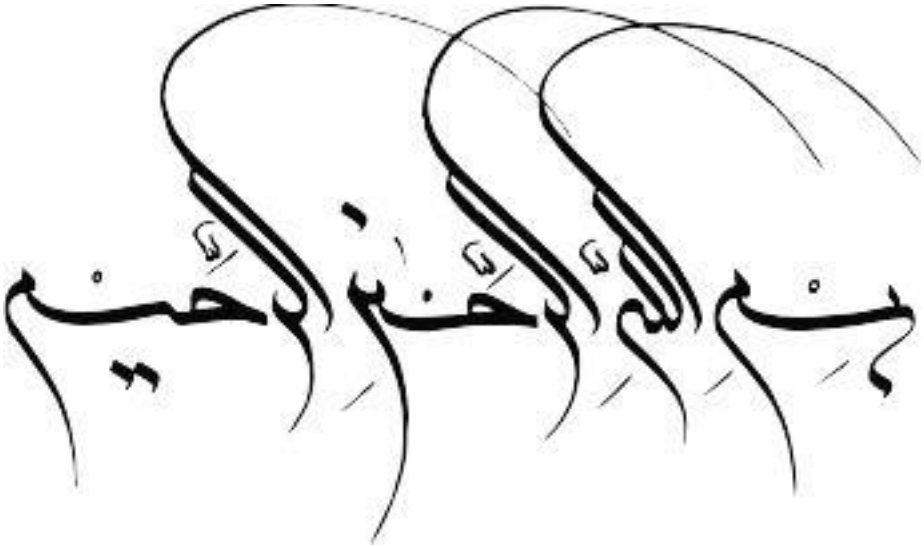

بسم الله الرحمن الرحيم

دانشکده: علوم پایه

گروه: شیمی

پیش‌بینی پارامتر حلالیت تعدادی حلال آلی با استفاده از روش‌های QSPR خطی و غیر خطی

دانشجو: هنگامه سلیمی

استاد راهنما:

دکتر ناصر گودرزی

استاد مشاور:

دکتر قدمعلی باقریان دهقی

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

تیر۱۳۸۹

تقدیم به هر آنکه در تاریکی‌ها

از روشنی به من آموخت

و با تشکر و قدردانی از:

پدر و مادر نازنینم و خواهر عزیزم ملیحه

اساتید بزرگوار و عزیزم در دانشگاه صنعتی شاهرود به ویژه استاد راهنمای گرانقدرم جناب آقای دکتر گودرزی که در طول انجام این پروژه مرا با نهایت دقت و دلسوزی راهنمایی کردند، استاد مشاور عزیزم جناب آقای دکتر باقریان دهقی و همچنین استاد دلسوز و مهربانم جناب آقای دکتر عرب چم جنگلی که در محضر علم و ادب ایشان نکته‌ها آموختم.

و دوستان مهربانم به ویژه هم اتاقی‌های عزیزم که آینه‌های پرتلألویی هستند از صداقت و صمیمیت و صفا

و تمام دوستان عزیزی که در انجام این پروژه مرا همیاری کردند به ویژه جناب آقای محمدرضایی که در مراحل مختلف این پروژه از هیچ کمکی فروگذار نکردند.

چکیده:

در این پروژه، مطالعات ارتباط کمّی ساختار - ویژگی بر روی پارامتر حلالیت ۷۰ حلال پلی‌مردر بخش اول، و شاخص بازداری ۷۰ ترکیبِ موجود در روغن‌های ضروری در بخش دوم صورت گرفت. پارامتر حلالیت خاصیت فیزیکو شیمیایی ذاتی یک ترکیب است. این خاصیت روش عددی ساده‌ای برای پیش‌بینی سریع خواص اساسی مولکول فراهم می‌کند. یکی از کاربردهای بسیار مهم پارامتر حلالیت، ارزیابی امتزاج‌پذیری مواد با هم می‌باشد.

روغن‌های ضروری گیاهی و ترکیبات آنها کاربرد گسترده‌ای در طب سنتی، طعم‌دار کردن غذا، و صنایع عطر سازی و دارویی دارند. رگرسیون مرحله‌ای برای انتخاب توصیف‌کننده‌ها به کار گرفته شد. مدل‌سازی از طریق دو روش خطی (رگرسیون خطی چندگانه؛ MLR) و غیر خطی (شبکه عصبی مصنوعی؛ ANN) انجام شد. اعتبار این مدل‌ها علاوه بر به کارگیری سری تست، توسط تکنیک‌های حذف مرحله‌ای تک تک و Y- تصادفی بررسی شد. هر دو روش خطی و غیرخطی توانایی پیش‌بینی خوبی دارند، اما مدل ANN در هر دو مورد نتایج صحیح‌تری دارد. در بخش اول این تحقیق، ضریب همبستگی سری تست توسط مدل‌های MLR و ANN به ترتیب ۰/۹۶۸ و ۰/۹۷۵ بودند. در بخش دوم، ضریب همبستگی سری تست توسط مدل‌های MLR و ANN به ترتیب ۰/۹۴۹ و ۰/۹۶۴ بودند.

کلیدواژه‌ها: ارتباط کمّی ساختار - ویژگی، رگرسیون خطی چندگانه، شبکه عصبی مصنوعی، پارامترحلالیت، شاخص بازداری

نتایج حاصل از این پایان‌نامه در پوستری تحت عنوان:

“Prediction of solubility of some polymers in different solvents using Linear and Nonlinear QSPR Methods”

در شانزدهمین سمینار شیمی تجزیه ایران در دانشگاه همدان

ودر پوستری تحت عنوان:

“Prediction of Retention indices of some essential oils using Linear and Nonlinear QSPR Methods”

در دومین سمینار کمومتریکس ایران در ارومیه پذیرفته شد.

