۳۲۹۷	۰/۱۹۵۵	۰/۱۶۶۹	۰/۱۱۷۴	۰/۲۴۰۹	۰/۰۹۸۶	۰/۲۴۷۴	۰/۱۰۸۲	۰/۱۷۱۳	۰/۱۰۶۰
	۸	۰/۳۶۷۱	۰/۱۹۵۲	۰/۱۶۲۸	۰/۱۵۶۶	۰/۰۸۸۸	۰/۴۹۵۵	۰/۰۷۵۵	۰/۰۸۳۲	۰/۲۷۹۵	۰/۰۷۹۷
	۹	۰/۳۵۵۸	۰/۲۲۷۴	۰/۱۵۷۵	۰/۱۳۰۵	۰/۱۲۰۱	۰/۱۲۱۲	۰/۰۸۳۴	۰/۰۸۰۶	۰/۰۹۵۷	۰/۰۹۷۱
	۱۰	۰/۳۲۴۷	۰/۱۹۵۰	۰/۱۸۵۸	۰/۱۲۶۸	۰/۰۹۲۲	۰/۰۹۱۴	۰/۵۵۷۶	۰/۲۶۴۸	۰/۰۷۶	۰/۱۴۷۹



شکل (۳-۶) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

جدول (۳-۱۰) - مقادیر میانگین مربعات خطای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانزانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد متغیرهای ورودی شبکه (تعداد توصیفگرها)											
تعداد گره های لایه پنهان		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
	۲	۰/۳۳۱۸	۰/۱۹۴۵	۰/۱۵۹۷	۰/۱۲۲۱	۰/۰۸۹۶	۰/۰۸۸۰	۰/۰۷۱۸	۰/۲۰۱۰	۰/۰۹۷۷	۰/۱۹۹۸
	۳	۰/۳۸۴۸	۰/۱۹۶۱	۰/۱۵۸۰	۰/۱۲۴۷	۰/۳۳۴۶	۰/۰۸۳۸	۰/۰۷۳۹	۰/۰۹۵۲	۰/۸۷۰۷	۰/۱۳۲۲
	۴	۰/۳۴۵۹	۰/۲۰۲۶	۰/۱۵۷۶	۰/۱۲۳۸	۰/۰۹۹۲	۰/۸۱۰۱	۰/۰۷۴۸	۰/۰۸۱۹	۰/۰۵۸۷	۰/۰۶۳۷
	۵	۰/۳۳۹۶	۰/۱۹۳۹	۰/۱۵۴۷	۰/۱۲۱۲	۰/۰۹۶۷	۰/۰۸۵۷	۰/۰۶۶۴	۰/۰۸۰۶	۰/۶۸۳۷	۰/۰۹۶۸
	۶	۰/۳۲۷۲	۰/۲۱۰۵	۰/۳۳۱۰	۰/۱۲۳۲	۰/۰۹۶۶	۰/۱۱۰۵	۰/۰۹۱۹	۰/۰۸۲۰	۰/۱۲۰۸	۰/۰۹۸۹
	۷	۰/۳۶۰۷	۰/۱۹۳۹	۰/۱۶۹۲	۰/۱۳۹۴	۰/۱۵۹۸	۰/۰۹۳۱	۰/۰۷۱۱	۰/۰۸۷۴	۰/۰۶۹۴	۰/۰۵۹۱
	۸	۰/۳۳۲۰	۰/۲۳۲۹	۰/۳۱۸۶	۰/۲۷۱۲	۰/۰۹۶۱	۰/۰۸۵۶	۰/۰۷۰۶	۰/۰۸۲۸	۰/۰۷۰	۰/۱۰۰۷
	۹	۰/۳۲۹۹	۰/۲۰۰۷	۰/۱۷۸۸	۰/۱۴۳۸	۰/۰۹۵۰	۰/۰۸۳۸	۰/۰۹۰۴	۰/۰۷۴۲	۰/۰۷۳۵	۰/۰۹۰۱
	۱۰	۰/۳۴۰۷	۰/۲۷۹۷	۰/۱۵۷۰	۰/۲۰۳۵	۰/۱۲۳۹	۰/۰۸۵۱	۰/۱۰۳۷	۰/۰۷۵۸	۰/۰۷۵۰	۰/۰۵۹۳



شکل (۳-۷) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانزانت سیگموئیدی

با توجه به نتایج به دست آمده، الگوریتم آموزشی تنظیم بایزین نسبت به الگوریتم آموزشی لونبرگ مارکوارت دارای میانگین مربع خطا (MSE) کمتری است. علاوه بر آن دارای قدرت تعمیم‌پذیری بالاتری است بدین معنی که قادر است میان ترکیباتی که در سری آموزش استفاده شده و آن‌هایی که در سری آموزش استفاده نشده‌اند تطابق خوبی برقرار کند. همچنین تابع تانژانت سیگموئیدی دارای خطای کمتری نسبت به لگاریتم سیگموئیدی است. بنابراین تابع انتقال تانژانت سیگموئیدی به عنوان تابع انتقال لایه پنهان استفاده شد. طبق نتایج به دست آمده ۱۰ توصیفگر به عنوان توصیفگرهایی که بیشترین ارتباط با فعالیت بازدارندگی دارند، انتخاب شدند که جدول (۱۱-۳) لیست این توصیفگرها به همراه طبقه آنها را نشان می‌دهد.

جدول (۱۱-۳) - توصیفگرهای انتخاب شده توسط شبکه عصبی

No	Symbol	Class	Meaning
۱	nPhX	Functional group	Number of X-C on aromatic ring
۲	BEHp8	BCUT	Highest eigenvalue n.8 of Burden matrix
۳	DECC	Topological	Eccentric
۴	RDF105m	RDF	Radial Distribution Function -10.5
۵	Mor 26 p	3D MORSE	3D -MORSE-signal 26
۶	CIC2	Topological	Complementary information content
۷	R5u_A	GETAWAY	R autocorrelation of lag 5
۸	MATS 5p	2D autocorrelation	Moran auto correlation -lag 5
۹	Mor 15p	3D MORSE	3D -MORSE-signal 15
۱۰	Mor 28u	3D MORSE	3D -MORSE-signal 28

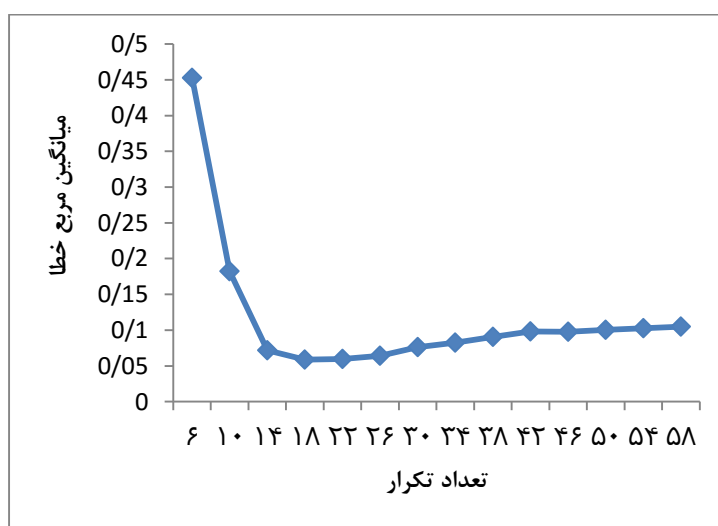
۳-۱-۷-۳- انتخاب تعداد دورهای آموزش

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. به این منظور شبکه با الگوریتم تنظیم بایزین، ۱۰ توصیفگر و تابع

انتقال تانژانت سیگموئیدی در تعداد دوره‌های متفاوت آموزش داده شد و میانگین مربع خطاها برای سری آموزش به روش ارزیابی تقاطعی ثبت گردید جدول (۳-۱۲) و شکل (۳-۸) نشان می‌دهد که با افزایش تعداد دوره‌های آموزش میانگین مربع خطا (MSE) کاهش می‌یابد تا جایی که در ۱۸ دور تکرار به حداقل خطا می‌رسد و سپس با افزایش بیشتر تعداد دوره‌های یادگیری، خطاها افزایش می‌یابد بنابراین ۱۸ بار تکرار به عنوان تعداد تکرارهای بهینه برای یادگیری شبکه انتخاب شد.

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

میانگین مربع خطا	تعداد دوره‌های آموزش	میانگین مربع خطا	تعداد دوره‌های آموزش
۰/۰۸۲۴	۳۴	۰/۴۵۲۷	۶
۰/۰۹۰۷	۳۸	۰/۱۸۲۵	۱۰
۰/۰۹۸۱	۴۲	۰/۰۷۱۷	۱۴
۰/۰۹۷۸	۴۶	۰/۰۵۱۶	۱۸
۰/۱۰۰۶	۵۰	۰/۰۵۹۷	۲۲
۰/۱۰۲۵	۵۴	۰/۰۶۴۱	۲۶
۰/۱۰۴۷	۵۸	۰/۰۷۶۱	۳۰



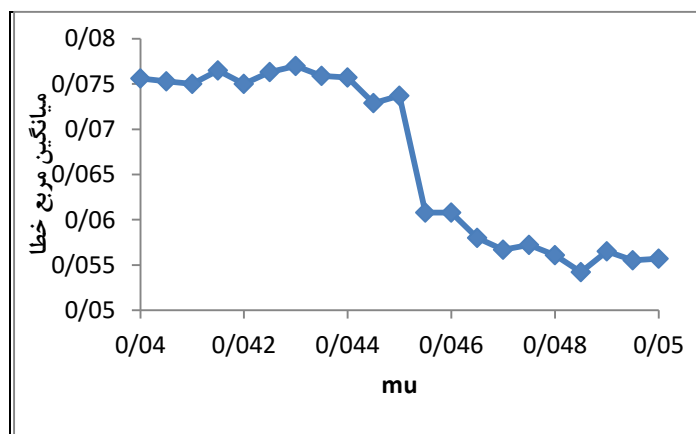
شکل (۳-۸) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش متفاوت

۳-۱-۷-۴- انتخاب مقدار بهینه μ

جهت بهینه کردن مقدار μ ، ساختار شبکه با ۱۰ متغیر ورودی، ۴ گره در لایه پنهان و الگوریتم آموزشی تنظیم بایزین در نظر گرفته شد. سپس مقدار μ از ۰/۰۴ تا ۰/۰۵ با گام‌های ۰/۰۰۰۵ تغییر داده شد و آنگاه برای هر مورد مقدار میانگین مربع خطای سری آموزش به روش ارزیابی تقاطعی محاسبه گردید. در نهایت مقدار میانگین مربع خطا بر حسب μ رسم گردید. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد که طبق جدول (۳-۱۳) و شکل (۳-۹) مقدار بهینه ۰/۰۴۸۵ به دست آمد.

جدول (۳-۱۳) - میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ

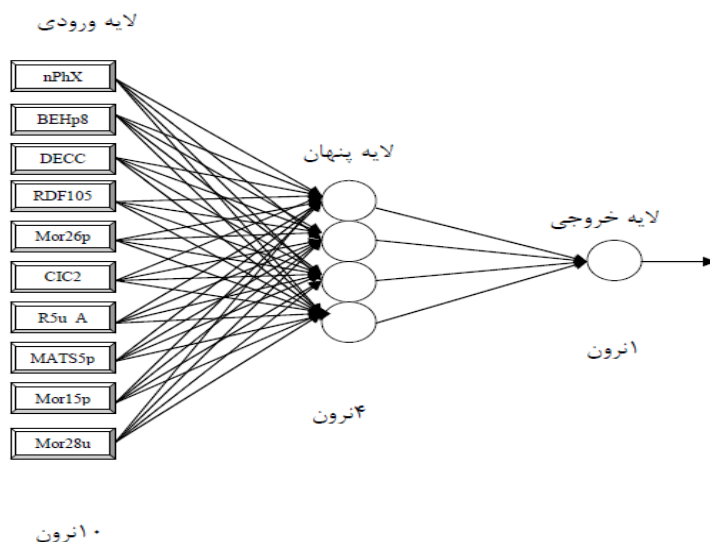
میانگین مربع خطا	مقادیر μ	میانگین مربع خطا	مقادیر μ
۰/۰۶۰۸	۰/۰۴۵۵	۰/۰۷۵۶	۰/۰۴
۰/۰۶۰۸	۰/۰۴۶	۰/۰۷۵۳	۰/۰۴۰۵
۰/۰۵۸	۰/۰۴۶۵	۰/۰۷۵	۰/۰۴۱
۰/۰۵۶۷	۰/۰۴۷	۰/۰۷۶۵	۰/۰۴۱۵
۰/۰۵۷۲	۰/۰۴۷۵	۰/۰۷۵	۰/۰۴۲
۰/۰۵۶۱	۰/۰۴۸	۰/۰۷۶۳	۰/۰۴۲۵
۰/۰۵۴۳	۰/۰۴۸۵	۰/۰۷۷	۰/۰۴۳
۰/۰۵۶۵	۰/۰۴۹	۰/۰۷۵۹	۰/۰۴۳۵
۰/۰۵۵۵	۰/۰۴۹۵	۰/۰۷۵۷	۰/۰۴۴
۰/۰۵۵۷	۰/۰۵	۰/۰۷۲۹	۰/۰۴۴۵
		۰/۰۷۳۷	۰/۰۴۵



شکل (۳-۹) - نمودار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر μ

۳-۱-۷-۵- 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 (۳-۱۰) نشان داده شده است.



شکل (۳-۱۰) - ساختار شبکه عصبی بهینه شده

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۳-۱۴) ارائه شده است.

جدول (۳-۱۴) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

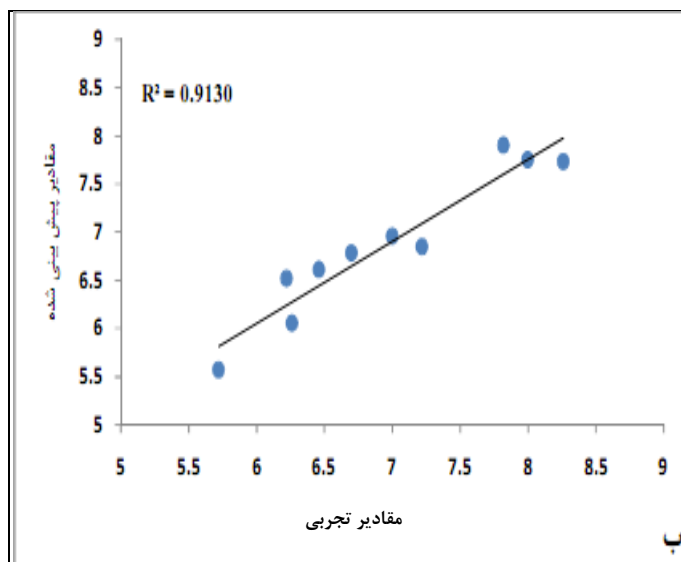
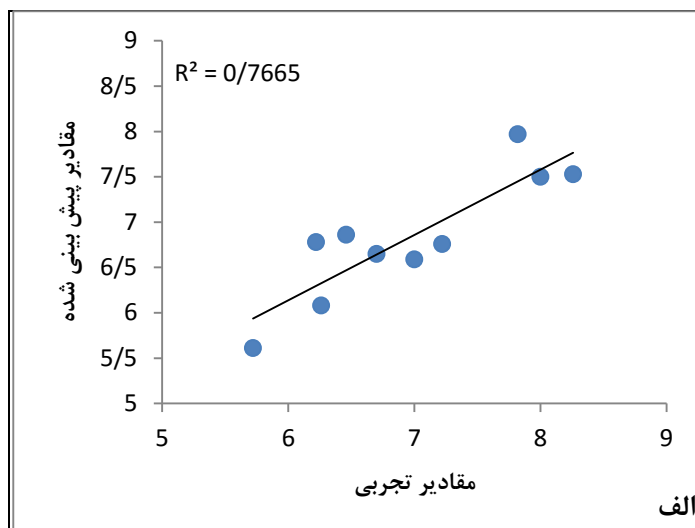
train br	تابع آموزش
tansig	تابع انتقال لایه پنهان
pure line	تابع انتقال لایه خروجی
۴	تعداد نرون لایه پنهان
۱۰	تعداد متغیرهای ورودی
۱۸	تعداد دوره‌های آموزش
۰/۰۴۸۵	پارامتر μ

۳-۱-۸- ارزیابی مدل‌های خطی و غیرخطی

۳-۱-۸-۱- ارزیابی مدل‌های خطی و غیرخطی با استفاده از سری تست

اعتبار و اهمیت معادلات پیشگویی وقتی مشخص می‌گردد که خواص مولکول‌های مجهولی که در دسته‌ی آموزش موجود نیستند را پیشگویی کند. بدین منظور مدل‌های منتخب خطی و غیرخطی جهت پیش‌بینی فعالیت بازدارندگی ۱۰ ترکیب که در فرایند مدل‌سازی استفاده نشده‌اند، به کار گرفته شد. که در مدل خطی با استفاده از ضرائب مدل برتر و در مدل غیرخطی با استفاده از شبکه عصبی بهینه شده، مقادیر pIC_{50} این ترکیبات پیش‌بینی شد. نتایج مربوط به پیش‌بینی فعالیت بازدارندگی این ترکیبات در جدول (۳-۱۵) نشان داده شده است. همچنین شکل (۳-۱۱)، ضریب تعیین مقدار پیش‌بینی شده بر حسب مقدار تجربی برای سری تست را با استفاده از هر دو مدل خطی و غیرخطی نشان می‌دهد. ضریب تعیین بیانگر این است که روش غیرخطی در پیش‌بینی فعالیت دارویی ترکیبات مورد بررسی، توانمند است.