فهرست مطالب

فهرست اشکال

فهرست جداول

فصل اول

مقدمه

۱-۱- تعریف تحقیق

کمومتریکس شاخه‌ای از شیمی است که ابزارهای ریاضی و آمار را برای سامان بخشیدن به داده‌های شیمیایی و به‌دست آوردن اطلاعات بیشتر از آنها به کار می‌گیرد. از قابلیت‌های کمومتریکس می‌توان به نگرش چند متغیره به مسئله، ساخت و ارزیابی مدل‌هایی با قابلیت پیش‌بینی، مقایسه نتایج به‌دست آمده از روش‌های مختلف و تعریف و استفاده از شاخص‌هایی که قادرند کیفیت اطلاعات استخراج شده و مدل‌های به‌دست آمده را بسنجند، اشاره کرد[۲،۱]. پیشرفت کامپیوترها و منابع نرم افزاری در سال‌های اخیر باعث رشد روزافزون استفاده از این قابلیت‌ها در شاخه‌های مختلف شیمی و به ویژه شیمی تجزیه شده است.

یکی از شاخه‌های بسیار مهم کمومتریکس که در این پایان‌نامه به کار گرفته می‌شود، ارتباط کمّی ساختار - ویژگی[۱] است. این تکنیک در این پایان‌نامه برای پیش‌بینی پارامتر حلالیت[۲] تعدادی حلال پلی‌مر در بخش اول، و پیش‌بینی شاخص‌های بازداری[۳] تعدادی از ترکیبات تشکیل‌دهنده روغن‌های ضروری[۴] در بخش دوم به کار گرفته شده است. از جمله روش‌هایی که به منظور مطالعه ارتباط خطی ساختار ویژگی مورد استفاده قرار می‌گیرند می‌توان به روش‌های رگرسیون خطی چندگانه[۵]، رگرسیون اجزای اصلی[۶] و روش حداقل مربعات جزئی[۷] اشاره کرد. روش‌های دیگری مانند شبکه‌های عصبی مصنوعی[۸] نیز ارتباط غیر خطی میان ساختار و ویژگی‌های ترکیبات را مورد مطالعه قرار می‌دهند. در

۱.Quantitative Structure-Property Relationship (QSPR)

۲.Solubility Parameter

۳.Retention Indices

۴. Essential Oils (Eos)

۵.Multiple Linear Regression (MLR)

۶. Principal Components Regression

۷. Partial Least Square

۸.Artifitial Neural Network (ANN)

تحقیق حاضر مطالعه ارتباط کمّی ساختار - ویژگی ترکیبات مـوردنظر بـا اسـتفاده از روش‌هـای خطـی و غیرخطی انجام گرفته و نتایج با هم مقایسه شده اند. در بخش اول رگرسیون خطی چندگانـه بـه منظـور ایجاد مدل مناسبی برای بیان ارتباط خطی بین پارامتر حلالیت و ساختار ۷۰ مولکول حلال به‌کار گرفتـه شد. در ادامه، شبکه عصبی مصنوعی به منظور بررسی روابط غیرخطی بین ساختار این ترکیبات و خواص مورد نظر به کار رفت. در بخش دوم این پایان‌نامه نیز به کمک این روش‌ها ارتباط بین ساختار ۷۰ ترکیب از روغن‌های ضروری و شاخص بازداری آنها بررسی شد. مدل‌های به دست آمده با روش‌های ارزیابی[۱] مورد بررسی قرار گرفته و پارامترهای آماری مربوطه گزارش شد.

در ادامه پارامتر حلالیت و روغن‌های ضروری و ترکیبات آنها معرفی می‌شوند.

۱-۱-۱- پارامتر حلالیت

فرایند تفکیک یک پلی‌مر آمورف در یک حلال تابع انرژی آزاد اختلاط می‌باشد (معادله ۱-۱).

$$\Delta G_m = \Delta H_m - T\Delta S_m$$ معادله ۱-۱

در معادله فوق ΔG_m تغییرات انرژی آزاد گیبس در حین فرایند اختلاط، ΔH_m ، تغییرات انتـالپی در حین فرایند اختلاط و ΔS_m تغییرات انتروپی در حین فرایند اختلاط می‌باشد. منفی بـودن انـرژی آزاد اختلاط نشان‌دهنده خود به خودی بودن فرایند اختلاط می‌باشد. در غیر این صورت در فرایند اختلاط، دو یا سه فاز تشکیل خواهد شد. از آنجاییکه تفکیک یک پلی‌مر با وزن ملکولی بالا همیشه باعث تغییرات کم یا متوسطی در انتروپی می‌گردد، عبارت مربوط به انتالپی (علامت و بزرگی ΔH_m) عامل تعیین کننده‌ی علامت تغییرات انرژی آزاد گیبس می‌باشد.

۱.validation techniques

هیلدبراند[۱]، اسکات[۲] و اسکاتچارد[۳] پیشنهاد کردند که :

$$\Delta H_m = V\left(\left(\frac{\Delta E_1^V}{V_1}\right)^{1/2} - \left(\frac{\Delta E_2^V}{V_2}\right)^{1/2}\right)^2 \emptyset_1\emptyset_2$$ معادله ۱-۲

در معادله فوق Vحجم مخلوط، ΔE^V_i انرژی تبخیر گونه i، V_i حجم مولی گونه i، و $\emptyset_i$ کسر مولی گونه **i** در مخلوط است. به عبارت دیگر، ΔE^V_i ، تغییرات انرژی در طول فرایند تبخیر ایزوترمال مایع اشباع به حالت گاز ایده‌آل در حجم بی‌نهایت می‌باشد. انرژی تبخیر به ازای هر سانتیمتر مکعب چگالی انرژی آزاد چسبندگی[۴] نامیده می‌شود.

پارامتر حلالیت به عنوان مجذور چگالی انرژی آزاد چسبندگی تعریف شده است و نیروی جاذبه بین ملکول‌های ماده را توصیف می‌کند (معادله ۱-۳).

$$\delta_i = \left(\frac{\Delta E_i^V}{V_i}\right)^{1/2}$$ معادله ۱-۳

دیمانسیون δ_i به صورت زیر است:

$(cal/cm^3)^{1/2} = (4.187J/10^{-6}\ m^3)^{1/2}$

پارامتر حلالیت را می‌توان به صورت "فشار داخلی حلال" نیز تفسیر کرد. δ_i در بعضی نوشتارها پارامتر هیلدبراند و همینطور پارامتر چسبندگی نیز نامیده می‌شود و نه تنها به امتزاج‌پذیری اجزای تشکیل دهنده بلکه به تعداد زیادی از خصوصیات فیزیکی و شیمیایی ترکیب مرتبط است. پارامتر حلالیت یک مخلوط معمولاً به صورت مجموع حاصلضرب‌های پارامترهای حلالیت اجزای تشکیل‌دهنده مخلوط در جزء مولی آنها، در نظر گرفته می‌شود (معادله ۱-۴).

$$\delta_{mixture} = \sum \delta_i\, \emptyset_i$$ معادله ۱-۴

---

۱. Hildebrand

۲.Scott

۳.Scotchard

۴.Cohesive Energy Density (CED)

رابطه بین δ_i و معادله ΔH_m و... را می‌توان بازنویسی کرد و گرمای اختلاط به ازای هر واحد حجمی برای یک مخلوط دوتایی را به دست آورد (معادله ۱-۵):

$$\frac{\Delta H_m}{V} = (\delta_1 - \delta_2)^2 \emptyset_1 \emptyset_2$$

برای امتزاج پذیر بودن پلی‌مر- حلال باید در معادله ۱-۱، گرمای اختلاط از عبارت انتروپی کوچکتر باشد. وقتی که δ_1 با δ_2 برابر باشد، انرژی آزاد اختلاط برای محلول‌های معمولی همواره از صفر کمتر خواهد بود و دو جزء با هر نسبتی در هم حل می‌شوند. در کل، برای امتزاج پذیری در تمام بازه حجم‌های جزئی، تفاوت پارامترهای حلالیت (δ_1وδ_2) باید کوچک باشد.

معمولاً پارامترهای حلالیت حلال‌ها را می‌توان مستقیماً از طریق اندازه‌گیری انرژی تبخیر تعیین کرد. اما پارامتر حلالیت پلی‌مرها را تنها به روش‌های غیر مستقیم می‌توان به دست آورد. این پارامتر تحت تاثیر ساختار پلی‌مرهای مورد نظر (برای مثال تعداد شاخه‌ها و نحوه توزیع آنها و همینطور گروه‌های استخلاف شده در طول زنجیره کربنی) قرار می‌گیرد[۳].

از کاربردهای پارامتر حلالیت می‌توان به انتخاب حلال‌های مناسب برای رزین‌های پوششی، اصلاح خواص پوشش‌ها، پیش‌بینی تورم ناشی از حلال در الاستومرهای اصلاح شده، تخمین فشار بخار حلال در محلول‌های پلی‌مری و همینطور پیش‌بینی تعادل فاز سیستم‌های پلی‌مر- پلی‌مری و حلال‌های چندجزئی اشاره کرد[۴].

۱-۱-۲- روغن‌های ضروری

روغن‌های ضروری (که روغن‌های اتری یا فرّار نیز نامیده می‌شوند) مایعات روغنی معطری هستند که از مواد خام گیاهی (گل‌ها، جوانه‌ها، دانه‌ها، برگ‌ها، چوب، میوه و ریشه‌ها) به دست می‌آیند. این روغن‌ها را می‌توان به روش‌های تخمیر، استخراج و همینطور تقطیر با بخار آب (که در تولید تجاری از همه رایج‌تر است) به دست آورد. خواص آنتی اکسیدانی، ضدمیکروبی، ضدویروسی و همینطور ضدانگلی

این ترکیبات شناخته شده است. با پیشرفت بهداشت عمومی و تلاش صنایع غذایی برای سالم‌سازی هرچه بیشتر فرآورده‌های غذایی، استفاده از این ترکیبات به عنوان جایگزینی برای مواد افزودنی سنتزی و همینطور نمک‌ها (که می‌توانند احتمال بیماری‌های قلبی را افزایش دهند) پیشنهاد شده است[۵].

پولیتئو[۱] و همکارانش در سال ۲۰۰۶ از کروماتوگرافی گازی - طیف سنجی جرمی برای تعیین کمّی ترکیبات موجود در ۱۲ گیاه حاوی روغن ضروری استفاده و شاخص بازداری مربوط به آنها را گزارش کردند[۶]. هدف نهایی آنها بررسی خواص آنتی‌اکسیدانی این ترکیبات بود. در تحقیق حاضر این نتایج به عنوان سری داده‌ها مورد استفاده قرار گرفته و کارامدی روش‌های QSPR برای پیش‌بینی شاخص بازداری این ترکیبات بررسی شده است.

۱-۱-۲-۱- شاخص بازداری

شاخص بازداری معیاری کمّی است که میزان بازداری نسبی هر جزء از اجزاء نمونه را نسبت به آلکان‌های نرمال بر روی یک فاز ساکن معین و در یک دمای مشخص نشان می‌دهد. شاخص بازداری، متغیرهای مختلف سیستم کروماتوگرافی را نرمالیزه می‌کند، به گونه‌ای که امکان مقایسه نتایج سیستم‌های مختلف فراهم می‌شود. مقدار شاخص بازداری از معادله ۱-۶ محاسبه می‌شود.

$$I = 100Y(Z-Y)\left[\frac{\log t'_r X - \log t'_r Y}{\log t'_r Z - \log t'_r Y}\right]$$ معادله ۱-۶