جدول (۳-۱۵) - نتایج حاصل از ارزیابی مدل‌های MLR و ANN با استفاده از سری تست					
شماره ترکیب	pIC_{50}				
	مقدار تجربی	مقدار پیش‌بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۲	۷/۲۲	۶/۷۶	۶/۸۵	-۶/۳۷	-۵/۱۲
۷	۶/۴۶	۶/۸۶	۶/۶۱	۶/۱۹	۲/۳۲
۱۲	۶/۲۲	۶/۷۸	۶/۵۲	۹/۰۰	۴/۸۲
۱۸	۶/۷۰	۶/۶۵	۶/۷۸	-۰/۷۵	۱/۱۹
۲۲	۸/۰۰	۷/۵۰	۷/۷۵	-۶/۲۵	-۳/۱۲
۲۷	۷/۰۰	۶/۵۹	۶/۹۶	-۵/۸۶	-۰/۵۷
۳۳	۶/۲۶	۶/۰۸	۶/۰۵	-۲/۸۸	-۳/۳۵
۳۶	۵/۷۲	۵/۶۱	۵/۵۷	-۱/۹۲	-۲/۶۲
۴۲	۸/۲۶	۷/۵۳	۷/۷۳	-۸/۸۴	-۶/۴۲
۴۹	۷/۸۲	۷/۹۷	۷/۹۰	۱/۹۲	۱/۰۲۲



شکل (۳-۱۱) - نمودار مقادیر پیش بینی شده pIC_{50} بر حسب مقادیر تجربی برای سری تست الف: با مدل MLR ب: با مدل ANN

۳-۱-۸-۲- ارزیابی مدل های خطی و غیرخطی توسط روش رد مرحله ای تک تک و رد مرحله ای گروهی^۱

به منظور بررسی بیشتر قدرت پیش گویی مدل های خطی و غیرخطی، علاوه بر تکنیک رد مرحله ای تک تک، روش رد مرحله ای گروهی نیز مورد استفاده قرار گرفت. در روش رد مرحله ای گروهی هر بار دسته ای از ترکیبات به طور کاملاً تصادفی از سری داده ها حذف شدند (در اینجا هر بار ۲۵٪ ترکیبات حذف شدند) و سپس با استفاده از مدل ساخته شده توسط بقیه ترکیبات، فعالیت دارویی ترکیبات

^۱ - Leave-group-out

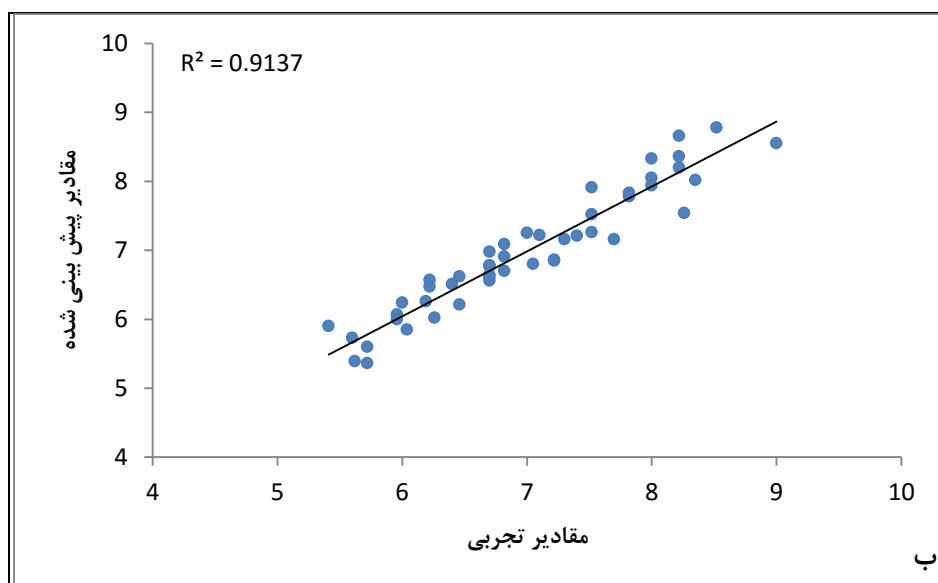
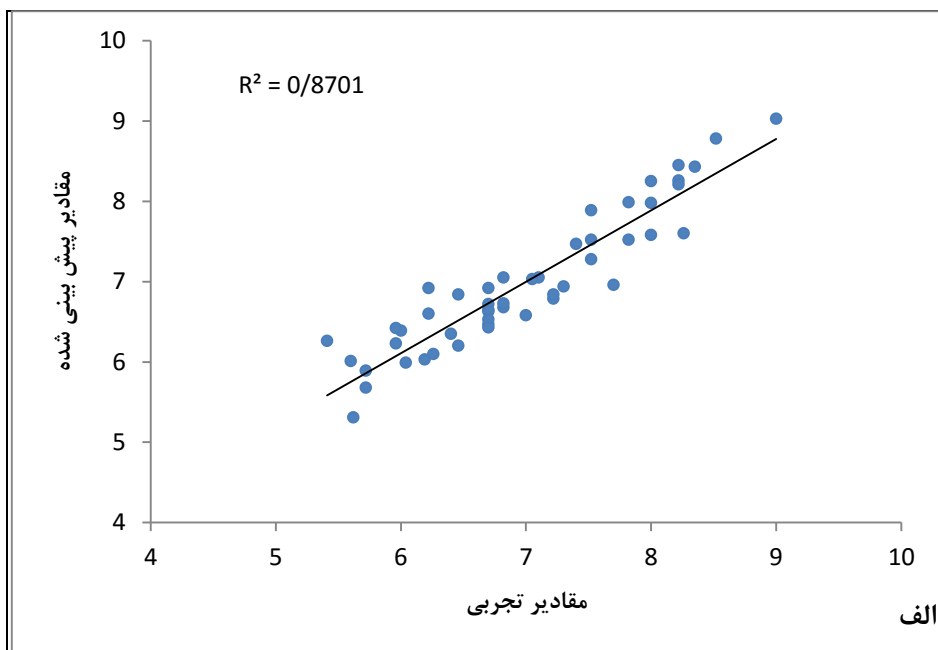
حذف شده، پیش‌بینی شد. این فرایند برای تمام اعضای سری داده‌ها تکرار شد. نتایج حاصل از رد مرحله‌ای تک‌تک و رد مرحله‌ای گروهی به ترتیب در جداول (۳-۱۶) و (۳-۱۷) ارائه شده است. این نتایج تعمیم‌پذیری بالای مدل خطی و غیرخطی طراحی شده را بیان می‌کند. ضریب تعیین مشاهده شده بین مقادیر پیش‌بینی شده pIC_{50} و مقادیر تجربی در شکل‌های (۳-۱۲) و (۳-۱۳) نشان دهنده میزان نزدیکی فعالیت پیش‌بینی شده توسط مدل‌ها نسبت به مقدار تجربی آن است.

جدول (۳-۱۶) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها

شماره ترکیب	مقدار تجربی	pIC_{50}			
		مقدار پیش‌بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۱	۶/۰۴	۵/۹۹	۵/۸۵	-۰/۸۳	-۳/۱۵
۲	۷/۲۲	۶/۷۹	۶/۸۵	-۵/۹۶	-۵/۱۲
۳	۶/۸۲	۶/۶۸	۶/۹۱	-۲/۰۵	۱/۳۲
۴	۶/۰۰	۶/۳۹	۶/۲۴	۶/۵۰	۴/۰۰
۵	۷/۰۴	۷/۴۷	۷/۲۱	۰/۹۵	-۲/۵۷
۶	۶/۸۲	۷/۰۵	۷/۰۹	۳/۳۷	۳/۹۶
۷	۶/۴۶	۶/۸۴	۶/۶۲	۵/۸۸	۲/۴۸
۸	۷/۱۰	۷/۰۵	۷/۲۲	-۰/۷۰	۱/۶۹
۹	۷/۵۲	۷/۵۲	۷/۲۶	۰/۰۰	-۳/۴۶
۱۰	۶/۸۲	۶/۷۳	۶/۷۰	-۱/۳۲	-۱/۷۶
۱۱	۵/۹۶	۶/۴۲	۶/۰۷	۷/۷۲	۱/۸۵
۱۲	۶/۲۲	۶/۹۲	۶/۵۷	۱۱/۲۵	۵/۶۳
۱۳	۷/۲۲	۶/۸۴	۶/۸۶	-۵/۲۶	-۴/۹۹
۱۴	۶/۷۰	۶/۶۵	۶/۶۴	-۰/۷۵	-۰/۹۰
۱۵	۶/۴۰	۶/۳۵	۶/۵۱	-۰/۷۸	۱/۷۲
۱۶	۵/۹۶	۶/۲۳	۶/۰۰	۴/۵۳	۰/۶۷
۱۷	۷/۰۵	۷/۰۳	۶/۸۰	-۰/۲۸	-۳/۵۵
۱۸	۶/۷۰	۶/۷۲	۶/۹۸	۰/۳۰	۴/۱۸
۱۹	۷/۵۲	۷/۸۹	۷/۵۲	۴/۹۲	۰/۰۰

ادامه جدول (۳-۱۶) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها

شماره ترکیب	مقدار تجربی	pIC ₅₀			
		مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۲۰	۸/۰۰	۷/۹۸	۸/۰۵	-۰/۲۵	۰/۶۳
۲۱	۸/۰۰	۸/۲۵	۸/۳۳	۳/۱۳	۴/۱۳
۲۲	۸/۰۰	۷/۵۸	۷/۹۴	-۵/۲۵	-۰/۷۵
۲۳	۸/۲۲	۸/۲۶	۸/۲۰	۰/۴۹	-۰/۲۴
۲۴	۶/۷۰	۶/۴۷	۶/۵۶	-۳/۴۳	-۲/۰۹
۲۵	۵/۷۲	۵/۸۹	۵/۶۰	۲/۹۷	-۲/۱۰
۲۶	۶/۱۹	۶/۰۳	۶/۲۶	-۲/۵۸	۱/۱۳
۲۷	۷/۰۰	۶/۵۸	۷/۲۵	-۶/۰۰	۳/۵۷
۲۸	۶/۲۲	۶/۶۰	۶/۴۷	۶/۱۱	۴/۰۲
۲۹	۶/۴۶	۶/۲۰	۶/۲۱	-۴/۰۲	-۳/۸۷
۳۰	۶/۷۰	۶/۹۲	۶/۷۸	۳/۲۸	۱/۱۹
۳۱	۵/۶۰	۶/۰۱	۵/۷۳	۷/۳۲	۲/۳۲
۳۲	۶/۷۰	۶/۶۳	۶/۷۷	-۱/۰۴	۱/۰۴
۳۳	۶/۲۶	۶/۱۰	۶/۰۲	-۲/۵۶	-۳/۸۳
۳۴	۶/۷۰	۶/۵۳	۶/۷۸	-۲/۵۴	۱/۱۹
۳۵	۵/۶۲	۵/۳۱	۵/۳۹	-۵/۵۲	-۴/۰۹
۳۶	۵/۷۲	۵/۶۸	۵/۳۶	-۰/۷۰	-۶/۲۹
۳۷	۵/۴۱	۶/۲۶	۵/۹۰	۱۵/۷۱	۹/۰۶
۳۸	۷/۷۰	۶/۹۶	۷/۱۶	-۹/۶۱	-۷/۰۱
۳۹	۷/۸۲	۷/۵۲	۷/۸۳	-۳/۸۴	۰/۱۳
۴۰	۷/۵۲	۷/۲۸	۷/۹۱	-۳/۱۹	۵/۱۹
۴۱	۶/۷۰	۶/۴۳	۶/۶۱	-۴/۰۳	-۱/۳۴
۴۲	۸/۲۶	۷/۶۰	۷/۵۴	-۷/۹۹	-۸/۷۲
۴۳	۸/۲۲	۸/۲۱	۸/۶۶	-۰/۱۲	۵/۳۵
۴۴	۹/۰۰	۹/۰۳	۸/۵۵	۰/۳۳	-۵/۰۰
۴۵	۸/۵۲	۸/۷۸	۸/۷۸	۳/۰۵	۳/۰۵
۴۶	۸/۳۵	۸/۴۳	۸/۰۲	۰/۹۶	-۳/۹۵
۴۷	۸/۲۲	۸/۴۵	۸/۳۶	۲/۸۰	۱/۷۰
۴۸	۷/۳۰	۶/۹۴	۷/۱۶	-۴/۹۳	-۱/۹۲
۴۹	۷/۸۲	۷/۹۹	۷/۷۸	۲/۱۷	-۰/۵۱



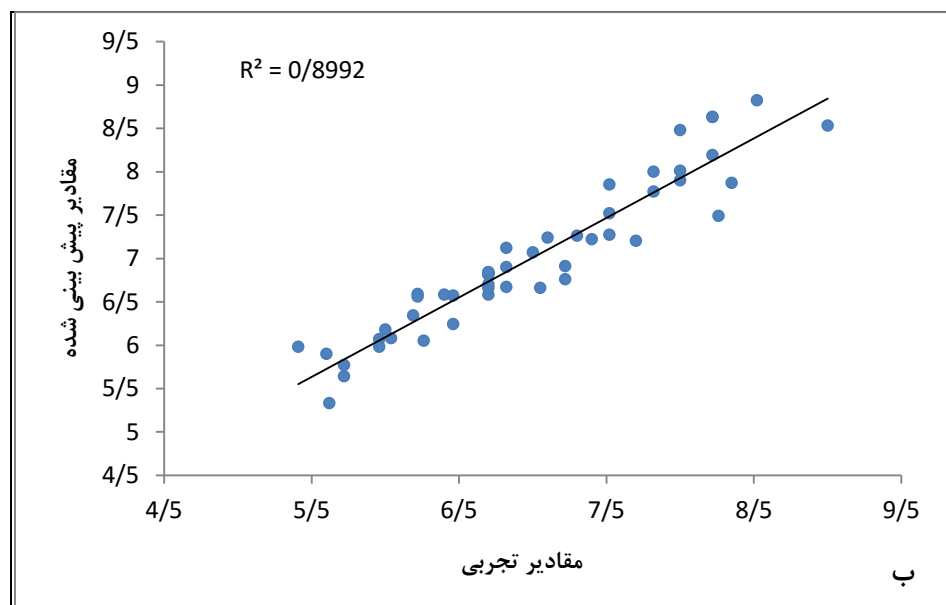
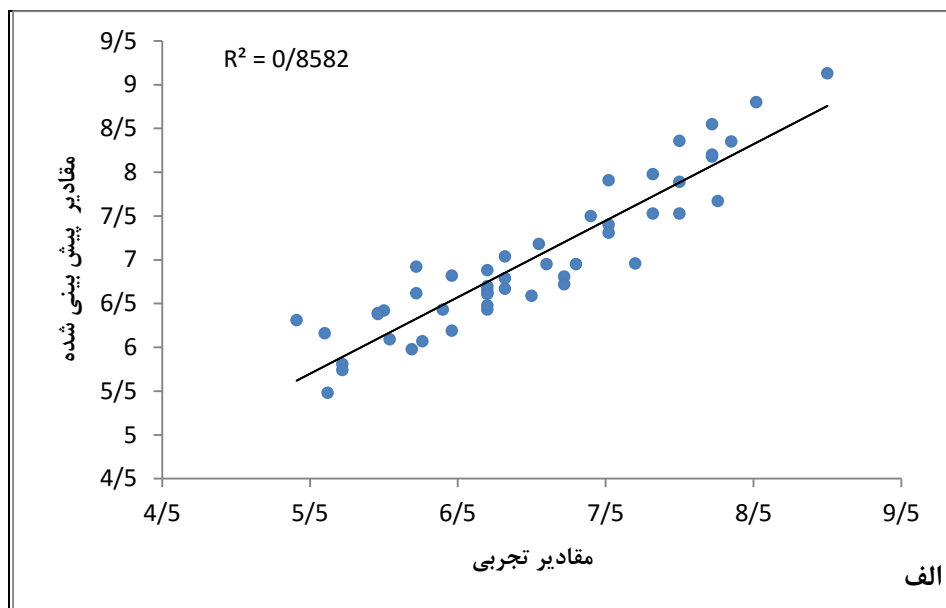
شکل (۳-۱۲) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی pIC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها
الف: با مدل MLR **ب:** با مدل ANN

جدول (۳- ۱۷) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از روش رد مرحله‌ای گروهی برای کل داده‌ها

شماره ترکیب	pIC ₅₀				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۳	۶/۸۲	۶/۷۹	۶/۹۰	-۰/۴۴	۱/۱۷
۷	۶/۴۶	۶/۸۲	۶/۵۷	۵/۵۷	۱/۷۰
۱۲	۶/۲۲	۶/۹۲	۶/۵۹	۱۱/۲۵	۵/۹۵
۱۵	۶/۴۰	۶/۴۳	۶/۵۸	۰/۴۷	۲/۸۱
۱۹	۷/۵۲	۷/۹۱	۷/۵۲	۵/۱۹	۰/۰۰
۲۱	۸/۰۰	۸/۳۶	۸/۴۸	۴/۵۰	۶/۰۰
۲۷	۷/۰۰	۶/۵۹	۷/۰۷	-۵/۸۶	۱/۰۰
۳۰	۶/۷۰	۶/۸۸	۶/۵۸	۲/۶۹	-۱/۷۹
۳۳	۶/۲۶	۶/۰۷	۶/۰۵	-۳/۰۴	-۳/۳۵
۳۹	۷/۸۲	۷/۵۳	۷/۷۷	-۳/۷۱	-۰/۶۴
۴۱	۶/۷۰	۶/۴۳	۶/۶۶	-۴/۰۳	-/۶۰
۴۷	۸/۲۲	۸/۵۵	۸/۶۳	۴/۰۱	۴/۹۹
۲	۷/۲۲	۶/۸۱	۶/۹۱	-۵/۶۸	-۴/۲۹
۶	۶/۸۲	۷/۰۴	۷/۱۲	۳/۲۳	۴/۴۰
۱۱	۵/۹۶	۶/۳۹	۶/۰۷	۷/۲۱	۱/۸۵
۱۴	۶/۷۰	۶/۷۰	۶/۸۱	۰/۰۰	۱/۶۴
۱۸	۶/۷۰	۶/۶۶	۶/۸۴	-۰/۶۰	۲/۰۹
۲۲	۸/۰۰	۷/۵۳	۷/۹۰	-۵/۸۸	-۱/۲۵
۲۵	۵/۷۲	۵/۸۱	۵/۷۷	۱/۵۷	۰/۸۷
۲۹	۶/۴۶	۶/۱۹	۶/۲۴	-۴/۱۸	-۳/۴۱
۳۴	۶/۷۰	۶/۶۲	۶/۸۲	-۱/۱۹	۱/۷۹
۳۸	۷/۷۰	۶/۹۶	۷/۲۰	-۹/۶۱	-۶/۴۹
۴۳	۸/۲۲	۸/۱۸	۸/۶۳	-۰/۴۹	۴/۹۹
۴۶	۸/۳۵	۸/۳۵	۷/۸۷	۰/۰۰	-۵/۷۵
۱	۶/۰۴	۶/۰۹	۶/۰۸	۰/۸۳	۰/۶۶
۵	۷/۴۰	۷/۵۰	۷/۲۲	۱/۳۵	-۲/۴۳
۹	۷/۵۲	۷/۴۰	۷/۲۷	-۱/۶۰	-۳/۳۲
۱۳	۷/۲۲	۶/۷۲	۶/۷۶	-۶/۹۳	-۶/۳۷

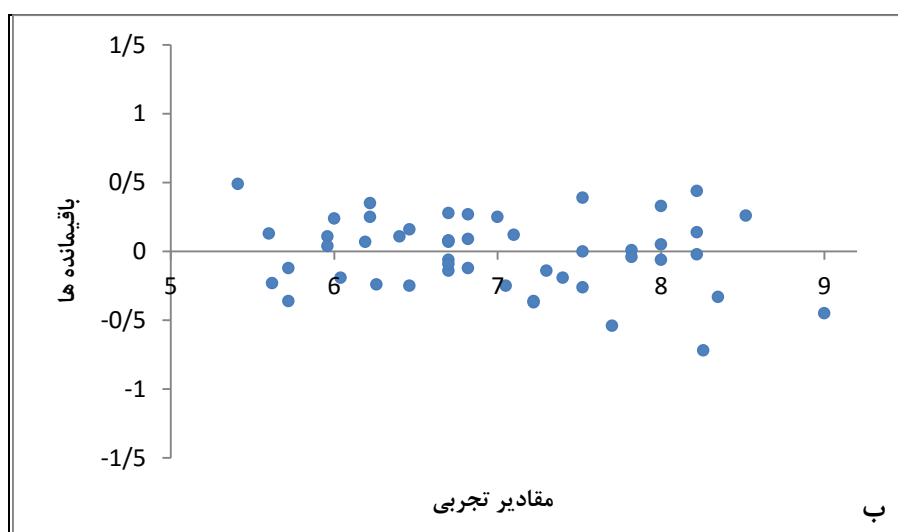
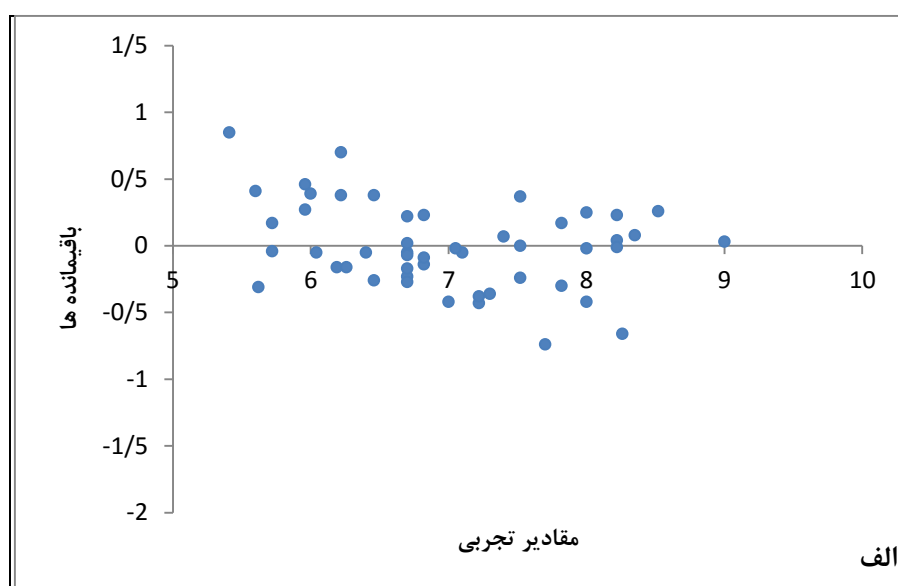
ادامه جدول (۳-۱۷) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از رد مرحله‌ای گروهی برای کل داده‌ها