در معادله بالا t'_rY زمان بازداری اصلاح شده برای آلکان نرمال که قبل از ترکیب X از ستون خارج می‌شود، t'_rX زمان بازداری اصلاح شده برای ترکیب مورد نظر، t'_rZ زمان بازداری اصلاح شده برای آلکان نرمال که بعد از ترکیب X از ستون خارج می‌شود، Y تعداد کربن‌های آلکان نرمال که قبل از ترکیب مورد نظر از ستون خارج می‌شود و Z نیز تعداد کربن‌های آلکان نرمال که بعد از ترکیب مورد نظر از ستون

۱. Politeo

خارج می‌شود است. به عنوان مثال، پیک ترکیبی با شاخص بازداری ۱۲۵۰ در کروماتوگرام دقیقاً در میان پیک‌های $C_{12}H_{26}$ (با شاخص بازداری ۱۲۰۰) و $C_{13}H_{28}$ (با شاخص بازداری ۱۳۰۰) قرار می‌گیرد. شاخص بازداری زمانی مفید است که پیش‌بینی ترتیب خروج نسبی اجزاء نمونه از ستون مورد نظر باشد. اگر از شاخص بازداری با چنین منظوری استفاده شود، باید در نظر داشت که شاخص بازداری در یک شرایط دمایی معین و برای یک ستون خاص قابل کاربرد است. از شاخص بازداری، برای مقایسه بازداری و گزینش‌پذیری ستون‌های مشابه (ستون‌هایی که از لحاظ جنس و ابعاد لوله موئین، نوع و ضخامت فاز ساکن و سایر موارد کاملاً مشابه اند) نیز می‌توان استفاده کرد[۷].

۱-۲- ضرورت تحقیق

در سال‌های اخیر پیشرفت‌های زیادی در جهت آسان‌تر شدن، دقیق‌تر شدن و سازگار بودن بیشتر روش‌های آزمایشگاهی با محیط زیست صورت گرفته است. با این‌حال این روش‌ها هنوز محدودیت‌هایی مانند عدم توانایی در اندازه‌گیری همه ترکیبات، هزینه بالا و در دسترس نبودن امکانات مورد نظر دارند. استفاده از روش‌های نظری می‌تواند در رفع این محدودیت‌ها مفید واقع شود. به کارگیری این روش‌ها می‌تواند علاوه بر پیش‌بینی خواص مورد نظر، به هدفمند و جهت دارتر ساختن آزمایش‌های تجربی و همینطور توضیح پارامترهای موثر بر نتایج این آزمایش‌ها و مکانیسم‌های درگیر کمک کند.

در این تحقیق سعی شده با به‌کارگیری روش‌های QSPR مدلی برای پیش‌بینی پارامتر حلالیت تعدادی حلال پلی‌مر در بخش اول، و نیز مدلی برای پیش‌بینی شاخص بازداری تعدادی از ترکیبات موجود در روغن‌های ضروری در بخش دوم طراحی شود. هنگامی که رابطه معتبری به دست آمد، امکان پیش‌بینی این خواص برای ترکیباتی با ساختار مشابه وجود خواهد داشت. در تحقیق حاضر تلاش شده تا مقایسه‌ای بین پاسخ سری داده‌ها به مدل خطی و مدل غیرخطی نیز صورت پذیرد.

۱-۳- پیشینه تحقیق

تانتیشایاکول[۱] و همکارانش در سال ۲۰۰۶ رگرسیون حداقل مربعات جزئی را برای پیش‌بینی پارامتر حلالیت ۵۱ ترکیب به کار گرفتند. R^2 گزارش شده برای کل سری داده‌ها با این روش ۰/۸۵۳ بود[۸].

در سال ۲۰۰۷، قاراقیزی روش ژنتیک الگوریتم- رگرسیون خطی چندگانه[۲] و نیز روش شبکه عصبی تخمین تابع تعمیم داده شده[۳] را برای مدل‌سازی پارامتر حلالیت ۱۲۲۸ ترکیب مختلف مورد استفاده قرار داد. مدل نهایی به دست آمده توسط این محقق، شامل ۷ توصیف کننده بود. مجذور ضریب همبستگی به دست آمده برای کل سری داده‌ها، برای روش GA-MLR ۰/۸۲ و برای روش GRNN ۰/۹۸ بود. این محقق به جای به کارگیری نسبت معمول شبکه عصبی یعنی ۶۰:۲۰:۲۰ برای سری آموزش، ارزیابی و تست فقط ۳۶ ترکیب را به عنوان هرکدام از سری‌های ارزیابی و تست مورد بررسی قرار داد. مجذور ضریب همبستگی به دست آمده برای سری تست توسط روش GRNN ۰/۸۲ بود[۹].

در سال ۲۰۰۸ نیز استفانیس[۴] و پانایوتو[۵] یک روش سهم گروه[۶] را برای پیش‌بینی پارامتر حلالیت تعدادی ترکیب آلی به کار گرفتند[۱۰].

ارتباطات کمی ساختار – ویژگی تا کنون در موارد گسترده‌ای برای توضیح مکانیسم‌های جداسازی و پیش‌بینی رفتار بازداری در کروماتوگرافی به کار گرفته‌شده است.

۱.Tantishaiyakul

۲.Genetic Algorithm-Multipel Linear Regression

۳.Generalized Regression Neural Network

۴.Stefanis

۵.Panayiotou

۶.Group contributions

لیائو[1] و همکارانش در سال ۲۰۰۷، روش MLR را برای پیش‌بینی زمان بازداری ۶۹ ترکیب روغن ضروری موجود در گونه خاصی از گلها به کار بردند[۱۱]. ضریب همبستگی ۰/۹۴ در این روش برای سری داده‌ها گزارش شده است.

در سال ۲۰۰۸، ریاحی و همکارانش روش GA-MLR را برای مدل‌سازی شاخص بازداری ۴۳ ترکیب روغن ضروری به کار گرفتند. R^2 گزارش شده برای سری آموزش متشکل از ۳۵ ترکیب، ۰/۹۷۷ و برای بقیه ترکیبات که به عنوان سری تست در نظر گرفته شدند، ۰/۹۷۸ بود[۱۲].

همین گروه تحقیقاتی در سال ۲۰۰۸، روش SVM[2] را برای پیش‌بینی شاخص بازداری صد ترکیب موجود در روغن‌های ضروری، با روش‌های خطی مورد مقایسه قرار دادند و آن را به عنوان روش بهتری نسبت به MLR گزارش کردند[۱۳].

در تحقیق حاضر، ارتباطات کمّی ساختار – ویژگی برای پیش‌بینی پارامتر حلالیت ۷۰ ترکیب حلال پلی‌مر و نیز شاخص بازداری ۷۰ ترکیب موجود در روغن‌های ضروری به کار گرفته شده است. مقایسه MLR و ANN در مورد این سری داده‌ها تا کنون انجام نشده و این کار به عنوان تحقیق جدیدی مطرح می‌باشد. در این پایان‌نامه رگرسیون خطی چندگانه و شبکه عصبی به دلیل توانایی بسیار بالا در یافتن ارتباطات خطی و غیر خطی ساختار – ویژگی برای ترکیبات مشابه و همینطور سهولت نسبی مورد استفاده قرار گرفته‌اند.

---

۱. Liao

۲.Supported Vector Machine (SVM)

فصل دوم

کمومتریکس

۲-۱- کمومتریکس

عبارت کمومتریکس توسط اسوانت ولد[۱] و بروس کووالسکی[۲] در اوایل ۱۹۷۰ تعریف شد. قبلاً عباراتی مانند بیومتریکس[۳] و اکونومتریکس[۴] نیز در زمینه‌های علوم زیستی و همینطور اقتصاد معرفی شده اند. پس از آن بود که انجمن بین‌المللی کمومتریکس تأسیس شد. از آن زمان تا کنون پیشرفت‌های زیادی در این زمینه به وجود آمده و اکنون کمومتریکس در شاخه های مختلف شیمی، به ویژه شیمی تجزیه مورد استفاده قرار می‌گیرد. برخی از کاربردهای آن عبارتند از: کنترل فرآیندها، تجزیه و تحلیل و شناخت الگوها، پردازش علائم، بهینه کردن شرایط.

یکی از تعاریفی که از کمومتریکس وجود دارد، "سازماندهی شیمیایی با استفاده از روش‌های ریاضی و آماری برای (الف) طراحی و انتخاب بهترین فرایندهای محاسباتی و آزمایشگاهی و (ب) فراهم نمودن بیشترین اطلاعات شیمیایی با بررسی داده‌های در دسترس" می‌باشد[۱۴].

هاوری[۵] و هیرش[۶] در اوایل سال ۱۹۸۰ پیشرفت‌های کمومتریکس را در مراحل مختلف رده‌بندی کردند. اولین مرحله قبل از ۱۹۷۰ است که در آن تعدادی از روش‌های ریاضی توسعه یافتند و در زمینه‌های مختلف ریاضی، علم رفتار و مهندسی به‌کار گرفته شدند. در این محدوده زمانی، شیمیدانان خود را به روش‌های آنالیز داده‌ها، یعنی محاسبه پارامترهای آماری مانند میانگین، انحراف استاندارد و حدود اطمینان محدود کرده بودند. هاوری و هیرش تحقیقات خود را به طور ویژه روی همبستگی تعداد زیادی از داده های شیمیایی و ویژگی‌های مولکولی مربوطه، متمرکز کردند. این اقدامات اولیه اساس یکی

۱.Svant Wold
۲.Bruce R.Kowalski
۳.Biometrics
۴.Econometrics
۵.Howery
۶.Hirsch

از شاخه های مهم کمومتریکس یعنی ارتباط کمّی ساختار- فعالیت (QSAR) را فراهم کرد. دومین مرحله تکامل کمومتریکس در سال ۱۹۷۰ واقع شد، هنگامی که کمـومتریکس توجـه شـیمیدانان، بـه خصـوص شیمیدانان تجزیه را که به این ترتیب روش‌های بیشتری بـرای آنـالیز داده‌هـای در دسـترس و همچنـین روش‌های نوینی برای به دست آوردن اطلاعات پیش روی خود می‌دیدند، به خود جلب کرد. مرحله بعدی تکامل کمومتریکس که ابتدا توسط هاوری و هیرش و سپس توسط براون[۱] پیش‌بینـی شـده بـود از اوایـل ۱۹۸۰ آغاز شد. هم‌اکنون کمومتریکس ابزاری قوی و کارامد برای شیمیدانان به شمار رفته و به عنوان یک زیرشاخه رشته شیمی در بسیاری از دانشگاه‌های آمریکا و اروپا و بعضی دانشگاه‌های چین و سایر کشورها تدریس می‌شود. پیشرفت‌ها در علوم کامپیوتر، فناوری اطلاعات، آمار و ریاضی کاربردی عناصر جدیـدی را در کمومتریکس معرفی کرده است[۱۵،۱۶].

۲-۱-۱- کمومتریکس در ایران

اولین مقاله منتشر شده در زمینه کمومتریکس در ایران به سال ۱۹۹۸ برمی‌گـردد[۱۷]. بـا رشـد سریع کمومتریکس در جهان، استفاده‌ی شیمی‌تجزیه‌دانـان ایرانـی از کمـومتریکس نیـز در تحقیقاتشـان افزایش یافت. تا سال ۲۰۰۵ حدود ۲۰۰ مقاله علمی توسط محققین کمومتریکس در ایران منتشر و بیش از ۸ گروه تحقیقاتی فعال در این زمینه مشغول فعالیت می‌باشند. انجمن کمومتریکس ایران نیز به عنـوان یکی از شاخه‌های فعال انجمن شیمی ایران در سال ۲۰۰۰ تأسیس شد و همچنین سـمینارهای داخلـی و بین المللی در زمینه کمومتریکس در ایران برگزار می‌گردد. زمینه‌هایی که تاکنون بیشتر در ایـران مـورد توجه قرار گرفته‌اند مطالعات (QSPR/QSAR[۲])، کالیبراسیون چندمتغیره[۳] و هوش مصنوعی[۴] هستند[۱۸].