شماره ترکیب	pIC ₅₀				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۱۷	۷/۰۵	۷/۱۸	۶/۶۶	۱/۸۴	-۵/۵۳
۲۴	۶/۷۰	۶/۴۸	۶/۸۴	-۳/۲۸	۲/۰۹
۲۶	۶/۱۹	۵/۹۸	۶/۳۴	-۳/۳۹	۲/۴۲
۳۲	۶/۷۰	۶/۶۱	۶/۷۰	-۱/۳۴	۰/۰۰
۳۶	۵/۷۲	۵/۷۴	۵/۶۴	۰/۳۵۰	-۱/۴۰
۳۷	۵/۴۱	۶/۳۱	۵/۹۸	۱۶/۶۴	-۱۰/۵۴
۴۲	۸/۲۶	۷/۶۷	۷/۴۹	-۷/۱۴	-۹/۳۲
۴۵	۸/۵۲	۸/۸۰	۸/۸۲	۳/۲۹	۳/۵۲
۴۹	۷/۸۲	۷/۹۸	۸/۰۰	۲/۰۵	۲/۳۰
۴	۶/۰۰	۶/۴۲	۶/۱۸	۷/۰۰	۳/۰۰
۸	۷/۱۰	۶/۹۵	۷/۲۴	-۲/۱۱	۱/۹۷
۱۰	۶/۸۲	۶/۶۷	۶/۶۷	-۲/۲۰	-۲/۲۰
۱۶	۵/۹۶	۶/۳۸	۵/۹۸	۷/۰۵	۰/۳۴
۲۰	۸/۰۰	۷/۸۹	۸/۰۱	-۱/۳۸	۰/۱۲
۲۳	۸/۲۲	۸/۲۰	۸/۱۹	-۰/۲۴	-۰/۳۶
۲۸	۶/۲۲	۶/۶۲	۶/۵۶	۶/۳۴	۵/۴۷
۳۱	۵/۶۰	۶/۱۶	۵/۹۰	۱۰/۰۰	۵/۳۶
۳۵	۵/۶۲	۵/۴۸	۵/۳۳	-۲/۴۹	-۵/۱۶
۴۰	۷/۵۲	۷/۳۱	۷/۸۵	-۲/۷۹	۴/۳۹
۴۴	۹/۰۰	۹/۱۳	۸/۵۳	۱/۴۴	-۵/۲۲
۴۸	۷/۳۰	۶/۹۵	۷/۲۶	-۴/۷۹	-۰/۵۵



شکل (۳-۱۳) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی pIC₅₀ به روش رد مرحله‌ای گروهی برای کل داده‌ها
 الف: با مدل MLR ب: با مدل ANN

نمودار باقیمانده‌ها معیاری برای شایستگی مدل به دست آمده، می‌باشد و اگر باقیمانده‌ها به طور یکنواخت حول محور افقی پراکنده باشند، نشان می‌دهد که مدل مناسبی به دست آمده و هیچ خطای سیستماتیکی وجود ندارد. شکل (۳-۱۴) نمودار باقیمانده‌ها را که از تفاضل مقدار تجربی فعالیت دارویی از مقدار محاسبه شده توسط مدل‌های خطی و غیرخطی به دست می‌آید، بر حسب مقدار تجربی نشان می‌دهد. تقارن پراکندگی نقاط در دو طرف محور افقی عدم وجود خطای سیستماتیک را نشان می‌دهد.



شکل (۳ - ۱۴) - نمودار باقیمانده‌ها بر حسب مقدار تجربی pIC_{50} برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN

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

مطابق جدول (۳-۱۸) هفت پارامتر آماری، جهت ارزیابی توانایی پیشگویی مدل‌های ساخته شده به روش خطی (MLR) و غیرخطی (ANN) به کار گرفته شد. نتایج رضایت بخش مدل غیرخطی نشان می‌دهد که مشارکت بعضی توصیفگرها برای بازدارندگی آنزیم نسخه‌بردار معکوس ممکن است، غیرخطی باشد.

جدول (۳-۱۸) - پارامترهای آماری برای مدل‌های برتر طراحی شده توسط MLR و ANN

پارامتر	سری تست (N=10)		کل داده‌ها (N=49)	
	MLR	ANN	MLR	ANN
MAE	۰/۳۵۵	۰/۲۱۳	۰/۲۴۶	۰/۲۱۱
MSE	۰/۱۷۱	۰/۰۶۷۴	۰/۱۰۱۲	۰/۰۶۸
PRESS	۱/۷۱	۰/۰۶۷۴	۴/۹۵۹	۳/۳۳۲
SEP	۰/۴۱۳	۰/۲۶۰	۰/۳۱۸	۰/۲۶۱
R ²	۰/۷۶۶	۰/۹۱۳۰	۰/۸۷۰	۰/۹۱۳۷
REP (%)	۵/۹۳۶	۳/۷۳	۴/۵۴۸	۳/۷۳
MRE	۴/۹۹۷	۳/۰۵۷	۳/۶۵۹	۳/۰۲۹
R ² (Leave Group Out)			۰/۸۵۸	۰/۸۹۹

۳-۱-۱۰- ارزیابی مدل ارائه شده توسط شبکه عصبی با استفاده از آزمون Y- تصادفی^۱

این تکنیک ارزیابی مدل با هدف بررسی هر گونه ارتباط تصادفی بین داده‌ها انجام شد. در این آزمون مقادیر تصادفی از متغیر وابسته تولید گردید. مدل QSAR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر وابسته توسعه یافت. اگر مدل اصلی هیچ گونه ارتباط تصادفی نداشته باشد، تفاوت قابل توجهی بین مقدار ضریب تعیین مدل اصلی و مدل QSAR که با پاسخ‌های تصادفی توسعه یافته، وجود دارد. نتایج حاصل از چندین بار اجرای آزمون Y- تصادفی در جدول (۳-۱۹) نشان داده شده است. مقادیر کوچک ضریب تعیین (R^2) بیانگر عدم وجود ارتباط شانس یا وابستگی ساختاری به سری آموزش در مدل توسعه یافته توسط شبکه می‌باشد.

^۱ - Y- randomization test

جدول (۳-۱۹) - مقادیر R^2 برای سری تست با استفاده از آزمون Y- تصادفی

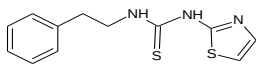
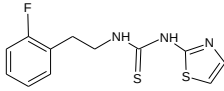
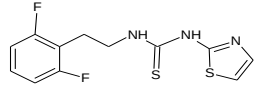
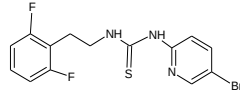
تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2	۰/۰۰۱۵	۰/۰۶۸۲	۰/۰۰۸۹	۰/۱۵۳۶	۰/۰۳۶۵	۰/۱۰۰۵	۰/۰۶۶۷	۰/۰۰۲۲	۰/۰۲۳۴	۰/۱۳۳۹

۳-۱-۱۱- بررسی ارتباط توصیفگرهای منتخب با فعالیت ضد HIV

الف- توصیفگرهای Functional group

اولین توصیفگری که در مدل ظاهر شده است، توصیفگر nPhX^۱ از دسته توصیفگرهای گروه عاملی^۲ است. این توصیفگر تعداد اتمهای هالوژن پیوند داده شده به اتمهای کربن موجود در حلقه آروماتیک را نشان می‌دهد [۶۹ و ۷۰] که دارای ضریب مثبت در مدل برتر است و نشان می‌دهد که افزایش تعداد استخلافهای هالوژن در حلقه آروماتیک منجر به افزایش pIC₅₀ می‌شود. به عبارت دیگر بازداری آنزیم نسخه‌بردار معکوس در فرایند رونویسی RNA، توسط مولکول‌هایی با استخلاف هالوژن زیاد می‌شود. چند مثال از اثر این توصیفگر در جدول (۳-۲۰) ارائه شده است.

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

ترکیب	nphx	pIC ₅₀
	۰	۶/۰۴
	۱	۷/۲۲
	۲	۸/۰۰
	۳	۹/۰۰

^۱ - Number of X-C on aromatic ring

^۲ Functional group

ب- توصیفگرهای BCUT^۱

این توصیفگرها از دسته توصیفگرهای دو بعدی هستند و به عنوان مقدار ویژه یک ماتریس ارتباطی اصلاح شده^۲ (ماتریس بوردن) می‌باشند. ماتریس بوردن یک گراف تهی از هیدروژن را ارائه می‌دهد که در آن B_{ii} ، یعنی عناصر قطری در ارتباط با خاصیت اتمی مانند الکترونگاتیویته، حجم و اندروالس، جرم اتمی و عناصر غیر قطری در ارتباط با مرتبه پیوند دو اتم پیوند داده شده، هستند [۷۱ و ۷۲]. $BEHp8$ بالاترین مقدار ویژه از ماتریس بوردن که با قطبش‌پذیری وزن دار شده است، بزرگترین ضریب را در مدل به خود اختصاص داده است. علامت مثبت این توصیفگر در معادله نشان می‌دهد که مقادیر بالاتر از این توصیفگر، باعث افزایش برهمکنش آنزیم – دارو و در نتیجه کاهش سرعت تکثیر ویروس HIV می‌شود.

پ- توصیفگرهای Topological

توصیفگرهای توپولوژیکی که از گراف مولکولی به دست می‌آیند، به عنوان توصیفگرهای عمومی مطرح هستند، چون به طور مستقیم از ساختار مولکول محاسبه می‌شوند و مستقل از صورتبندی مولکول هستند. $DECC$ ^۳ و $CIC2$ ^۴ از جمله این توصیفگرها می‌باشند که در مدل برتر ظاهر شده‌اند. این توصیفگرها بیانی از اندازه و شاخه‌ای بودن ترکیب هستند [۷۳ و ۷۴] و با افزایش مقدار هر دو این توصیفگرها، مقدار pIC_{50} کاهش می‌یابد زیرا ممانعت‌های فضایی ایجاد شده در اثر شاخه‌دار شدن ترکیبات، باعث کاهش برهمکنش آنزیم نسخه‌بردار- بازدارنده (دارو) می‌شود و به موجب آن اثر دارو در جلوگیری از فرایند نسخه‌برداری و تکثیر ویروس HIV کمتر می‌گردد.

^۱ - Burden-CAS- university of texas eigenvalues

^۲ - Modified connectivity matrix

^۳ - Highest eigenvalue n.8 of Burden matrix weighted by polarizabilities

^۴ - Eccentric

^۵ - Complementary information content

ت-توصیفگرهای گروه RDF^۱

می‌توان تابع توزیع شعاعی یک ترکیب شامل N اتم را به احتمال یافتن یک اتم در یک حجم کرووی با شعاع r تعریف کرد و فرم کلی این تابع به صورت زیر است:

$$g(r) = f \sum_{i=1}^{N-1} \sum_{j>i}^N A_i A_j e^{\beta(r-r_{ij})^2} \quad (۷-۳)$$

که در این رابطه N تعداد اتم‌ها، A_i و A_j خاصیت اتمی (الکترونگاتیویته، جرم اتمی، قطبش پذیری اتمی و...)، r_{ij} شعاع بین اتم i و j، f فاکتور مقیاس، β فاکتور دما که باعث حرکت و جابجایی اتم‌ها است، می‌باشند [۷۵]. از میان این توصیفگرها RDF105m^۲ که با جرم اتمی وزن دار شده در مدل برتر دیده می‌شود. با توجه به رابطه فوق هر چه جرم اتمی بیشتر شود، مقدار این توصیفگر افزایش می‌یابد. علامت منفی مربوط به ضریب این توصیفگر در مدل برتر گویای این مطلب است که استخلافهای الکترون دهنده و الکترون کشنده کوچک دارای فعالیت دارویی مناسبتری هستند.

ث-توصیفگرهای 3D - MORSE^۳

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

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad (۸-۳)$$

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

¹ - Radial Distribution Function

² - Radial Distribution Function weighted by atomic mass

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

برخی از این توصیفگرهای سه بعدی مانند ¹Mor28u ، ²Mor15p و ³Mor26p در مدل برتر دیده می‌شوند. چون این توصیفگرها آرایش سه بعدی اتم‌ها را بدون اینکه به اندازه مولکول ارتباط داشته باشند، بیان می‌کنند بنابراین برای تعداد زیادی مولکول با تفاوت‌های ساختاری زیاد قابل کاربرد است [۷۶].

ج-توصیفگرهای GETAWAY^۴

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

$$H = M (M^T M) \cdot M^T \quad (۹-۳)$$

که M ماتریس مختصات اتمی، بالانویس T به معنای ماتریس ترانپوز^۶ و H ماتریس تأثیر مولکولی است. عناصر قطری (h_{ii}) ماتریس H لوریج^۷ نام دارند که هر یک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. به عنوان مثال اتم‌های دورتر نسبت به اتم‌های نزدیکتر به مرکز اتم، دارای h_{ii} بزرگتر هستند.

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

¹ - 3D –MORSE-signal 28 un weighted

² - 3D –MORSE-signal 15 weighted by polarizability

³ - 3D –MORSE-signal 26 weighted by polarizability

⁴ - Geometrical, Topology and atom Weighted assembly

⁵ - Molecular Influence Matrix

⁶ - Transpose matrix

⁷ - Leverage

$$[R]_{ij} = \left[\frac{\sqrt{h_{ii} h_{jj}}}{r_{ij}} \right]_{ij} \quad i \neq j \quad (10-3)$$

که h_{ii} و h_{jj} لوریج (قدرت نفوذ) دو اتم i و j و r_{ij} فاصله آن دو اتم است [۷۷]. $R5u^+$ (از دسته توصیفگرهای R- GETAWAY) از لوریج‌های دو اتم که در فاصله توپولوژیکی ۵ قرار دارند و با ویژگی خاصی از مولکول وزن دار نشده، محاسبه می‌شود. این توصیفگرها از آن جهت مهم هستند که به ساختار سه بعدی مولکول حساس هستند و معمولاً برای مقایسه مولکول‌ها و حتی صورتبندی‌ها از نظر شکل، تقارن، اندازه اتم به کار می‌روند. این توصیفگر دارای علامت منفی در مدل است و نشان می‌دهد با افزایش این توصیفگر مقدار pIC_{50} کاهش پیدا می‌کند.

چ-توصیفگرهای Moran Autocorrelation

توصیفگرهای گروه Moran Autocorrelation توصیفگرهای دوبعدی هستند که از طریق رابطه زیر محاسبه می‌شوند:

$$I(d) = \frac{\frac{1}{2} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - \bar{w})(w_j - \bar{w})}{\frac{1}{A} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad (11-3)$$

که w یک ویژگی اتم، $\bar{w}$ میانگین مقدار آن روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است و δ_{ij} نیز - که به تابع کرونکر^۲ معروف است - در حالتی که $d_{ij} = d$ باشد یک است، در غیر اینصورت صفر است. مقدار این توصیفگر که از جنس فاصله است، از $+1$ تا -1 متغیر است [۷۲ و ۷۸]. $MATS5p$ ^۳ از این گروه توصیفگرهاست که در مدل برتر ظاهر شده است و نقش قطبش‌پذیری اتمی ترکیبات را در جهت‌گیری به سوی مولکول هدف نشان می‌دهد. علامت مثبت این توصیفگر بیانگر این است که مقادیر بالاتر از این توصیفگر برای بازدارندگی آنزیم نسخه‌بردار معکوس مطلوب است.

^۱ - R autocorrelation of lag 5 un weighted

^۲ - Kronecker

^۳ - Moran autocorrelation-lag15 weighted by atomic polarizabilities

۳-۱-۱۲- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی

میزان مشارکت توصیفگرهای منتخب به صورت زیر تعیین شد:

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

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

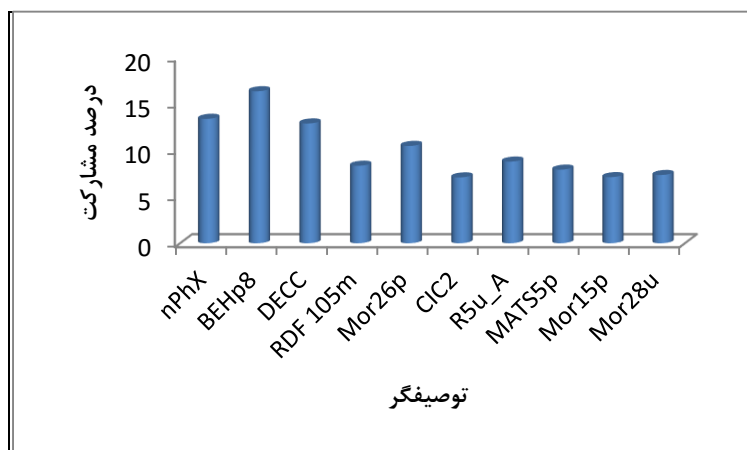
۳- میانگین خطای مطلق حاصل از ارزیابی تقاطعی ترکیبات سری آموزش محاسبه گردید.

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

۵- سرانجام درصد مشارکت هر توصیفگر توسط رابطه (۳-۱۲) برآورد شد [۷۹-۸۱].

$$c_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad (۳-۱۲)$$

در این رابطه c_i درصد مشارکت توصیفگر حذف شده Δm_i ، N تعداد توصیفگرهای مدل و Δm_i میانگین خطای مطلق حاصل از ارزیابی تقاطعی سری آموزش در غیاب توصیفگر Δm_i را نشان می‌دهد. که بر این اساس درصد مشارکت توصیفگرهای منتخب در ترکیبات مورد بررسی به شکل زیر می‌باشد.



شکل (۳-۱۵)- مشارکت توصیفگرها در شبکه عصبی بهینه

بر اساس نمودار فوق توصیفگرهای nPhX و DECC، BEHp8 دارای بیشترین اثر مشارکت می‌باشند.

توصیفگرهای DECC و BEHp8 در برگیرنده اطلاعات دو بعدی مولکول هستند بنابراین می‌توان

گفت که فعالیت بازداری این ترکیبات وابستگی قابل ملاحظه‌ای به توپولوژی ترکیب بازدارنده دارد. همچنین توصیفگر nPhX از دسته توصیفگرهای یک بعدی است که نقش اتم‌های هالوژن در میزان بازدارندگی آنزیم نسخه‌بردار معکوس را نشان می‌دهد.

۳-۱-۱۳- نتیجه‌گیری نهایی

در این تحقیق برای اولین بار مدل‌های خطی و غیرخطی برای پیش‌بینی فعالیت ضد HIV یکسری از مشتقات تیواوره به کار گرفته شد. ارزیابی مدل بویژه با سری تست نشان دهنده صحت و اعتبار مدل‌های پیشنهادی است بنابراین بر اساس این مدل‌ها می‌توان فعالیت ضد HIV را برای همین طبقه از ترکیبات که هنوز سنتز نشده‌اند پیش‌بینی کرد. و در واقع این روش به طراحان دارو این امکان را می‌دهد که قبل از اینکه بخواهند یک ترکیب دارویی را سنتز کنند، ابتدا با استفاده از مدل معتبر خاصیت دارویی آن را پیشگویی 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.