۱.Brown

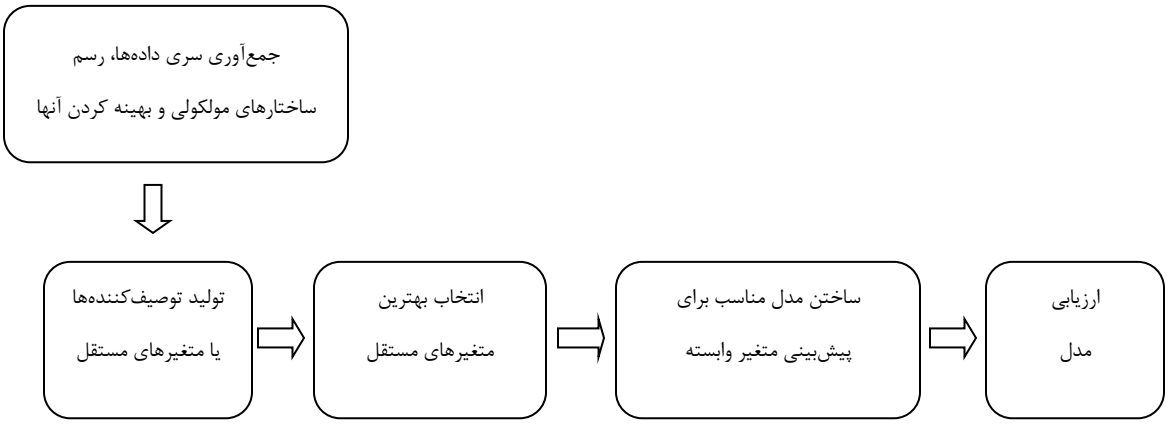
۲. Quantitative Structure - Activity Relationship (QSAR)

۳.Multivariate Calibration (MVC)

۴.Artificial Intelligence (AI)

۲-۲- ارتباط کمّی ساختار - ویژگی

وابستگی خواص شیمیایی، فیزیکی و همینطور زیستی ترکیبات به فرمول ساختاری آنها یکی از اصولی است که مدت‌هاست در شیمی مورد پذیرش قرار گرفته است. البته به دلایل بسیاری از جمله محدودیت‌های فیزیکی فناوری کامپیوتر و نداشتن پایه نظری به اندازه کافی قوی، محاسبه‌ی مستقیم خواص کمّی ترکیبات از اصول اولیه کار آسانی نیست. به همین دلیل توسعه روش‌هایی برای یافتن ارتباطات کمّی ریاضی بین ساختار یک مولکول و خواص قابل مشاهده آن اهمیت زیادی دارد[۱۹]. به این ترتیب، به جای ارتباط مستقیم ساختار و ویژگی، مراحل زیر انجام می‌شود (شکل ۲-۱):



شکل ۲-۱- شمای کلی مراحل به کار گرفته شده در یک فرایند QSPR

با توجه به نتایج چنین پیش‌بینی‌هایی می‌توان آزمایش‌های شیمیایی را در جهتی انجام داد که احتمال موفقیت بیشتر است.

۲-۲-۱- انتخاب سری داده‌ها و بهینه‌سازی ساختار هندسی

سری داده[۱]ها مجموعه‌ای از موارد توصیف شده توسط یک یا تعداد بیشتری از متغیرها است. همان‌گونه که قبلاً گفته شد متغیرهای سری داده‌ها می‌توانند از طریق نقششان در مدل‌ها به صورت متغیرهای مستقل و وابسته تمایز داده شوند. اولین مرحله جمع‌آوری و انتخاب یک سری ملکولی است که مقادیر تجربی خاصیت مورد نظر آنها (مثلاً پارامتر حلالیت یا شاخص بازداری) در دسترس باشد. هرچه سری داده‌ها بزرگتر و متنوع‌تر باشد مدل حاصل از آن معتبرتر بوده و قدرت پیش‌بینی بالاتری خواهد داشت. در ضمن لازم است اطلاعات راجع به سری داده‌ها در شرایط یکسان به دست آمده باشد؛ در غیر این صورت باید داده‌ها را دسته‌بندی کرد.

کمومتریکس از روش‌های مفیدی که قادر به استخراج اطلاعات موجود در داده هستند، استفاده می‌کند[۲۰]. به این منظور از شیمی محاسباتی[۲] استفاده می‌شود. شیمی محاسباتی ساختارهای مولکولی را به صورت پارامترهای عددی معرفی و رفتار آنها را با معادلات کوانتومی و فیزیک کلاسیک شبیه‌سازی می‌نماید. این امر به دانشمندان امکان می‌دهد که بتوانند از این طریق به اطلاعات مولکول از جمله ساختار هندسی، انرژی‌ها و خواص الکترونیکی و اثرات حلال دست پیدا کنند. در سال‌های اخیر روش‌های محاسباتی در بین شیمیدانان بسیار رواج یافته است.در این روش‌ها می‌توان به راحتی محاسبات را انجام داد، بدون اینکه از اصول اولیه و روش محاسبه آگاهی دقیق داشت. روش‌های محاسباتی در شیمی به چندین دسته تقسیم می‌شوند. روشی که در این تحقیق به کار گرفته‌شده، روش AM1[۳] است که جزء

۱.Data set

۲.Computational chemistry

۳.Austin Methods

روش‌های محاسباتی نیمه‌تجربی[1] می‌باشد. در روش‌های نیمه‌تجربی، که در برنامه‌هایی نظیر HyperChem وارد شده‌اند، محاسبات بر اساس مکانیک کوانتومی صورت می‌گیرد.

۲-۲-۲- به دست آوردن توصیف‌کننده‌ها

اگرچه چندین پارامتر مولکولی از زمان شروع شیمی کوانتومی و نظریه گراف تعریف شدند اما اصطلاح "توصیف کننده مولکولی" با پیشرفت مدل‌های ارتباط دهنده ساختار- خاصیت مشهور شد. تعداد پلات[2][۲۱] و شاخص وینر[3][۲۲] که درسال ۱۹۴۷ تعریف شدند، در بعضی مراجع به عنوان اولین توصیف‌کننده‌های مولکولی در نظر گرفته می‌شوند. توصیف‌کننده‌های مولکولی مقادیر عددی هستند که ساختار یا شکل مولکول را توصیف می‌کنند و به پیش‌بینی فعالیت و خصوصیات مولکول‌ها در آزمایش‌های پیچیده کمک می‌کنند. به عبارت دیگر توصیف‌کننده‌های مولکولی نتیجه نهایی یک فرآیند منطقی و ریاضی هستند که اطلاعات شیمیایی مربوط به ساختار یک مولکول را به اعداد تبدیل می‌کنند. این اعداد می‌توانند برای تفسیر خواص مولکولی استفاده شوند و یا برای پیش‌بینی تعدادی از ویژگی‌های مولکولی در یک مدل شرکت کنند.

برخی از ویژگی‌های لازم برای یک توصیف کننده مولکولی مناسب عبارتند از:[۲۳]

- ساده بودن
- مستقل بودن
- تفسیر ساختار مولکول (غنی بودن از نظر اطلاعات)
- عدم همبستگی با سایر توصیف کننده‌ها

۱.Semi-empirical methods

۲.Platt Number

۳.Wiener Index

- تغییر منظم با تغییر تدریجی در ساختارها
- وابستگی صحیح به اندازه مولکول
- تمایز بین ایزومرهای مختلف مولکول

توصیف‌کننده‌های مولکولی به دو دسته اصلی تقسیم می‌شوند: توصیف‌کننده‌های حاصل از اندازه‌گیری‌های تجربی مانند قطبش‌پذیری، ممان دوقطبی و توصیف‌کننده‌های مولکولی نظری که از ساختار نمادین مولکول مشتق شده و می‌توانند به دسته های بیشتری مطابق با انواع مختلفی از نمایش مولکولی[۱] یا بر اساس ابعاد توصیف‌کننده مولکولی تقسیم شوند[۲۴]. نمایش مولکولی روشی است که با آن یک مولکول از طریق فرآیند قراردادی و قواعد اختصار به صورت نمادین ارائه می شود.

۲-۲-۲-۱- توصیف‌کننده‌های صفر بعدی

ساده‌ترین نمایش مولکولی فرمول شیمیایی است. این نمایش هیچ‌گونه اطلاعاتی از شکل فضایی مولکول ارائه نمی‌کند و از این رو توصیف کننده‌های مولکولی به دست آمده از فرمول‌های شیمیایی، توصیف‌کننده‌های صفر بعدی نامیده می‌شوند. این دسته از توصیف‌کننده‌ها تعداد و نوع اتم‌ها، جرم مولکولی و سایر خصوصیات اتمی را تعیین می‌کنند (مثلا مجموع حجم‌های اتمی واندروالس). همچنین این نوع توصیف‌کننده‌ها، به توصیف کننده‌های ساختاری نیز معروفند که علاوه بر موارد فوق شامل توصیف‌کننده‌های مربوط به نوع پیوندها و حضور حلقه‌ها در مولکول نیز می‌شوند. تعداد کل اتم‌ها در مولکول، تعداد یک عنصر شیمیایی خاص(C,N,O,H,F,…) در مولکول، تعداد گروه‌های عاملی خاص در مولکول، تعداد پیوندهای ساده، دوگانه، سه گانه، آروماتیک و... در مولکول، تعداد کل حلقه‌ها بر اساس

۱.Molecular representation

تعداد اتم‌های آنها (حلقه های شش تایی، پنج تایی و ...) در مولکول، وزن مولکولی و متوسط وزن اتمی مثال‌هایی از توصیف‌کننده‌های ساختاری می‌باشند.

۲-۲-۲-۲- توصیف‌کننده‌های یک بعدی

نمایش فهرست ویژگی‌های زیرساختاری[۱] می تواند به عنوان نمایشی یک بعدی از مولکول در نظر گرفته شود و شامل لیستی از اجزای ساختاری از یک مولکول (شامل اجزا، گروه‌های عاملی یا استخلافات مد نظر در یک مولکول) است. توصیف کننده های گروه‌های عاملی، قطعات اتم مرکزی، توصیف کننده‌های تجربی و خصوصیات مولکولی زیر گروه های این دسته می باشند.

۲-۲-۲-۳- توصیف‌کننده‌های دوبعدی

نمایش دو بعدی از مولکول چگونگی اتصال اتم‌ها در مولکول بر حسب حضور و ماهیت باندهای شیمیایی می‌باشد. مثلاً دیدگاه بر پایه گراف مولکولی یک نمایش دوبعدی را ارائه می‌کند و معمولاً به عنوان نمایش توپولوژیکی شناخته می شود. توصیف‌کننده‌های مولکولی به دست آمده از الگوریتم‌های به‌کار رفته برای یک نمایش توپولوژیکی، توصیف‌کننده‌های دو بعدی نامیده می‌شوند. این دسته از توصیف‌کننده‌ها شامل زیرگروه‌های توپولوژیکی، شمارش‌های مسیرهای مولکولی[۲]، توصیف‌کننده‌های BCUT[۳]، ضرایب بار[۴] و توصیف‌کننده‌های خود همبستگی دو بعدی[۵] می‌شود.