فصل چهارم

مطالعه ارتباط کمی ساختار-فعالیت مشتقات

تیوکربامات با استفاده از رگرسیون خطی چندگانه

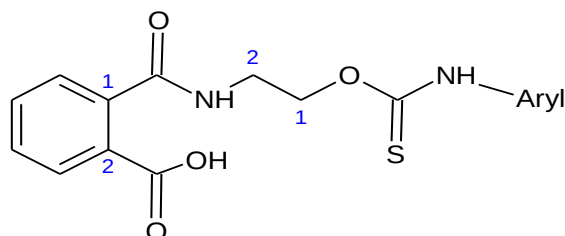
و شبکه عصبی مصنوعی

۴-۱- محاسبات و مدل سازی فعالیت دارویی مشتقات تیوکربامات به عنوان بازدارنده های غیرنوکلئوزیدی HIV

در راستای مطالعه ارتباط کمی ساختار- فعالیت ضد ایدز یک طبقه ی جدید از مشتقات تیوکربامات ها موسوم به O-2- (۲- هیدروکسی کربونیل بنزآمیدو) اتیل- N- آریل تیوکربامات^۱، از روش رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی (ANN) استفاده شد و قدرت پیش بینی مدل های به دست آمده از هر دو روش توسط پارامترهای آماری مختلف مورد سنجش قرار گرفت.

۴-۱-۱- سری داده ها

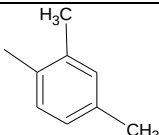
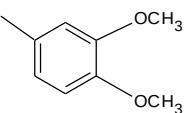
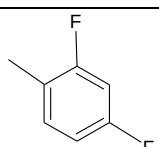
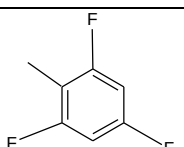
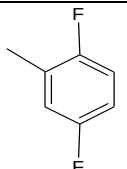
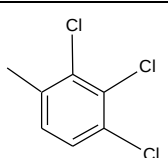
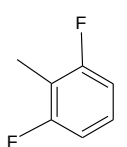
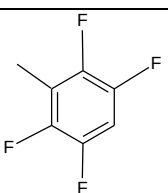
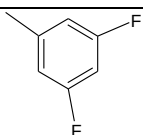
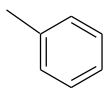
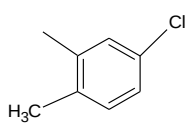
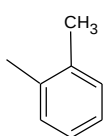
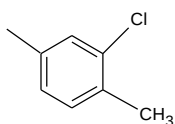
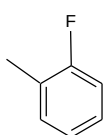
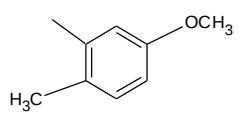
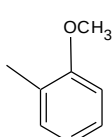
ترکیبات مورد استفاده در این تحقیق، داده های تجربی ۳۲ ترکیب از مشتقات تیوکربامات ها با عنوان O-2- (۲- هیدروکسی کربونیل بنزآمیدو) اتیل- N- آریل تیوکربامات می باشد. شکل (۴-۱) نشان دهنده ی ساختار پایه این ترکیبات است. جدول (۴-۱) ساختار استخلاف آریل و مقادیر تجربی pEC₅₀ بر حسب مولار که در شرایط یکسان اندازه گیری شده و از مرجع [۸۲ و ۸۳] گرفته شده، را نشان می دهد. EC₅₀ غلظتی از دارو است که تا ۵۰٪ سلول MT-4 را از HIV محافظت کند [۸۲].



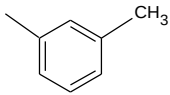
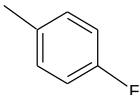
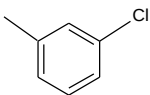
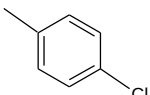
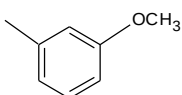
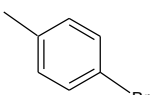
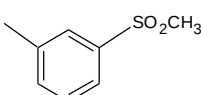
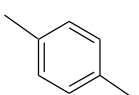
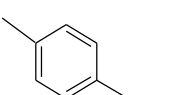
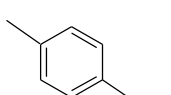
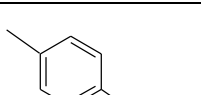
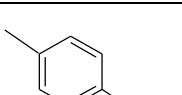
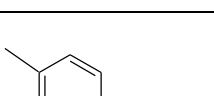
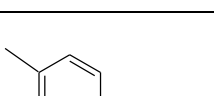
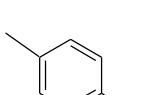
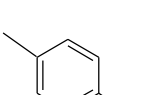
شکل (۴-۱) - اسکلت اصلی O-2- (۲- هیدروکسی کربونیل بنزآمیدو) اتیل - N- آریل تیوکربامات ها

^۱ - O- 2(2-hydroxy carbonyl bezamido) ethyl – N- aryl thiocarbamate

جدول (۱-۴) - ساختار و مقادیر تجربی pEC_{50} مشتقات ۲-۲-۰ (۲- هیدروکسی کربونیل بنزآمیدو) اتیل - N - آریل تیوکربامات

شماره	Aryl	pEC_{50}	شماره	Aryl	pEC_{50}
۱		۴/۲۸	۹		۴/۸۰
۲		۵/۴۲	۱۰		۵/۱۴
۳		۴/۲۹	۱۱		۴/۳۶
۴		۴/۳۵	۱۲		۴/۱۵
۵		۴/۲۱	۱۳		۵/۵۲
۶		۴/۵۲	۱۴		۴/۵۷
۷		۵/۳۷	۱۵		۶/۳۰
۸		۴/۰۶	۱۶		۴/۱۹

ادامه جدول (۱-۴) - ساختار و مقادیر تجربی pEC_{50} مشتقات ۲-۵ - (۲- هیدروکسی کربونیل بنزآمیدو) اتیل- N - آریل تیوکربامات

شماره	Aryl	pEC_{50}	شماره	Aryl	pEC_{50}
۱۷		۵/۳۰	۲۵		۶/۶۰
۱۸		۵/۵۲	۲۶		۷/۰۰
۱۹		۵/۵۸	۲۷		۷/۰۰
۲۰		۴/۹۶	۲۸		۶/۸۹
۲۱		۶/۴۰	۲۹		۶/۵۲
۲۲		۶/۵۲	۳۰		۶/۷۰
۲۳		۵/۴۳	۳۱		۵/۷۰
۲۴		۷/۰۰	۳۲		۶/۳۰

۴-۱-۲- رسم و بهینه‌سازی ساختار مولکول‌ها

ابتدا ساختار شیمیایی ۳۲ ترکیب در نرم افزار Hyperchem 7.0 ترسیم شد. پس از اتصال اتم‌های هیدروژن، این ساختارها با روش کوانتومی AM1 از دسته روش‌های نیمه تجربی بهینه شدند و زمانی بهینه‌سازی پایان یافت که گرادیان به 0.001 کیلو کالری بر مول رسید.

۴-۱-۳- محاسبه و انتخاب توصیفگرهای مهم

برای محاسبه توصیفگرها، ساختار بهینه شده مولکول‌ها به نرم افزار Dragon 2.1 منتقل شد و برای هر ترکیب هزار و چهارصد و هشتاد و یک توصیفگر محاسبه شد ولی تعداد توصیفگرهای زیاد فرایند مدل‌سازی را پیچیده و طولانی می‌کنند و تأثیری در بهبود نتایج ندارند. لذا برای انتخاب توصیفگرهایی که به طور معناداری فعالیت دارویی را به ساختار مولکول ارتباط دهند، مراحل زیر انجام گرفت:

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

ب- همبستگی میان توصیفگرها محاسبه گردید و از میان جفت توصیفگرهایی که ضریب همبستگی بزرگتر از 0.9 داشتند، آنکه که همبستگی کمتری با متغیر وابسته داشت، حذف شد.

ج- روش رگرسیون خطی چندگانه مرحله‌ای بر اساس انتخاب رو به جلو و حذف رو به عقب جهت غربال کردن بهترین توصیفگرهای مورد نیاز با استفاده از نرم افزار SPSS 17 انجام شد که در این مرحله ۹ توصیفگر انتخاب شدند که بیشترین همبستگی با فعالیت ضد HIV ترکیبات مورد مطالعه را دارند.

جدول (۴-۲) کل توصیفگرهای انتخاب شده و طبقه مربوط به آنها را نشان می‌دهد. همچنین ماتریس ضریب همبستگی این توصیفگرها که در جدول (۴-۳) ارائه شده، عدم همبستگی قابل ملاحظه بین توصیفگرها را نشان می‌دهد.

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

معانی	طبقه	نماد	تعداد
Mean square distance index	topological	MSD	۱
3D –MORSE-signal 17	3D -MORSE	Mor17u	۲
2st component symmetry directional WHIM index	WHIM	G2p	۳
Geary auto correlation	2D autocorrelation	GATS8v	۴
2st component accessibility directional WHIM	WHIM	E2m	۵
H auto correlation-lag4	GETAWAY	H4e	۶
3D –MORSE-signal 09	3D -MORSE	Mor09m	۷
3st component symmetry directional WHIM index	WHIM	G3u	۸
Topographic electronic	charge	TE2	۹

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

	MSD	Mor17u	G2p	GATS8v	E2m	H4e	Mor09m	G3u	TE2
MSD	۱								
Mor17u	-۰/۳۳۹	۱							
G2p	-۰/۱۰۶	۰/۵۰۱۰	۱						
GATS8v	-۰/۳۶۶	-۰/۲۴۴	-۰/۳۲۱	۱					
E2m	-۰/۶۴۷	۰/۲۰۱	۰/۰۴۴	۰/۴۴۶	۱				
H4e	۰/۱۵۶	-۰/۵۹۵	-۰/۶۴۶	۰/۳۳۱	-۰/۰۶۳	۱			
Mor09m	۰/۱۰۱	-۰/۰۴۸	۰/۰۱۱	۰/۳۱۵	۰/۱۷۴	-۰/۱۰۶	۱		
G3u	-۰/۰۶۹	۰/۲۵۵	-۰/۱۱۰	-۰/۰۲۰	۰/۱۵۴	۰/۰۸۲	۰/۱۲۳	۱	
TE2	-۰/۰۷۹	-۰/۱۱۱	-۰/۲۷۳	۰/۳۷۴	۰/۱۴۰	۰/۱۰۷	۰/۱۱۷	۰/۲۶۹	۱

۴-۱-۴- کشف پدیده همخطی بین کل توصیفگرهای انتخاب شده

به منظور بررسی وجود پدیده‌ی همخطی در بین توصیفگرهای انتخاب شده، عامل افزایش واریانس

(VIF) مربوط به هر توصیفگر محاسبه شد که نتایج آن در جدول (۴-۴) نشان داده شده است و با

توجه به اینکه برای کلیه متغیرهای مستقل، عامل افزایش واریانس از ۱۰ کمتر است، عدم وجود

همخطی در بین توصیفگرها اثبات می‌گردد.

جدول (۴-۴) - مقادیر عامل افزایش واریانس کل توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه

توصیفگر	MSD	Mor17u	G2p	GATS8v	E2m	H4e	Mor09m	G3u	TE2
VIF	۲/۴۲	۲/۲۶	۲/۰۷	۲/۲۳	۲/۱۰	۲/۳۶	۱/۴۳	۱/۴۵	۱/۴۲

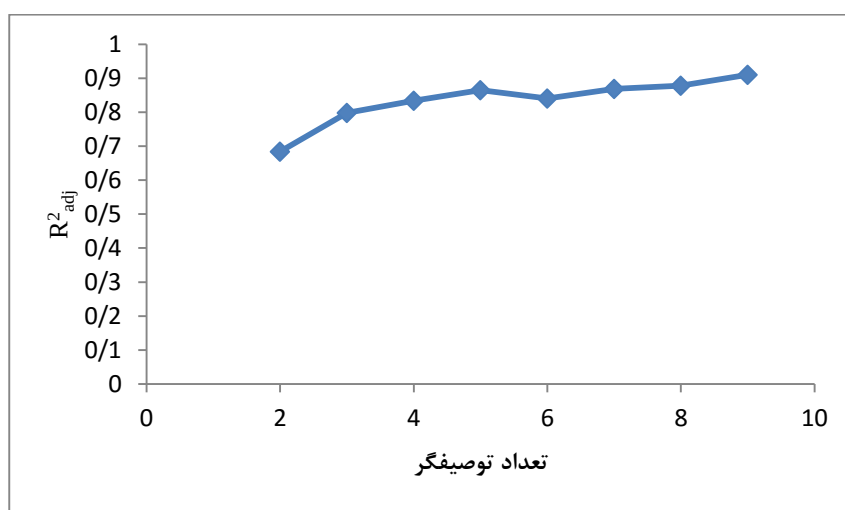
۴-۱-۵- مدل سازی و انتخاب بهترین مدل با روش رگرسیون خطی (MLR)

برای ساختن مدلی که بیانگر ارتباط ساختاری ترکیبات مورد بررسی با فعالیت ضد HIV آن‌ها باشد، از روش رگرسیون خطی چندگانه (MLR) به روش زیر استفاده شد:

توصیفگرهای محاسبه شده توسط نرم افزار Dragon به عنوان متغیر مستقل و pEC_{50} به عنوان متغیر وابسته وارد نرم افزار SPSS شدند سپس با استفاده از روش آنالیز، گزینه رگرسیون و انتخاب روش مرحله‌ای، مدل‌های مختلفی ایجاد گردید. حال کل داده‌ها به طور کاملاً تصادفی به سری آموزش (۲۵ ترکیب شیمیایی) و سری تست (۷ ترکیب شیمیایی) تقسیم شدند. سپس به روش ارزیابی تقاطعی بر روی سری آموزش، مدل‌های مختلف مورد بررسی قرار گرفتند. در این راستا یک ترکیب از سری داده‌های آموزش حذف شد آنگاه با بقیه داده‌های سری آموزش و استفاده از روش Enter ضرائب توصیف‌کننده‌های وارد شده در مدل استخراج شد و سپس با استفاده از این ضرائب مقدار فعالیت دارویی برای ترکیب حذف شده محاسبه گردید. این کار برای تک‌تک ترکیبات سری آموزش و با استفاده از کلیه مدل‌های به دست آمده تکرار شد و با توجه به پارامترهای آماری، مدلی که تعداد توصیفگر کمتر، FIT و ضریب تعیین تصحیح شده (R^2_{adj}) بیشتر دارد، به عنوان مدل برتر انتخاب می‌شود. جدول (۴-۵) مقادیر پارامترهای آماری مدل‌های مختلف را نشان می‌دهد و همچنین شکل (۴-۲) تأثیر تعداد توصیفگر بر مقدار R^2_{adj} ارائه می‌دهد.

جدول (۴-۵) - پارامترهای آماری برای مدل‌های به دست آمده از روش MLR

مدل	۲	۳	۴	۵	۶	۷	۸	۹
FIT	۱/۸۶	۲/۸۸	۳/۰۴	۳/۱۸	۲/۱۸	۲/۲۶	۲/۰۴	۲/۳۸
R^2_{adj}	۰/۶۸۴۳	۰/۷۹۸۴	۰/۸۳۴۳	۰/۸۶۵۳	۰/۸۴۱۰	۰/۸۶۹۰	۰/۸۷۸۰	۰/۹۱۰۴
R^2	۰/۷۱۰۶	۰/۸۲۳۶	۰/۸۶۱۹	۰/۸۹۳۴	۰/۸۸۱	۰/۹۰۷۸	۰/۹۱۹	۰/۹۴۴



شکل (۴-۲) - تأثیر تعداد توصیفگرها بر مقادیر R^2_{adj} مدل‌های MLR

بنابراین مدل ۵ به عنوان مدل برتر انتخاب شد. در جدول (۴-۶) توصیفگرهای مربوط به این مدل،

ضرائب غیراستاندارد و ضرائب استاندارد شده آنها نشان داده شده است.

جدول (۴-۶) - توصیفگرهای مدل منتخب و ضرائب آنها

توصیفگر	طبقه توصیفگر	ضرائب غیر استاندارد	ضرائب استاندارد شده
MSD	Topological	۹۷/۷۶۹	۰/۷۳۲
Mor17u	3D -MORSE	۱/۸۲۸	۰/۵۷۶
G2p	WHIM	-۸۰/۸۳۰	-۰/۴۱۶
GATS8v	2D autocorrelation	-۲/۳۰۷	-۰/۲۱۴
E2m	WHIM	-۲/۲۲۸	-۰/۲۰۰
intercept	_____	-۴/۳۵۷	_____

۴-۱-۶- مدل سازی به روش شبکه عصبی مصنوعی

در این مرحله، جهت مطالعه ارتباط کمی ساختار- فعالیت ضد ایدز O-۲- (۲- هیدروکسی کربونیل بنزآمیدو) اتیل - N - آریل تیوکربامات ها از شبکه عصبی مصنوعی استفاده شد. سری داده ها توسط دو ماتریس جداگانه به طور تصادفی به دو سری آموزش (۲۵ ترکیب) و تست (۷ ترکیب)، تقسیم شدند. سپس پارامترهای مؤثر بر آموزش موفق شبکه عصبی پس انتشار از قبیل تعداد ورودی ها، تعداد لایه های پنهان، تعداد گره های لایه پنهان، نوع تابع انتقال، نوع تابع آموزش، تعداد دوره های آموزش و پارامتر μ با استفاده از سری آموزش و با روش ارزیابی تقاطعی بهینه شدند. در طی فرایند آموزش مقادیر وزن های ارتباطی تغییر می کند و معیار برای پایان دادن آموزش و بهینه شدن مقادیر وزن ها جهت پیش بینی پارامتر مورد بررسی، مقدار خطای شبکه است که طبق تابع کارایی تعیین شده، محاسبه می شود که در این روش، تابع کارایی، میانگین مربع خطا (MSE) در سری آموزش می باشد.

۴-۱-۶-۱- انتخاب تعداد لایه های پنهان

متأسفانه هیچ گونه نتایج تئوری قابل دسترس یا قوانین تجربی رضایت بخش که در تعیین تعداد لایه های پنهان به ما کمک کند، وجود ندارد با این وجود برای اغلب کاربردهای شبکه عصبی در شیمی، به نظر می رسد که یک لایه مخفی کافی باشد [۲۱ و ۶۸]. به این ترتیب شبکه عصبی تشکیل شده، یک شبکه سه لایه شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی است که تعداد ورودی ها (توصیفگرها) با تعداد نرون های لایه ورودی برابر است و لایه خروجی دارای یک نرون است که نشان دهنده PEC_{50} متناظر با هر ترکیب می باشد. ولی تعداد نرون های لایه پنهان باید بهینه شود.

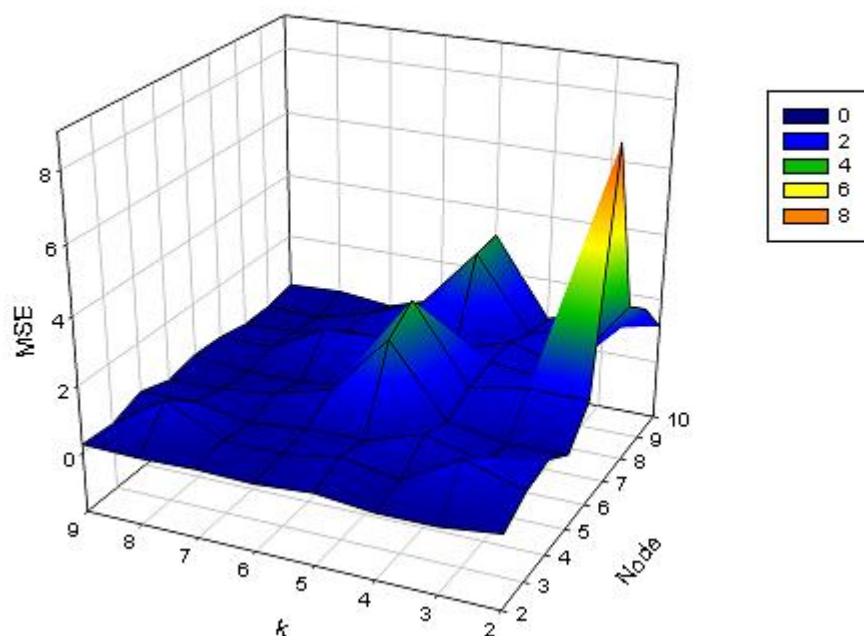
۴-۱-۶-۲- انتخاب تعداد ورودی‌های شبکه، نوع تابع آموزش، نوع تابع انتقال و

تعداد نرون‌های لایه پنهان

برای بهینه‌سازی تعداد ورودی‌های شبکه، نوع تابع آموزش، نوع تابع انتقال و تعداد گره‌های لایه پنهان، شبکه با ورودی‌های از ۲ تا ۹ توصیفگر 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 ($\tan \text{sig}$) و لگاریتم سیگموئیدی ($\log \text{sig}$) به عنوان تابع انتقال استفاده شد. در تمامی این شبکه‌ها تعداد دورهای آموزش ثابت و برابر ۲۰ در نظر گرفته شد. همچنین از تابع انتقال خطی (purelin) در لایه خروجی استفاده گردید. بدین ترتیب در مجموع ۲۸۸ شبکه سه لایه آموزش داده شد و در روند بهینه‌سازی پارامترهای فوق، به حداقل رساندن میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش به عنوان معیار انتخاب شد. نتایج بهینه‌سازی این پارامترها در جداول (۴-۷)، (۴-۸)، (۴-۹) و (۴-۱۰) گردآوری شده است. نمودار سه بعدی این جداول نیز در شکل‌های (۴-۳)، (۴-۴)، (۴-۵) و (۴-۶) ارائه شده است. بر اساس این نتایج، تعداد متغیر ورودی ۹، تعداد ۴ گره در لایه پنهان و تابع لگاریتم سیگموئیدی به عنوان تابع انتقال لایه پنهان انتخاب شد.

جدول (۷-۴) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال تانزانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

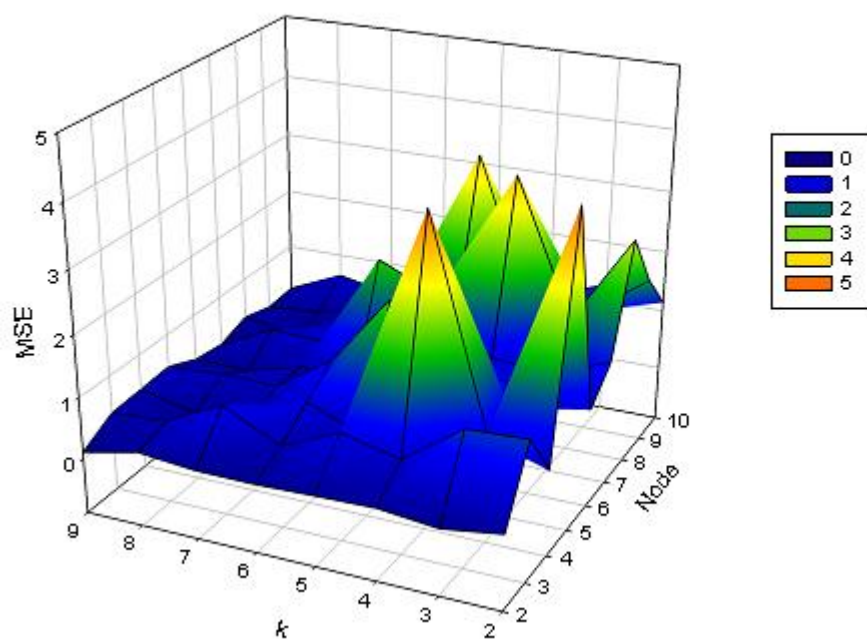
تعداد گره های لایه پنهان	تعداد توصیفگرها								
		۲	۳	۴	۵	۶	۷	۸	۹
	۲	۰/۴۰۴۶	۰/۱۹۰۴	۰/۱۷۹۵	۰/۳۵۵۶	۰/۲۲۹۳	۰/۲۷۲۶	۰/۲۲۲۸	۰/۲۵۸۹
	۳	۰/۸۵۷۹	۱/۳۹۶۷	۰/۳۹۷۸	۰/۶۸۲۷	۰/۳۸۷۰	۰/۴۲۴۴	۱/۲۶۸۸	۰/۱۹۱۹
	۴	۱/۰۳۰	۰/۵۶۹۸	۱/۰۱۸	۰/۴۷۲۸	۰/۴۴۵۸	۰/۲۸۴۵	۰/۵۶۸۵	۰/۵۲۰
	۵	۰/۴۶۰۱	۰/۴۰۹۲	۰/۷۱۹۱	۲/۸۰۰۷	۰/۹۷۶۵	۰/۵۱۲۳	۰/۲۸۹۶	۰/۲۲۸۰
	۶	۱/۳۸۳۹	۱/۳۹۵۶	۱/۰۷۷	۳/۳۷۳۷	۰/۴۶۱۱	۰/۲۱۶۳	۰/۲۰۷۳	۰/۴۰۹۹
	۷	۸/۱۵۲۷	۱/۰۱۳۹	۱/۲۱۹۸	۰/۳۱۹۶	۱/۳۳۰۷	۱/۲۵۰	۰/۳۶۴۷	۰/۳۶۰
	۸	۲/۹۵۵۵	۰/۶۸۶۷	۱/۲۲۲۱	۰/۶۲۹۷	۰/۵۵۶۳	۰/۶۰۶۱	۰/۲۷۰۵	۰/۱۷۸۸
	۹	۲/۲۸۶۹	۱/۳۴۸۳	۰/۸۹۶۵	۳/۰۹۸۱	۱/۲۱۴۹	۰/۵۹۶۵	۰/۱۳۴۶	۰/۱۸۳۹
	۱۰	۱/۱۵۸۷	۱/۲۵۵۲	۰/۷۶۹۶	۳/۱۵۳۵	۰/۷۷۷۶	۰/۲۴۲۲	۰/۴۲۴۰	۰/۳۴۸



شکل (۴-۳) - نمودار سه بعدی شبکه‌هایی با تابع آموزش لونبرگ-مارکوارت و تابع انتقال تانزانت سیگموئیدی

جدول (۴-۸) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

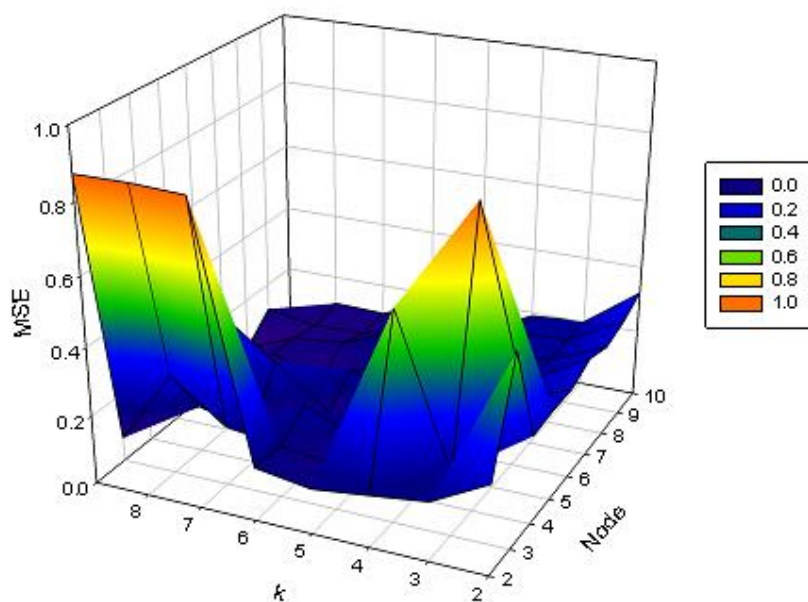
تعداد گره های لایه پنهان	تعداد توصیفگرها								
		۲	۳	۴	۵	۶	۷	۸	۹
	۲	۰/۳۱۵۳	۰/۱۷۴۶	۰/۲۷۵۷	۰/۲۵۰۶	۰/۲۱۹۹	۰/۲۲۷۰	۰/۳۱۴۵	۰/۱۱۱۸
	۳	۱/۳۶۱۷	۱/۳۲۲۴	۰/۶۴۴۳	۰/۸۶۴۹	۰/۴۴۲۳	۰/۹۰۰	۰/۴۶۷۲	۰/۳۸۹۱
	۴	۰/۴۹۵۴	۰/۸۶۲۲	۴/۱۰۳۵	۰/۴۰۱۱	۰/۷۹۵۹	۰/۳۲۴۱	۰/۲۰۶۸	۰/۴۲۱۹
	۵	۴/۱۶۹۳	۰/۹۱۶۲	۰/۷۰۴۶	۲/۲۶۱۳	۱/۱۲۱	۰/۴۹۶۳	۰/۵۱۷۵	۰/۴۵۸۹
	۶	۰/۷۰۵۳	۰/۹۰۴۱	۰/۸۹۵۵	۰/۷۸۳۰	۰/۴۹۶۵	۰/۴۷۳۵	۰/۲۳۳۳	۰/۲۴۱۱
	۷	۱/۱۲۸۵	۰/۷۶۲۱	۰/۸۲۸۸	۱/۷۱۹۳	۲/۱۳۱۸	۰/۵۷۱۸	۰/۶۰۳۰	۰/۱۶۶۲
	۸	۲/۷۸۳۳	۱/۳۲۸۴	۳/۵۲۰۳	۱/۱۰۲۵	۱/۲۳۸۷	۰/۵۴۲۵	۰/۱۷۴۲	۰/۳۲۷۳
	۹	۱/۸۳۳۱	۱/۱۳۲۴	۰/۸۹۷۷	۳/۴۴۰۷	۱/۰۵۷۸	۰/۴۵۷۸	۰/۲۷۴۷	۰/۱۸۹۹
	۱۰	۱/۱۴۵۰	۱/۲۷۷۴	۰/۹۶۷۲	۱/۶۱۶۲	۰/۶۶۲۰	۰/۴۴۵۱	۰/۶۲۸۵	۰/۲۲۸۲



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

جدول (۴-۹) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانزانت سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

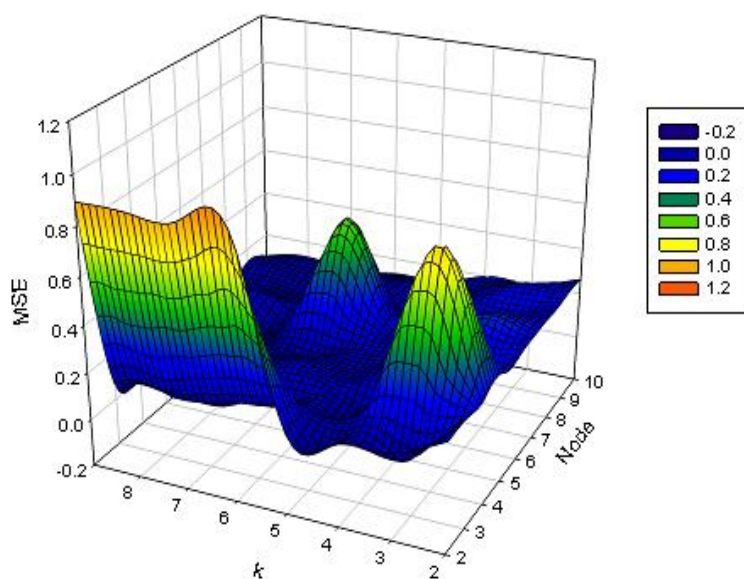
تعداد گره های لایه پنهان	تعداد توصیفگرها								
		۲	۳	۴	۵	۶	۷	۸	۹
	۲	۰/۲۵۸۸	۰/۱۷۱۳	۰/۱۵۰۹	۰/۱۳۱۴	۰/۱۵۶۰	۰/۸۶۸۲	۰/۸۷۵۲	۰/۸۷۳۹
	۳	۰/۵۵۳۴	۰/۱۸۰۷	۰/۶۰۵۲	۰/۱۳۰۴	۰/۱۵۱۶	۰/۱۷۳۸	۰/۲۹۰۶	۰/۰۶۸۹
	۴	۰/۲۵۶۴	۰/۸۶۲۷	۰/۱۶۴۵	۰/۱۴۲	۰/۲۰۶۲	۰/۰۹۰۷	۰/۱۳۴۷	۰/۰۵۰۹
	۵	۰/۲۶۳۶	۰/۱۸۴۸	۰/۱۷۸۱	۰/۱۱۶	۰/۱۱۹۵	۰/۲۴۳۵	۰/۳۵۳۱	۰/۰۴۷۵
	۶	۰/۲۶۲۲	۰/۱۹۷۱	۰/۱۸۱۷	۰/۱۳۷۶	۰/۱۹۲۰	۰/۰۷۵۱	۰/۰۹۰۸	۰/۱۵۶۶
	۷	۰/۲۹۲۲	۰/۱۹۷۶	۰/۱۱۲۷	۰/۱۴۲	۰/۱۳۰۹	۰/۰۷۰۸	۰/۱۱۶۳	۰/۰۴۹
	۸	۰/۲۶۷۶	۰/۱۹۷۴	۰/۱۹۰۲	۰/۱۶۶۲	۰/۱۹۰۰	۰/۱۲۴۷	۰/۱۰۹۹	۰/۰۴۵
	۹	۰/۲۸۷۰	۰/۲۰۱۵	۰/۱۹۰۵	۰/۱۱۰۵	۰/۱۵۶۰	۰/۰۹۶۴	۰/۰۹۳۳	۰/۱۱۸۹
	۱۰	۰/۳۲۰۹	۰/۲۰۰۳	۰/۱۸۹۸	۰/۱۰۹۱	۰/۱۶۹۳	۰/۱۰۷۶	۰/۱۱۱۴	۰/۰۴۷۵



شکل (۴-۵) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال تانزانت سیگموئیدی

جدول (۴-۱۰) - مقادیر میانگین مربعات خطاهای حاصل از ارزیابی تقاطعی سری آموزش برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد متفاوت توصیفگرهای ورودی و تعداد مختلف گره در لایه پنهان

تعداد گره های لایه پنهان	تعداد توصیفگرها								
		۲	۳	۴	۵	۶	۷	۸	۹
	۲	۰/۲۶۲۲	۰/۱۶۰۵	۰/۱۴۷۹	۰/۱۳۶۷	۰/۸۹۸۳	۰/۸۹۱۲	۰/۸۹۰۵	۰/۸۹۴۶
	۳	۰/۲۶۲۴	۰/۱۷۳۰	۰/۱۵۶۵	۰/۱۴۱۲	۰/۱۴۸۴	۰/۱۰۸۶	۰/۰۸۲۳	۰/۰۵۸۶
	۴	۰/۲۶۰۰	۰/۸۱۳۲	۰/۱۵۸۷	۰/۱۳۶۳	۰/۱۴۴۶	۰/۰۸۵۲	۰/۲۳۳۳	۰/۰۴۵۶
	۵	۰/۲۶۵۰	۰/۱۸۳۰	۰/۱۷۷۲	۰/۱۳۳۱	۰/۱۵۳۱	۰/۱۲۰۸	۰/۰۸۰۳	۰/۱۱۳۲
	۶	۰/۲۵۵۶	۰/۲۰۵۴	۰/۱۸۵۵	۰/۱۴۷۰	۰/۱۴۰۲	۰/۱۰۷۳	۰/۰۹۱۵	۰/۰۶۹۸
	۷	۰/۲۶۴۵	۰/۳۰۷۷	۰/۱۸۸۸	۰/۱۹۱۳	۰/۵۵۸۷	۰/۱۰۱۰	۰/۱۹۶۲	۰/۱۱۷۴
	۸	۰/۲۶۳۴	۰/۱۹۹۰	۰/۱۸۳۶	۰/۱۲۳۱	۰/۱۹۲۹	۰/۱۱۴۸	۰/۱۰۷۸	۰/۰۶۹۸
	۹	۰/۲۶۹۸	۰/۲۶۱۰	۰/۱۸۸۸	۰/۱۲۲۴	۰/۱۳۰۲	۰/۰۹۷۴	۰/۱۲۳۳	۰/۱۳۸۹
	۱۰	۰/۲۶۳۳	۰/۲۰۱۷	۰/۱۹۷۳	۰/۱۳۱۱	۰/۸۹۸۳	۰/۱۰۳۱	۰/۱۳۰۸	۰/۰۹۷۶



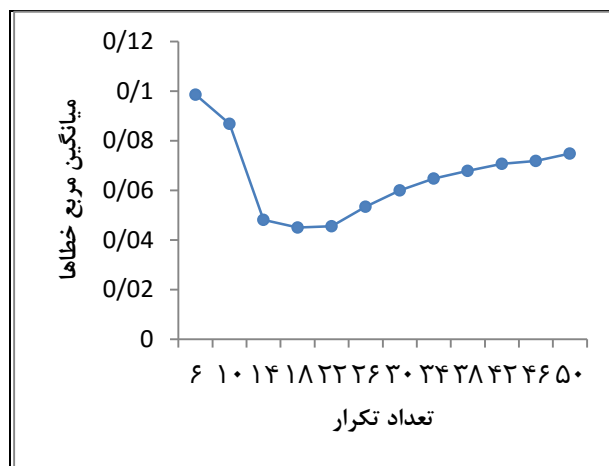
شکل (۴-۶) - نمودار سه بعدی شبکه‌هایی با تابع آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی

۴-۱-۶-۳- انتخاب تعداد دوره‌های آموزش

در طی آموزش شبکه، دقت تقریبی ANN برای داده‌های سری آموزش افزایش می‌یابد در حالی که توانایی پیشگویی برای ورودی‌های خارجی^۱ به شدت کاهش می‌یابد که این حالت را آموزش بیش از حد می‌گویند [۲۴]. برای جلوگیری از این حالت، از تکنیک توقف زودرس استفاده شد. بدین منظور با ۹ توصیفگر منتخب، تابع انتقال لگاریتم سیگموئیدی و قرار دادن ۴ گره در لایه پنهان، تعداد چرخه آموزش تغییر داده شد و برای هر چرخه میانگین مربع خطاهای سری آموزش به روش ارزیابی تقاطعی برآورد شد (جدول (۴-۱۱)) سپس منحنی تعداد چرخه‌های آموزش بر حسب میانگین مربع خطاها بر طبق شکل (۴-۷) ترسیم شد. مقدار بهینه تعداد تکرار جایی است که میزان این خطا حداقل باشد.

جدول (۴-۱۱)- مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در تعداد دوره‌های آموزش مختلف

تعداد تکرار	میانگین مربع خطا	تعداد تکرار	میانگین مربع خطا	تعداد تکرار	میانگین مربع خطا
۶	۰/۰۹۸۶	۲۲	۰/۰۴۵۵	۳۸	۰/۰۶۷۹
۱۰	۰/۰۸۶۸	۲۶	۰/۰۵۳۴	۴۲	۰/۰۷۰۷
۱۴	۰/۰۴۸۱	۳۰	۰/۰۵۹۹	۴۶	۰/۰۷۱۹
۱۸	۰/۰۴۵۰	۳۴	۰/۰۶۴۷	۵۰	۰/۰۷۴۸



شکل (۴-۷)- نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب تعداد دوره‌های آموزش متفاوت

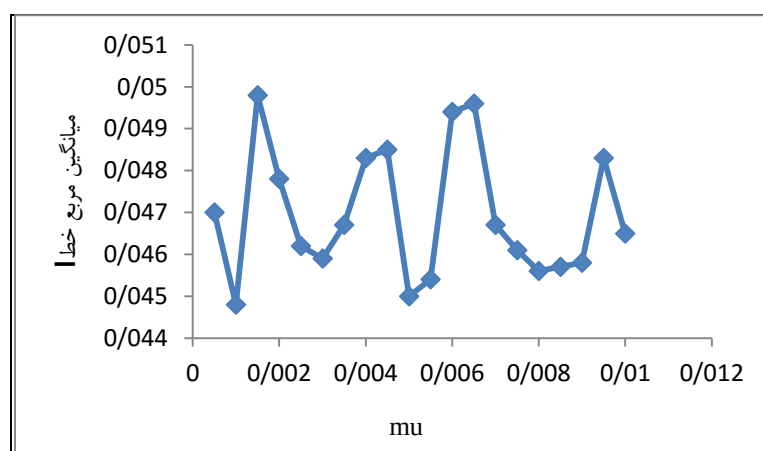
^۱ - External inputs

۴-۶-۱- بهینه کردن مقدار μ

جهت یافتن مقدار بهینه μ ، ساختار شبکه با ۹ متغیر ورودی، ۴ گره در لایه پنهان و الگوریتم آموزشی تنظیم بایزین در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۰۵ تا ۰/۰۱ با گام‌های ۰/۰۰۰۵ تغییر داده شد و آنگاه برای هر مورد، مقدار میانگین مربع خطای حاصل از ارزیابی تقاطعی سری آموزش محاسبه شد. 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 (۴-۸) مقدار بهینه ۰/۰۰۱ به دست آمد.

جدول (۴-۱۲) - مقادیر میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش در مقادیر متفاوت μ

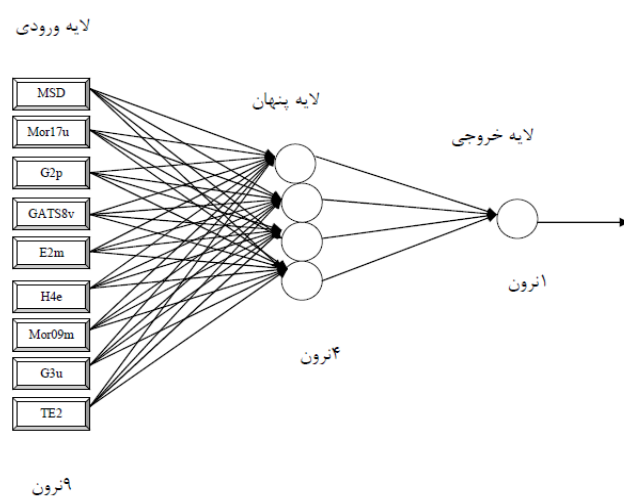
ممنتوم	میانگین مربع خطا	ممنتوم	میانگین مربع خطا
۰/۰۰۰۵	۰/۰۴۷	۰/۰۰۵۵	۰/۰۴۵۴
۰/۰۰۱	۰/۰۴۴۸	۰/۰۰۶	۰/۰۴۹۴
۰/۰۰۱۵	۰/۰۴۹۸	۰/۰۰۶۵	۰/۰۴۹۶
۰/۰۰۲	۰/۰۴۷۸	۰/۰۰۷	۰/۰۴۶۷
۰/۰۰۲۵	۰/۰۴۶۲	۰/۰۰۷۵	۰/۰۴۶۱
۰/۰۰۳	۰/۰۴۵۹	۰/۰۰۸	۰/۰۴۵۶
۰/۰۰۳۵	۰/۰۴۶۷	۰/۰۰۸۵	۰/۰۴۵۷
۰/۰۰۴	۰/۰۴۸۳	۰/۰۰۹	۰/۰۴۵۸
۰/۰۰۴۵	۰/۰۴۸۵	۰/۰۰۹۵	۰/۰۴۸۳
۰/۰۰۵	۰/۰۴۵۰	۰/۰۱	۰/۰۴۶۵



شکل (۴-۸) - نمودار میانگین مربع خطاهای حاصل از ارزیابی تقاطعی سری آموزش بر حسب پارامتر μ

۴-۱-۶-۵- معماری شبکه عصبی بهینه شده

با توجه به روند بهینه سازی که در قسمت های قبلی ذکر شد، شبکه عصبی مصنوعی که بتواند فعالیت ضد HIV ترکیبات ۲-O - (۲- هیدروکسی کربونیل بنز آمیدو) اتیل - N - آریل تیوکربامات را پیش بینی کند، دارای ۹ توصیفگر، ۴ گره در لایه پنهان، تابع آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی است. شکل (۴-۹) ساختار شبکه عصبی بهینه نهایی را نشان می دهد.



شکل (۴-۹) - ساختار شبکه عصبی بهینه شده

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۴-۱۳) ارائه شده است.

جدول (۴-۱۳) - توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

Train br	تابع آموزش
Logsig	تابع انتقال لایه پنهان
pure line	تابع انتقال لایه خروجی
۴	تعداد نرون لایه پنهان
۹	تعداد متغیرهای ورودی
۱۸	تعداد دوره های آموزش
۰/۰۰۱	پارامتر μ

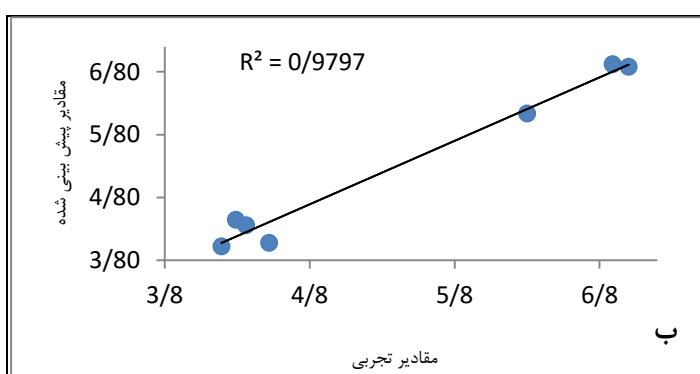
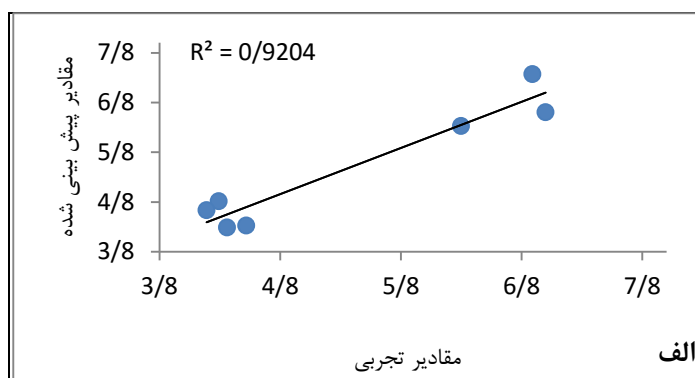
۷-۱-۴- ارزیابی مدل‌های خطی و غیرخطی

۷-۱-۴-۱- ارزیابی مدل خطی و غیرخطی با استفاده از سری تست

به منظور مطالعه قدرت پیش‌بینی مدل‌های منتخب خطی و غیرخطی به دست آمده، سری تست مورد استفاده قرار گرفت. در این راستا در مدل خطی با استفاده از ضرائب مدل برتر و در مدل غیرخطی با استفاده از شبکه عصبی بهینه شده، مقادیر pEC_{50} مربوط به ۷ ترکیب که در ساخت مدل استفاده نشده بودند، پیش‌بینی شد (جدول (۴-۱۴)). مقادیر ضریب تعیین در شکل (۴-۱۰)، توانایی بالای مدل خطی و غیرخطی در پیش‌بینی فعالیت دارویی این ترکیبات را تصدیق می‌کند.

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

شماره ترکیب	pIC_{50}				
	مقدار تجربی	مقدار پیش‌بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۳	۴/۲۹	۴/۸۲	۴/۴۴	۱۲/۳۵	۳/۵۰
۶	۴/۵۲	۴/۳۳	۴/۰۸	-۴/۲	-۹/۷۳
۱۱	۴/۳۶	۴/۲۹	۴/۳۶	-۱/۶۱	۰/۰۰
۱۶	۴/۱۹	۴/۶۴	۴/۰۲	۱۰/۷۴	-۴/۰۶
۲۴	۷/۰۰	۶/۶۱	۶/۸۸	-۵/۵۷	-۱/۷۱
۲۸	۶/۸۹	۷/۳۷	۶/۹۲	۶/۹۷	۰/۴۴
۳۲	۶/۳۰	۶/۳۳	۶/۱۴	۰/۴۸	-۲/۵۴



شکل (۴-۱۰) - نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی pEC_{50} برای تست الف: با مدل MLR ب: با مدل ANN

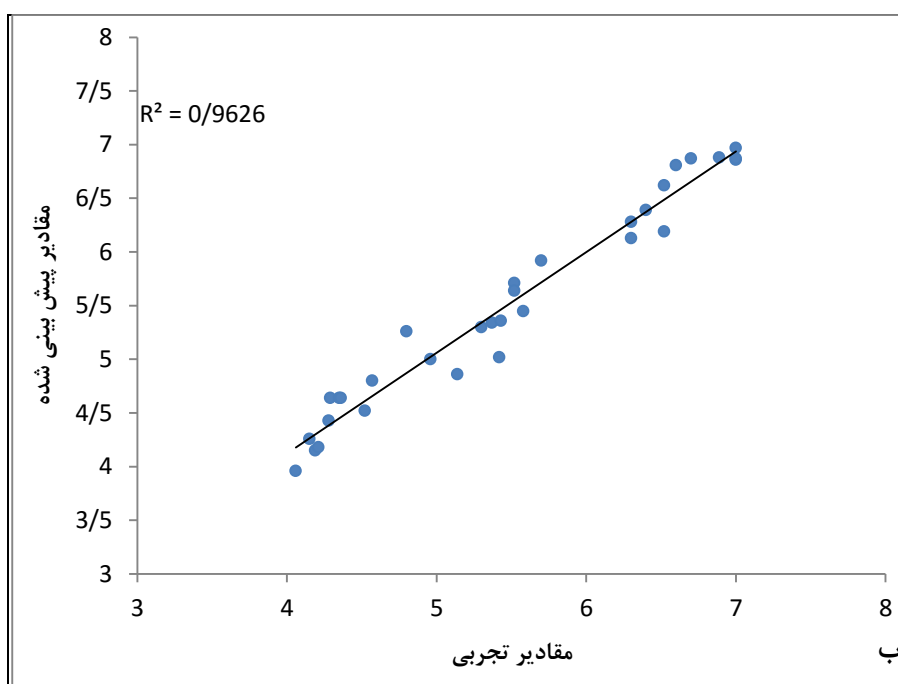
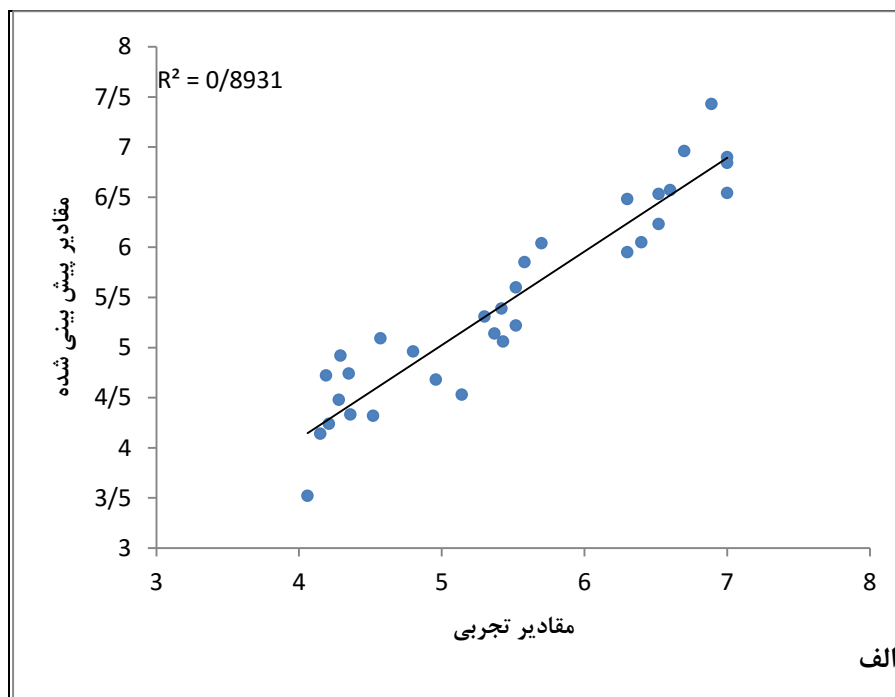
۴-۷-۲- ارزیابی مدل‌های خطی و غیرخطی به روش رد مرحله‌ای تک تک و رد مرحله‌ای

گروهی

از دیگر روش‌هایی که برای ارزیابی بیشتر مدل‌های برتر ارائه شده توسط MLR و ANN به کار گرفته شد، تکنیک رد مرحله‌ای تک تک (برای کل داده‌ها) و رد مرحله‌ای گروهی می‌باشد که نتایج حاصل از این دو روش به ترتیب در جداول (۴-۱۵) و (۴-۱۶) ارائه شده است. این نتایج تعمیم‌پذیری بالای مدل طراحی شده توسط هر دو روش خطی و غیرخطی را بیان می‌کند. ضریب تعیین مشاهده شده بین مقادیر پیش‌بینی شده pEC_{50} و مقادیر تجربی در شکل‌های (۴-۱۱) و (۴-۱۲)، بر نزدیکی مقادیر پیش‌بینی شده توسط مدل‌ها نسبت به مقادیر تجربی آن‌ها دلالت دارد. همچنین در شکل (۴-۱۳) مقادیر باقیمانده‌ها بر حسب مقادیر تجربی pEC_{50} ترکیبات مورد بحث، ترسیم شده است که عدم وجود خطای سیستماتیک را نشان می‌دهد. تمرکز نقاط به سمت محورهای افقی نشان‌دهنده‌ی نزدیک بودن مقادیر پیش‌بینی شده به مقادیر تجربی pEC_{50} است.

جدول (۴-۱۵) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از روش رد مرحله‌ای تک‌تک

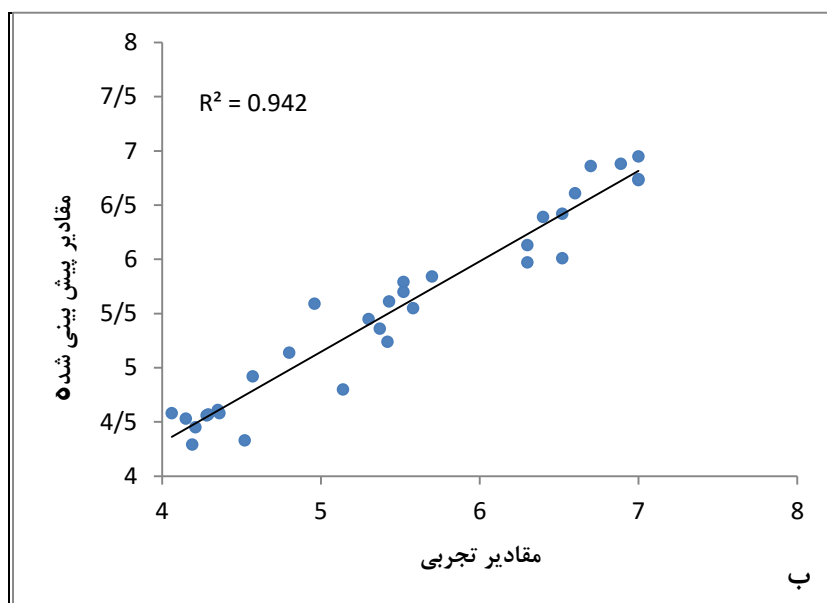
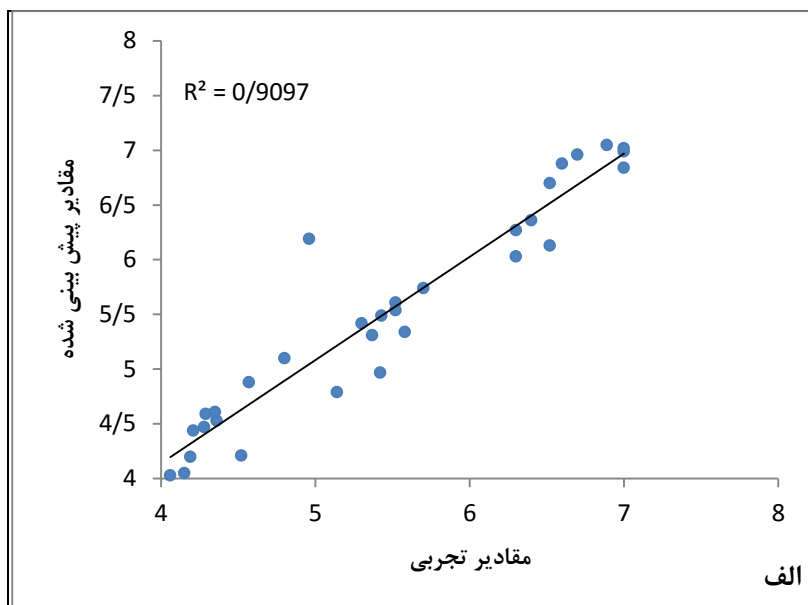
شماره ترکیب	pIC ₅₀				
	مقدار تجربی	مقدار پیش بینی شده		درصد خطا	
		MLR	ANN	MLR	ANN
۱	۴/۲۸	۴/۴۸	۴/۴۳	۴/۶۷	۳/۵۰
۲	۵/۴۲	۵/۳۹	۵/۰۲	-۰/۵۵	-۷/۳۸
۳	۴/۲۹	۴/۹۲	۴/۶۴	۱۴/۶۹	۸/۱۶
۴	۴/۳۵	۴/۷۴	۴/۶۴	۸/۹۷	۶/۶۷
۵	۴/۲۱	۴/۲۴	۴/۱۸	۰/۷۱	-۰/۷۱
۶	۴/۵۲	۴/۳۲	۴/۵۲	-۴/۴۲	۰/۰۰
۷	۵/۳۷	۵/۱۴	۵/۳۴	-۴/۲۸	-۰/۵۶
۸	۴/۰۶	۳/۵۲	۳/۹۶	-۱۳/۳۰	-۲/۴۶
۹	۴/۸۰	۴/۹۶	۵/۲۶	۳/۳۳	۹/۵۸
۱۰	۵/۱۴	۴/۵۳	۴/۸۶	-۱۱/۸۷	-۵/۴۵
۱۱	۴/۳۶	۴/۳۳	۴/۶۴	-۰/۶۹	۶/۴۲
۱۲	۴/۱۵	۴/۱۴	۴/۲۶	-۰/۲۴	۲/۶۵
۱۳	۵/۵۲	۵/۲۲	۵/۷۱	-۵/۴۳	۳/۴۴
۱۴	۴/۵۷	۵/۰۹	۴/۸۰	۱۱/۳۸	۵/۰۳
۱۵	۶/۳۰	۵/۹۵	۶/۱۳	-۵/۵۶	-۲/۷۰
۱۶	۴/۱۹	۴/۷۲	۴/۱۵	۱۲/۶۵	-۰/۹۵
۱۷	۵/۳۰	۵/۳۱	۵/۳۰	۰/۱۹	۰/۰۰
۱۸	۵/۵۲	۵/۶۰	۵/۶۴	۱/۴۵	۲/۱۷
۱۹	۵/۵۸	۵/۸۵	۵/۴۵	۴/۸۴	-۲/۳۳
۲۰	۴/۹۶	۴/۶۸	۵/۰۰	-۵/۶۵	۰/۸۱
۲۱	۶/۴۰	۶/۰۵	۶/۳۹	-۵/۴۷	-۰/۱۶
۲۲	۶/۵۲	۶/۲۳	۶/۱۹	-۴/۴۵	-۵/۰۶
۲۳	۵/۴۳	۵/۰۶	۵/۳۶	-۶/۸۱	-۱/۲۹
۲۴	۷/۰۰	۶/۵۴	۶/۸۶	-۶/۵۷	-۲/۰۰
۲۵	۶/۶۰	۶/۵۷	۶/۸۱	-۰/۴۵	۳/۱۸
۲۶	۷/۰۰	۶/۹۰	۶/۸۷	-۱/۴۳	-۱/۸۶
۲۷	۷/۰۰	۶/۸۴	۶/۹۷	-۲/۲۹	-۰/۴۳
۲۸	۶/۸۹	۷/۴۳	۶/۸۸	۷/۸۴	-۰/۱۵
۲۹	۶/۵۲	۶/۵۳	۶/۶۲	۰/۱۵	۱/۵۳
۳۰	۶/۷۰	۶/۹۶	۶/۸۷	۳/۸۸	۲/۵۴
۳۱	۵/۷۰	۶/۰۴	۵/۹۲	۵/۹۶	۳/۸۶
۳۲	۶/۳۰	۶/۴۸	۶/۲۸	۲/۸۶	-۰/۳۲



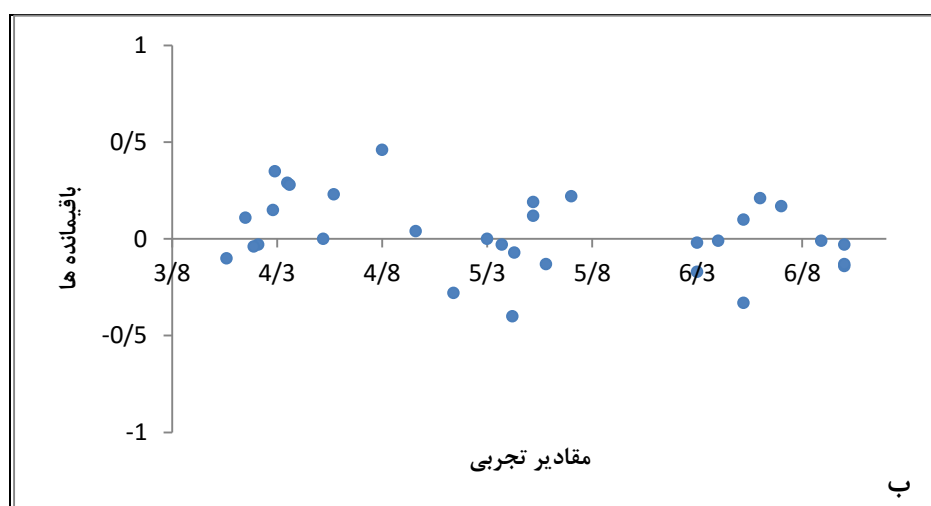
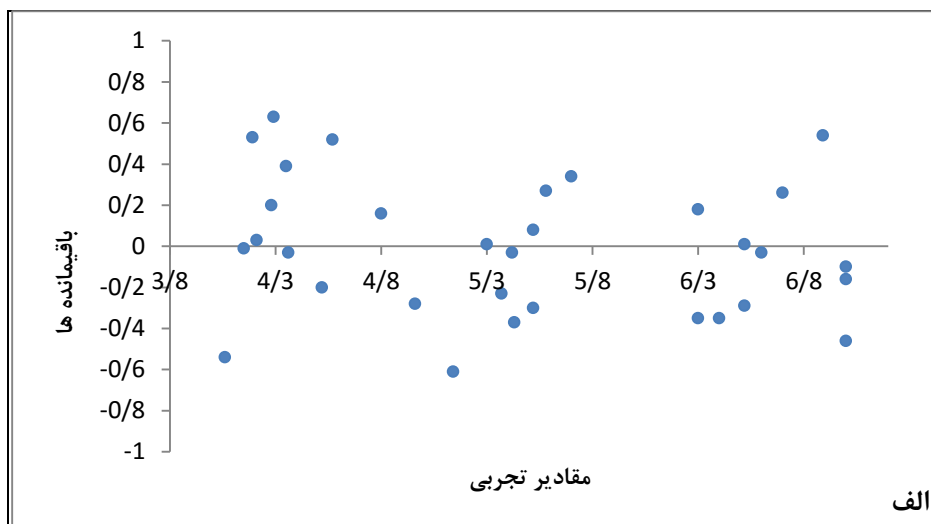
شکل (۴-۱۱) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی pEC_{50} به روش رد مرحله‌ای تک‌تک برای کل داده‌ها
الف: با مدل MLR ب: با مدل ANN

جدول (۴-۱۶) - نتایج حاصل از ارزیابی مدل برتر ارائه شده توسط MLR و ANN با استفاده از روش رد مرحله‌ای گروهی

شماره ترکیب	مقدار تجربی	pIC ₅₀		درصد خطا	
		مقدار پیش بینی شده			
		MLR	ANN	MLR	ANN
۳	۴/۲۹	۴/۵۹	۴/۵۷	۶/۹۹	۶/۵۳
۷	۵/۳۷	۵/۳۱	۵/۳۶	-۱/۱۱۲	-۰/۱۹
۱۰	۵/۱۴	۴/۷۹	۴/۸۰	-۶/۸۱	-۶/۶۱
۱۴	۴/۵۷	۴/۸۸	۴/۹۲	۶/۷۸	۷/۶۶
۱۸	۵/۵۲	۵/۶۱	۵/۷۹	۱/۶۳	۴/۸۹
۲۴	۷/۰۰	۶/۸۴	۶/۷۳	-۲/۲۹	-۳/۸۶
۲۸	۶/۸۹	۷/۰۵	۶/۸۸	۲/۳۲	-۰/۱۵
۳۲	۶/۳۰	۶/۲۷	۶/۱۳	-۰/۴۸	-۲/۷۰
۲	۵/۴۲	۴/۹۷	۵/۲۴	-۸/۳۰	-۳/۳۲
۸	۴/۰۶	۴/۰۳	۴/۵۸	-۰/۷۴	۱۲/۸۱
۱۲	۴/۱۵	۴/۰۵	۴/۵۳	-۲/۴۱	۹/۱۶
۱۵	۶/۳۰	۶/۰۳	۵/۹۷	-۴/۲۹	-۵/۲۴
۱۹	۵/۵۸	۵/۳۴	۵/۵۵	-۴/۳۰	-۰/۵۴
۲۳	۵/۴۳	۵/۴۹	۵/۶۱	۱/۱۰	۳/۳۱
۲۶	۷/۰۰	۶/۹۹	۶/۷۴	-۰/۱۴	-۳/۷۱
۳۱	۵/۷۰	۵/۷۴	۵/۸۴	۰/۷۰	۲/۴۶
۱	۴/۲۸	۴/۴۷	۴/۵۶	۴/۴۴	۶/۵۴
۶	۴/۵۲	۴/۲۱	۴/۳۳	-۶/۸۶	-۴/۲۰
۱۱	۴/۳۶	۴/۵۳	۴/۵۸	۳/۹۰	۵/۰۵
۱۳	۵/۵۲	۵/۵۴	۵/۷۰	۰/۳۶	۳/۲۶
۱۷	۵/۳۰	۵/۴۲	۵/۴۵	۲/۲۶	۲/۸۳
۲۱	۶/۴۰	۶/۳۶	۶/۳۹	-۰/۶۲	-۰/۱۶
۲۷	۷/۰۰	۷/۰۲	۶/۹۵	۰/۲۹	-۰/۷۱
۳۰	۶/۷۰	۶/۹۶	۶/۸۶	۳/۸۸	۲/۳۹
۴	۴/۳۵	۴/۶۱	۴/۶۱	۵/۹۸	۵/۹۸
۵	۴/۲۱	۴/۴۴	۴/۴۵	۵/۴۶	۵/۷۰
۹	۴/۸۰	۵/۱۰	۵/۱۴	۶/۲۵	۷/۰۸
۱۶	۴/۱۹	۴/۲۰	۴/۲۹	۰/۲۴	۲/۳۹
۲۰	۴/۹۶	۶/۱۹	۵/۵۹	۲۴/۸۰	۱۲/۷۰
۲۲	۶/۵۲	۶/۱۳	۶/۰۱	-۵/۹۸	-۷/۸۲
۲۵	۶/۶۰	۶/۸۸	۶/۶۱	۴/۲۴	۰/۱۵
۲۹	۶/۵۲	۶/۷۰	۶/۴۲	۲/۷۶	-۱/۵۳



شکل (۴-۱۲) - نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی PEC_{50} به روش رد مرحله‌ای گروهی برای کل داده‌ها
الف: با مدل MLR ب: با مدل ANN



شکل (۴-۱۳) - نمودار باقیمانده‌ها بر حسب مقادیر تجربی pEC_{50} برای کل داده‌ها الف: با مدل MLR ب: با مدل ANN

۴-۱-۸- ارزیابی مدل‌های برتر خطی و غیرخطی با استفاده از پارامترهای آماری

در راستای بررسی کارایی مدل‌های طراحی شده توسط MLR و ANN چندین پارامتر آماری محاسبه گردید که نتایج آن در جدول (۴-۱۷) ارائه شده است. مقادیر این پارامترها نشان می‌دهد که مدل‌های ارائه شده از قدرت پیشگویی مناسبی برخوردار هستند.

جدول (۴-۱۷) - پارامترهای آماری برای مدل‌های برتر طراحی شده توسط MLR و ANN

پارامتر	سری تست (N=7)		کل داده‌ها (N=32)	
	MLR	ANN	MLR	ANN
MAE	۰/۳۰۵۷	۰/۱۵۲۹	۰/۲۶۵	۰/۱۵۱
MSE	۰/۱۲۹۷	۰/۰۴۰۸	۰/۱۰۶	۰/۰۳۸
PRESS	۰/۹۰۷۹	۰/۲۸۵۶	۳/۳۹۵	۱/۲۱۶
SEP	۰/۳۶۰۱	۰/۲۰۲۰	۰/۳۲۶	۰/۱۹۵
R ²	۰/۹۲۰	۰/۹۷۹	۰/۸۹۳	۰/۹۶۲
REP(%)	۶/۷۱	۳/۷۶۵	۵/۹۵۷	۳/۵۶۴
MRE	۵/۹۸۸	۳/۱۴۰	۵/۰۹۵	۲/۹۱۷
R ² (Leave Group Out)			۰/۹۰۹	۰/۹۴۲

۴-۱-۹ - آزمون Y - تصادفی

برای اثبات قدرت مدل ایجاد شده و اینکه نتایج به دست آمده شانسی نبوده است، آزمون Y - تصادفی انجام گرفت. در این راستا pEC₅₀ (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 جدول (۴-۱۸) آورده شده که مقادیر ضریب تعیین کوچک، نشان‌دهنده‌ی عدم وابستگی در مدل توسعه یافته توسط شبکه عصبی مصنوعی است.

جدول (۴-۱۸) - مقادیر R² سری تست با استفاده از آزمون Y - تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R ²	۰/۲۷۴	۰/۰۴۱	۰/۲۵۵	۰/۰۲۰	۰/۰۶۵	۰/۰۰۰	۰/۱۷۸	۰/۱۴۵	۰/۰۳۶	۰/۰۱۳

۴-۱-۱۰ - بررسی ارتباط توصیفگرهای وارد شده در مدل با فعالیت بازدارندگی

در این قسمت با توجه به توصیفگرهای وارد شده در مدل MLR و ANN یک بررسی اجمالی روی اثرات این متغیرها بر فعالیت دارویی ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده شامل ۹ توصیفگر است که هر کدام بیانگر خصوصیت دو بعدی، هندسی و سه بعدی ترکیبات مورد بررسی می‌باشند.

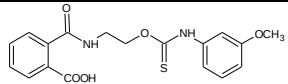
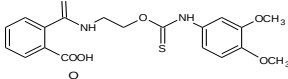
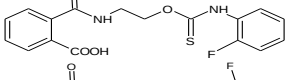
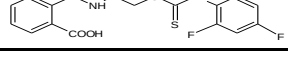
الف- توصیفگرهای Topological

این دسته توصیفگرها که از گراف مولکولی تپی از هیدروژن محاسبه می‌شوند، از ساده‌ترین توصیفگرهای مولکولی بوده و به ساختار فضایی مولکول ارتباطی ندارند. توصیفگرهای توپولوژیکی نشان‌دهنده‌ی نوع اتم، نوع و تعداد پیوند و چگونگی ارتباط با یکدیگر بوده و به سادگی از روی ساختمان دوبعدی مولکول قابل محاسبه‌اند. از میان این توصیفگرها، متوسط مربع شاخص فاصله (MSD)^۱ در مدل ظاهر شده است که به صورت زیر محاسبه می‌شود:

$$MSD = \frac{[\sum_{i=1}^A \sum_{j=1}^A (D_{ij})^2]}{A \cdot (A - 1)} \quad (۱-۴)$$

که A تعداد اتم‌ها به غیر از اتم‌های هیدروژن، D_{ij} ورودی سطر i ام و ستون j ام در ماتریس فاصله است. MSD با افزایش میزان شاخه‌دار شدن کاهش می‌یابد [۸۴]. همچنین با توجه به ضریب استاندارد، این توصیفگر دارای بیشترین اثر روی pEC_{50} می‌باشد. بنابراین اگر تعداد شاخه‌ها بیشتر شود MSD کاهش یافته و به موجب آن pEC_{50} نیز کاهش می‌یابد. به بیان دیگر، با افزایش تعداد استخلاف‌ها میزان برهمکنش متقابل آنزیم نسخه‌بردار معکوس- بازدارنده کمتر شده و به موجب آن تکثیر ویروس HIV سریعتر رخ می‌دهد. چند مثال از اثر این توصیفگر در جدول (۴-۱۹) آورده شده است.

جدول (۴-۱۹)- اثر توصیفگر MSD بر فعالیت دارویی

نام ترکیب	MSD	pEC_{50}
	۰/۲۹۲	۴/۵۷
	۰/۲۸۶	۴/۲۸
	۰/۲۹	۶/۳۰
	۰/۲۸۰	۵/۱۴

^۱ - Mean Square Distance index

ب- توصیفگرهای 3D MORSE

همانطور که در بخش (۳-۱-۱۱) اشاره شد، توصیفگرهای 3D-MORSE از معادله‌ای که در مطالعات پراش الکترونی بکار می‌رود و امکان نمایش سه بعدی مولکول به صورت مقادیری ثابت فراهم می‌کند، قابل محاسبه‌اند. این توصیفگرها قادرند ارتباط بین ساختار سه بعدی ترکیبات آلی و خصوصیات فیزیکی، شیمیایی و بیولوژیکی را فراهم سازند [۷۲]. در تحقیق حاضر، برای پیش‌بینی فعالیت دارویی، توصیفگر Mor17u و Mor09m از این گروه توصیفگرها وارد مدل برتر شد که با توجه به علامت مثبت آنها در مدل می‌توان گفت که با افزایش مقادیر این توصیفگرها pEC₅₀ نیز افزایش می‌یابد.

پ-توصیفگرهای WHIM^۱

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 توصیفگر از رابطه (۴-۲) به دست می‌آید:

$$S_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_j) (q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad (۴-۲)$$

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

^۱ -Weighted Holistic Invariant Molecular descriptors

۱ - حالت بدون وزن (u) ۲- وزن دار شده با جرم اتمی ۳- وزن دار شده با حجم و اندروالس ۴- وزن دار شده با الکترون گاتیویته ساندرسون ۵- وزن دار شده با حالت الکتروتوپولوژیکی کییر و هال^۱ ۶- وزن دار شده با قطبش پذیری [۷۲ و ۸۵]

از این نوع توصیفگرها E2m, G2p, G3u توسط مدل انتخاب گردید که به تأثیر صورتبندی دارو (بازدارنده) در طی برهمکنش آن با آنزیم نسخه بردار معکوس اشاره دارد.

ت - توصیفگرهای 2D-autocorrelation

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

$$C(d) = \frac{\frac{1}{2\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - w_j)^2}{\frac{1}{A-1} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad (3-4)$$

که w یک ویژگی اتم، $\bar{w}$ میانگین مقدار آن ویژگی روی مولکول، A تعداد اتمها و d فاصله‌ی توپولوژیکی است و δ_{ij} نیز - که به تابع کرونکر^۲ معروف است - در حالتی که $d_{ij} = d$ باشد یک است، و در غیر اینصورت صفر است و Δ هم مجموعه‌ی δ هاست. مقدار این توصیف‌کننده که از جنس فاصله است، از صفر تا بی‌نهایت متغیر است [۷۲] از میان این توصیفگرها GATS8v که با حجم و اندروالس وزن دار شده در مدل انتخاب شده و نشان می‌دهد که حجم کلی استخلاف، فاکتوری تعیین کننده برای فعالیت دارویی آن ترکیب می‌باشد و افزایش این توصیفگر باعث کاهش pEC_{50} می‌شود.

¹ - Kier-hall

² - Kronecker

ث- توصیفگرهای GETAWAY

H4e از گروه توصیفگرهای H- GETAWAY می‌باشد که از ماتریس تأثیر مولکولی محاسبه گردیده و مجموع مشارکت جفت اتمهایی که در فاصله توپولوژیکی ۴ قرار گرفته‌اند و با الکترونگاتیویته ساندerson وزن دار شده اند را در نظر می‌گیرد. این توصیفگر سه بعدی بیانی از شکل کلی و اندازه بازدارنده می‌باشد [۷۲ و ۸۶] و بر اساس مدل طراحی شده با افزایش مقدار این توصیفگر مقدار pEC_{50} افزایش می‌یابد.

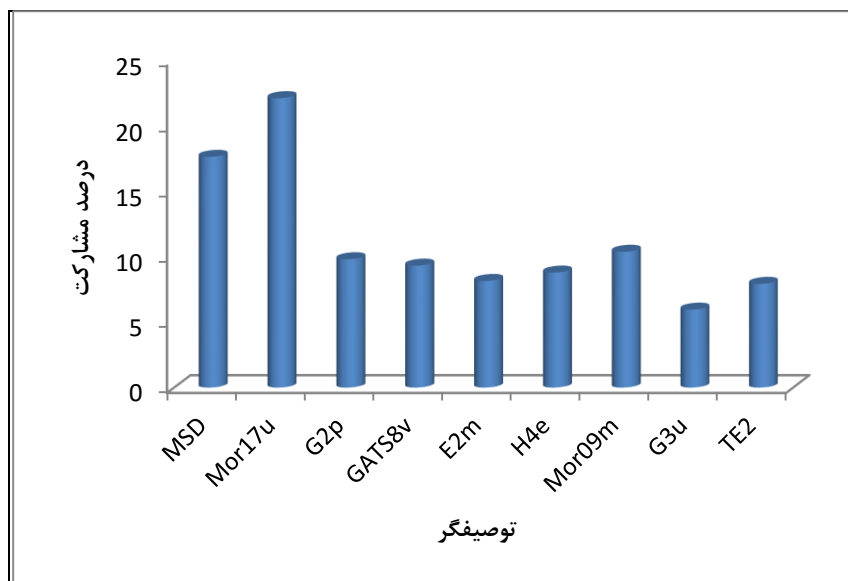
ج- توصیفگرهای Charge

با توجه به تئوری کلاسیک، برهم‌کنشهای بین مواد از نظر ماهیت، الکتروستاتیک یا کووالانسی است. تغییرات بار الکتریکی در مولکول به عنوان نیروی پیش‌برنده‌ی برهم‌کنشهای الکتروستاتیک است. از جمله این توصیفگرها که در مدل برگزیده شده، توصیفگر الکترونی توپوگرافیکی^۱ است که می‌تواند توزیع بار را در بازدارنده توصیف کند. 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 [۸۷].

۴-۱-۱۱- بررسی میزان مشارکت توصیفگرهای منتخب در شبکه عصبی

بر اساس روشی که در بخش (۳-۱-۱۲) توضیح داده شد، میزان مشارکت توصیفگرهای منتخب در پیش‌بینی فعالیت دارویی موردنظر برآورد شد که نتایج آن در شکل (۴-۱۴) ارائه شده است.

^۱ - Topographic Electronic descriptor



شکل (۴-۱۴) - مشارکت توصیفگرها در شبکه عصبی بهینه

طبق شکل فوق توصیفگر Mor17u دارای بیشترین اثر مشارکت می‌باشد. این توصیفگر در برگیرنده اطلاعات سه بعدی مولکول است که این اطلاعات به صورتبندی دارو ارتباط دارد بنابراین می‌توان گفت که فعالیت بازدارنده این ترکیبات وابستگی قابل ملاحظه‌ای به تغییرات صورتبندی دارو دارد. توصیفگر دیگری که درصد مشارکت بالایی دارد، توصیفگر MSD از دسته توصیفگرهای توپولوژیکی است که اهمیت شاخه دار شدن ترکیبات را در میزان بازدارندگی بیان می‌کند.

۴-۱-۱۲- نتیجه‌گیری نهایی

همانطور که گفته شد هدف اصلی از کار حاضر، توسعه روشهای خطی و غیرخطی جهت مطالعه ارتباط کمی بین ساختار و فعالیت ضد HIV ترکیبات ۰-۲- (۲- هیدروکسی کربونیل بنزامیدو) اتیل-N- آریل تیوکربامات می‌باشد. نتایج نشان می‌دهند که مدل‌های خطی و غیرخطی توسعه یافته می‌توانند به عنوان روش‌های موفق برای مدل‌سازی و پیشگویی فعالیت بازدارندگی ترکیبات مورد مطالعه باشند و این مدل‌ها می‌توانند به محققان در طراحی داروهایی با فعالیت بازدارندگی قویتر کمک کنند.

آینده نگری

- ❖ پژوهشگران می‌توانند نتایج حاصل از این مطالعات را در طراحی ترکیبات (داروها) جدید با فعالیت بازدارندگی قویتر نسبت به آنزیم نسخه‌بردار معکوس به کار ببرند.
- ❖ از آن جایی که در تحقیق حاضر توصیفگرهای به دست آمده از روش رگرسیون خطی چندگانه به عنوان ورودی شبکه عصبی مصنوعی قرار گرفته است و چون در روشهای رگرسیون خطی چندگانه تنها برهم‌کنشهای خطی بین پارامترها لحاظ می‌شود، لذا می‌توان از روشهایی همچون ژنتیک الگوریتم، طرح‌ریزی متوالی^۱ برای انتخاب توصیفگرهای مهم استفاده کرد و می‌توان روشهایی مانند SVM^۲ و LS-SVM^۳ را به جای شبکه عصبی به کار برد.
- ❖ می‌توان از روشهای خطی دیگر مانند حداقل مربعات جزئی و آنالیز اجزای اصلی را جایگزین روش رگرسیون خطی چندگانه کرد و نتایج به دست آمده را با هم مقایسه کرد.
- ❖ از روشهای نوین مانند CoMFA^۴ و CoMSIA^۵ برای پیش‌بینی فعالیت بازدارندگی ترکیبات مورد مطالعه استفاده کرد.

^۱ - Successive Projection Algorithm

^۲ - Supported Vector Machine

^۳ - Least Square Supported Vector Machine

^۴ - Comparative Molecular Field Analysis

^۵ - Comparative Molecular Similarity Indices Analysis

- [1]-<http://micro.magnet.fsu.edu/cells/viruses/hivvirus.html>.
- [2]-http://en.wikipedia.org/wiki/Discovery_and_development_of_nonnucleoside_reverse_transcriptase.
- [3]- <http://www.iranhiv.com/HIVbasic.htm>.
- [4]- [http://www.takbarg.ir/ARTICLE 580.html](http://www.takbarg.ir/ARTICLE_580.html).
- [5]-[http:// www.iranhiv.com/treatment](http://www.iranhiv.com/treatment).
- [6]-http://www.aftab.ir/articles/view/health_therapy/illness/c13c1173519504_hiv_p1.php
- [7]-<http://www.soton.ac.uk/~ceb/teaching/2005/206-8.htm>.
- [8]- www.iranhiv.com/HIV_cycle.html.
- [9]- [http://www.takbarg.ir/ARTICLE 1960.html](http://www.takbarg.ir/ARTICLE_1960.html).
- [10]- <http://www.avert.org/hiv-virus.htm>.
- [11]- <http://ray.tums.ac.ir/information/hiv.htm>.
- [12]- www.nature.com/nature/journal/v410/n6831/fig_tab/410995a0_F3.html.
- [13]- www.nature.com/nature/journal/v410/n6831/fig_tab/410995a0_F4.html.
- [14]- Sosnik A., Chieappetta D.A, Carcabose A.M, (2009), "*Drug delivery systems in HIV pharmacotherapy: what has been done and the challenges standing ahead*", J. controlled released, 138, pp 2-15.
- [15]- Jalali-Heravi M., Asadollahi-Baboli, (2008), "*QSAR analysis of platelet-derived growth inhibitors using GA-ANN and shuffling cross validation*", J. QSAR. Comb. Sci., 6, pp 750-757
- [16]- Jalali-Heravi M.; Parastar F., (2000), "*Use of artificial neural network in a QSAR study of anti-HIV activity for a large group of HEPT derivatives*", J. Chem. Inf. Comput. Sci., 40, pp 147-154.
- [17]- Bazoui H., Zahouily M., Sebti S., (2002), "*Structure- cytotoxicity relationships for a series of HEPT derivatives*", J. Mol. Model, 8, pp 1-7.
- [18]- Douali L., Villemin D., Cherqaoui D.; (2004), "*Exploring QSAR of non-nucleoside reverse transcriptase inhibitors by neural networks: TIBO derivatives*", Int. Mol. Sci., 5, pp 48-55.
- [19]- Roy K., Leonard J.T., (2005), "*QSAR by LFER model of cytotoxicity data of anti-HIV 5-phenyl-1-phenylamino-1H-imidazole derivatives using principal component factor analysis and genetic function approximation*", J. Bioorg. Med. Chem. 13, pp 2967–2973.

- [20]- Arab Chamjangali M., Beglari M., Bagherian G., (2007), "*Prediction of cytotoxicity data (CC₅₀) of anti-HIV 5-phenyl-1-phenylamino-1H-imidazole derivatives by artificial neural network trained with Levenberg-Marquardt algorithm*", J. Mol. Graph. Model, 26, pp 360–367.
- [21]-Zahouily M., Rakik J., Lazar M., Banlaoui M.A., Rayadh A., Komiha N., (2007), "*Exploring QSAR of non-nucleoside reverse transcriptase inhibitors by artificial neural networks: HEPT derivatives*", ARKIVOC, pp 245-256.
- [22]- Zarei K., Atabati M., (2009), "*QSAR study of anti-HIV activities against HIV-1 and some of their mutant strain for a group of HEPT derivatives*", J. Chinese Chem. Sci., 56, pp 206-213.
- [23]- Dessalew N., (2009), "*Investigation of the structural requirement for inhibiting HIV integrase: QSAR study*", J. Acta. Pharm., 59, pp 31-43.
- [24]- Arab Chamjangali M., (2009), "*Modeling of cytotoxicity data (CC₅₀) of anti-HIV 1-[5-chlorophenyl] sulfonyl-1H-pyrrole derivatives using calculated molecular descriptors and Levenberg- Marquardt artificial neural network*", J.Chem. Bio. Drug. Des., 73, pp 456-465.
- [25]- Almerico A.M, Tutone M., Lauria A., (2009), "*A QSAR study investigating the potential anti-HIV-I effect of some Acyclovir and Ganciclovir analogs*", ARKIVOC, pp 85-94.
- [26]- Jalali-Heravi M., Mani-Varnosfaderani A., (2009), "*QSAR modeling of 1-(3,3-diphenylpropyl)-piperidinyl amides as CCR₅ modulators using multivariate adaptive regression spline and Bayesian regularization genetic neural networks*", J. QSAR Comb. Sci., 9, pp 946-958.
- [27]-Patel A., Gupta L., Karthikeyan C., Trivedi P.(2010), "*QSAR studies on dihydro-alkoxy-benzyl-oxopyrimidines(DABOs) derivatives, as new series of potent, broad-spectrum non-nucleoside reverse transcriptase inhibitors*" J. Current Pharmaceutical Research , 1, pp 19-25
- [28]- Wold S. (1995) "*Chemometrics; what do we mean with it and what do we want from it*", J. Chemolab., 30, pp 109-115.
- [29]-<http://www.wordiq.com/definition/chemometrics>
- [30]-Manssnat D. L., Vandeginste B. G., Deming S. N. and Kaufman L. (1998) "*Chemometrics, A Text Book*", Elsevier, Amesterdom.
- [31]- Jurs P. C. (2005) "*Assessing the reliability of a QSAR model's predictions*", J.Mol Graph Model, 23(6), pp 503.

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

[33]- Brown C., Frazer T., (1868), "on the connection between chemical constitution and physiological action of salt of the ammonia bases derived from strychnine, Brucia, Thebaia, Codeia, Morphia and Nicotina", j. Trans Royal Soc Edinburg, 25, pp 151-203.

[34]- I. M. Kapetanovic, (2008), "Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach", J.Chem.Biol.Intract., 171(2), pp 165-176.

[35]- Hunsch C., Fujita T., (1964), "Rho-sigma-pi analysis. A method for the correlation of biological activity and chemical structure", J.Am. Chem. Soc., 86, pp 1616-1626.

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

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

[38]- HyperChem7.0 Toronto, Canada: HyperCube Inc, <http://www.hyper.com>.

[39]- لواین ای. ان (۱۳۸۷)، "شیمی کوآنتومی"، پارسا فرغ، مقاری ع، نجفی ب، جلد سوم چاپ اول، موسسه انتشارات فاطمی.

[40]. Atkins P.W., Friedman R.S. (1996), "Molecular Quantum Mechanics", 3rd Ed, Oxford University Press, New York, pp 307.

[41]- عرب چم جنگلی م. (۱۳۸۶) "پیش‌بینی فعالیت دارویی ضد ایدز (سیتوتوکسیسی) مشتقات ۵-فنیل-۱-فنیل آمینو-۱-H- / ایمیدازول به وسیله شبکه عصبی مصنوعی"، دانشگاه صنعتی شاهرود، گزارش طرح پژوهشی.

[42]- www.iasbs.ac.ir/chemistry/chemometrics/.../8th/qsar_introduction.ppt

[43]- Habibi A.; Danandeh M., (2007), "Prediction of acidity constant for substituted acetic acids in water using artificial neural networks", Indian journal of chemistry 46B, pp 478-487.

[44]- فردوسی م.ع. (۱۳۸۹) پایان نامه کارشناسی ارشد، "پیش‌بینی ثابت‌های هنری بعضی از ترکیبات آلی با استفاده از روش‌های خطی و غیرخطی QSPR"، دانشگاه صنعتی شاهرود.

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

[46] - مرکز انفورماتیک و توسعه، (۱۳۷۳)، "راهنمای نرم‌افزار SPSS/PC"، چاپ اول، انتشارات راهبرد.

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

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

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

[50]- منہاج م، (۱۳۸۷) "مبانی شبکه های عصبی (هوش محاسباتی)" جلد اول، چاپ پنجم، مرکز نشر دانشگاه صنعتی امیرکبیر، تهران.

[51]- هاگان. ت. ه.، دیموث. ه.، بیل م.، "طراحی شبکه های مصنوعی"، کیام. انتشارات کیان رایانه سبز.

[52]- کیام، (۱۳۸۷) "شبکه های عصبی در MATLAB"، چاپ دوم، انتشارات کیان رایانه سبز.

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

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

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

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

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

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

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

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

[61]- Goodarzi M., Deshpande S., Murugesan V., Katti S.B., Prabhakar Y.S., (2009), "Is feature selection essential for ANN modeling?", J. QSAR Com. Sci., 28, pp 1487-1499.

[62]- Mercader A.G., Pomilio A.B., (2010), "QSAR study of flavonoids and biflavonoids as influenza H1N1 virus neuraminidase inhibitors", Eur. J. Med. Chem., 45, pp1724-1730.

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

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

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

- [66]- Bell F.W., Cantrell A.S., Hogberg M., Jaskunas S.R., Johansson N.G., Jordan C.L., Kinnick M.D., Lind P., Morin Jr. J.M., Noreen R., Oberg B., Palkowitz H.A., Parrish C.A., Pranc P., Sahiberg C., Ternansky R.J., Vasileff R.T., L. Vrang, West S.J., Zhang H., Zhou X.X., (1995), "*Phenethylthiazolethiourea (PETT) compounds, a new class of HIV-1 reverse transcriptase inhibitors.1.Synthesis and basic structure–activity relationship studies of PETT analogs*", J. Med. Chem., 38, pp 4929 – 4936.
- [67]. Garkani-Nejad Z., Karlovits M., Demuth W., Stimpfl T., Vycudilik W., Jalali-Heravi M., Varmuza K., (2004), "*Prediction of gas chromatographic retention indices of a diverse set of toxicologically relevant compounds*", J. Chrom. A, 1028, pp 287-295.
- [68]- Zupan J.J., Gasteiger J., (1993), "*Neural networks for chemists an introduction*", VCH Publishers, Weinheim, Germany.
- [69]- Melo. E.B.Ferreira M.M., (2009), "*Nonequivalent effects of diverse LogP algorithm in three QSAR studies*", J.QSAR Comb.Sci., 28, pp1156-1165.
- [70]- Bansal R., Karthikeyan C., (2007),"QSAR analysis of som phthalimide analoges based inhibitors of HIV-Integrase", ARKIVOC, pp 66-81.
- [71]- Marjanovic M., Kralj M., Supek F., Frkanec L., Piantunida I., Smuc T., Tusek-Bozic L., (2007), "*Antitumor potential of crown ether: structure-activity relationship, cell cycle disturbances, and cell death studies of a series of Ionophores*", J. Med. Chem., 50, pp 1007-1018.
- [72]- Todeschini R., Consonni V., (2000), "*Handbook of molecular descriptors*", Wiley-VCH, Weinheim, Germany.
- [73]- Sun M., Chen J, Wei H., Yin S., Yang Y., Ji M., (2009),"Quantitative structure-activity relationship and classification analysis of diarylureas against vascular endothelial growth factor receptor-2 kinase using linear and non-linear methods", J. Bio. Drug Des., 73, pp 644-654.
- [74]- Yao X., Liu H., Zhang R., Liu M., Hu Z., Panaye A., Doucet J.P., Fan B., (2005), "*QSAR and classification study of 1,4dihydropyridine calcium channel antagonists based on least square support vector machines*", J. Mol. Pharm., 2, pp 348-358
- [75]- Hemmer M.C., Steinhauer V., Gasteiger J., (1999),"The prediction of the 3D structure of organic molecules from their infrared spectra", J. Vib. Spectrosc. ,19, pp 151-164.
- [76]- Schuur J.H., Selzer P., Gasteiger J., (1996), "*The coding of the three dimensional structure of molecules by molecular transform and its application to structure- spectra correlation and studies of biological activity*", J. Chem. Inf. Comput. Sci , 36, pp 334-344.
- [77]- 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*", J. Chem. Inf. Comput. Sci., 42, pp 682-692.
- [78]- P. Broto, G. Moreau, C. Vandycke, (1984), "*Molecular structures: Perception, autocorrelation descriptor and SAR studies*", Eur. J. Med. Chem., 19, pp 66–70.

- [79]- Douali L., Villemin D., Cherqaoui D.; (2003), “*Neural networks: Accurate nonlinear QSAR model for HEPT derivatives*”, J. Chem. Inf. Comput. Sci., 43, pp 1200-1207.
- [80]- Cabarello J., Garriga M., Fernandez M., (2006), “*2D autocorrelation modeling of negative inotropic activity of calcium entry blockers using Bayesian-regularized genetic neural networks*”, J. Bioorg. Med. Chem., 14, pp 3330-3340.
- [81]- chen Z., Zhang Y., Fu W., (2010), “*QSAR study of carboxylic acid derivatives as HIV-1 integrase inhibitors*”, Eur. J. Med. Chem., 45, pp 3970-3980.
- [82]- Spallarossa A., Cesarini S., Ranise A., Bruno O., Schenone S., Colla P.L., Collu G., Sanna G., Secci B., Loddo R., (2009), “*Novel modification in the series of O-(2-phthalimidoethyl)-N-substituted thiocarbamates and their ring-opened congeners as non-nucleoside HIV-1 reverse transcriptase inhibitors*”, Eur. J. Med. Chem., , 44, pp 1650–1663.
- [83]- Ranise A., Spallarossa A., Cesarini S., Bondavalli F., Schenone S., Bruno O., Menozzi G., Fossa P., Mosti L., Colla M.L., Sanna G., Murreduu M., Collu G., Loddo R., (2005), “*structure- based design, parallel synthesis, structure-activity relationship, and molecular modeling studies of thiocarbamates, new potent non-nucleoside isoesters of phenethylthiazolylthiourea derivatives*”, J. Med. Chem , 48 (11), pp.3858 – 3873.
- [84]- Moorthy H.N., Karthikeyan C., Trivedi P.,(2007) “*QSAR studies of cytotoxic acridine 5,7-diones: A comparative study using P-VSA descriptors and topological descriptors*”, Indian. J. Chem., 46B, pp 177-187.
- [85]- Todeschini R., Gromatica P., (1997), “*3D- modeling and prediction by WHIM descriptors. Part5. Theory development and chemical meaning of WHIM descriptors*”, J. Quant. Struct., Acta., Relat., 16, pp 113-119.
- [86]- Duchowicz P.R.; Fernandez M., Caballero J., Castro E.A., Fernandez F.M., (2006), “*QSAR for non- nucleoside inhibitors of HIV-1 reverse transcriptase*”, Bioorg. Med. Chem., 14, pp 5876-5889.
- [87]- Karelson M., Lobanov V.S., Katritzky A.R. (1996), “*Quantum-chemical descriptors in QSAR/ QSPR studies*”, J. Chem Rev., 96, pp 1027–1043.

Abstract

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are a class of anti retroviral drugs used to treat HIV infection. NNRTIs inhibit activity of reverse transcriptase (RT), an enzyme that controls the replications of the genetic material of HIV. In the first section, Quantitative structure – activity relationship (QSAR) models were constructed for predicting the Anti-HIV activity of a set of phenethyl thiazole thiourea (PETT) analogs, as potent HIV-1 reverse transcriptase inhibitors by calculated descriptors. The stepwise multiple linear regression method was used to select the most important descriptors. Then significant descriptors were used as input for QSAR model generation using multiple linear regression (MLR) and artificial neural network (ANN). Data set was randomly divided into training set and test set including 39 and 10 compounds, respectively. Training set was used in the selection the best MLR and ANN models by cross validation technique. The validation study of the MLR and ANN models was performed using test set and leave-one-out technique. The results obtained for prediction of anti HIV activity of the test set by MLR and ANN models showed squared correlation coefficients of 0.766 and 0.913 respectively.

In the second section, MLR and ANN methods were used for modeling and accurate prediction of Anti-HIV activities for a set of O-[2-(2- hydroxyl carbonyl benzamido) ethyl]-N-aryl thiocarbamate derivatives. The data set was randomly divided into training and test set containing 25 and 7 chemical, respectively. Training set was used in the selection the best MLR and ANN models by cross validation technique. The prediction ability of the proposed models was evaluated by test set and leave-one-out method. The squared correlation coefficients obtained for test set by MLR and ANN models were 0.920 and 0.979, respectively. The results obtained showed proper prediction power of the proposed models.

Keyword: QSAR, Anti -HIV, Artificial neural network, Multiple Linear Regression



Shahrood University of Technology

Faculty of Chemistry

**Quantitative Structure – Activity Relationship
Study Of Thiocarbamates Derivatives As New Class
Of Non – Nucleoside HIV - Inhibitors**

Motahareh Ashrafi

Supervisor:

Dr. M. Arab Chamjangali

Advisor:

Dr. Gh. Bagherian dehaghi

Date: Dec – 2010