۱.Substructure List Representation
۲.Molecular Walk Count
۳.Burden-CAS-University of Texas eigenvalues
۴.Charge indices
۵.2D autocorrelation

۲-۲-۲-۴- توصیف‌کننده‌های سه بعدی

نمایش سه بعدی یک مولکول را معمولاً به عنوان یک شیء هندسی انعطاف ناپذیر در فضا می‌بینند و نه تنها یک نمایش از ماهیت و اتصال اتم‌ها، بلکه از صورت بندی فضائی مولکول نیز فراهم می‌کند که به آنها نمایش‌های هندسی[۱] گفته می‌شود. توصیف‌کننده‌های مولکولی مشتق شده از این نمایش، توصیف‌کننده‌های سه بعدی نامیده می‌شوند. زیرگروه‌های این دسته عبارتند از: شاخص‌های آروماتیکی[۲]، خصوصیات مولکولی راندیک[۳] و توصیف‌کننده‌های 3D-MoRSE[۴].

۲-۳- انتخاب بهترین توصیف‌کننده‌ها

تعداد زیاد توصیف کننده‌ها گمراه کننده است و فرایند برازش و مدل‌سازی را طولانی و مدل نهایی را پیچیده می‌کند؛ بنابراین از کاهش متغیر[۵] و انتخاب متغیر[۶] استفاده می شود که کیفیت اطلاعات و مدل (مخصوصاً قدرت پیش‌بینی آن) را بهبود می‌بخشد.

کاهش متغیر انتخاب زیر مجموعه‌ای از متغیرها (توصیف‌کننده‌ها) است که اطلاعات اساسی مربوط به کل سری داده را در بر دارد و متغیرهایی که با هم، همبستگی خیلی بالایی دارند را حذف می‌کند.

بنابراین به طور خلاصه پس از محاسبه توصیف‌کننده‌ها باید از بین آنها، توصیف‌کننده‌هایی که با متغیر وابسته همبستگی بیشتری دارند را انتخاب نمود. برای این منظور چند فاکتور وجود دارند که عبارتند از:

۱.Geometrical Representation

۲ .Aromaticity indices

۳.Randic molecular properties

۴. 3-D MOlecular Representation of Structures based on Electron diffraction

۵.Variable reduction

۶.Variable selection

- توصیف‌کننده‌هایی که بیش از ۹۰٪ مقادیر آنها یکسان باشند حذف می‌شوند.
- از بین توصیف‌کننده‌هایی که با هم همبستگی بالای ۰/۹ دارند توصیف کننده‌ای که با متغیر وابسته همبستگی کمتری دارد، حذف می شود.
- توصیف‌کننده‌هایی که محاسبه و تفسیر آنها مشکل است حذف می‌شوند.

۲-۴- آنالیز مدل‌های آماری و انتخاب مدل مناسب

هدف از انتخاب متغیر دستیابی به یک مدل مناسب که شامل تعدادی از متغیرهای مستقل باشد، برای پیش‌بینی مقدار متغیر وابسته است. مدل‌های رگرسیون که شامل تعداد کمی از متغیرهای مستقل هستند بر سایر مدل‌ها برتری دارند، زیرا این مدل‌ها می‌توانند با سهولت تفسیر شوند و همچنین قدرت پیش‌بینی بالایی دارند. اشکال عمده فرآیندهای انتخاب متغیر در احتمال انتخاب متغیرهای با همبستگی شانسی[۱] است. از این رو برای اجتناب از همبستگی شانسی باید به تکنیک‌های اعتبارسنجی توجه ویژه‌ای شود. متداول‌ترین روش‌های انتخاب متغیر شامل روش‌های رگرسیون مرحله‌ای[۲]، الگوریتم ژنتیک[۳] و آنالیز کلاستر[۴] می‌باشند.

مدل یک رابطه ریاضی است که بیانگر رابطه بین متغیر وابسته و متغیر(های) مستقل می‌باشد و با استفاده از روش‌های آماری گوناگون، می‌توان آن را به دست آورد. یکی از این روش‌ها رگرسیون خطی چندگانه می‌باشد که در آن چندین روش رگرسیون مختلف وجود دارد که به برخی از آنها اشاره می‌شود:

۱.Chance Correlation
۲.Stepwise regression
۳.Genetic Algorithm (GA)
۴.Claster Analysis

۲-۴-۱- روش ورود اجباری[۱]

فرآیندی برای انتخاب متغیرها است که طی آن تمام توصیف‌کننده‌ها در یک مرحله وارد می‌شوند. از آنجا که در این روش اثرات مثبت و منفی متغیرهای مستقل برای ورود در مدل نادیده گرفته می‌شوند، روش مطلوبی نیست.

۲-۴-۲- روش جلو برنده[۲]

تکنیکی مرحله‌ای برای انتخاب متغیرها است که طی آن متغیرها یکی پس از دیگری وارد مدل می‌شوند. اولین متغیری که برای ورود به مدل انتخاب می‌شود دارای بالاترین همبستگی مثبت یا منفی با متغیر وابسته است. این روند تا زمانی که دیگر متغیری وجود نداشته باشد که دارای شرط ورود باشد و باعث بهبود مدل گردد، ادامه دارد.

۲-۴-۳- روش عقب برنده[۳]

در این تکنیک تمام توصیف‌کننده‌ها وارد معادله می‌شوند و سپس متغیر با کمترین همبستگی با متغیر وابسته حذف می‌شود. پس از این که اولین متغیر حذف شد از بین متغیرهای باقیمانده در معادله متغیری که کوچکترین همبستگی را با متغیر وابسته دارد برای حذف در نظر گرفته می‌شود. این فرآیند زمانی متوقف خواهد شد که حذف متغیری که از شرط فرآیند پیروی کند باعث کاهش ارزش مدل گردد.

---

۱. Enter

۲. Forward

۳. Backward

۲-۴-۴- روش مرحله‌ای[۱]

این روش مشابه روش انتخاب جلو برنده است، با این تفاوت که در این روش در هر مرحله برای متغیرهای وارد شده در مدل، آماره F محاسبه می‌شود و متغیر با کوچک‌ترین آماره F حذف خواهد شد و مدل مناسب به دست می آید[۲۵].

۲-۵- مدل‌سازی به کمک روش MLR

انتخاب نوع آنالیز چند متغیره تحت تأثیر نوع توصیف‌کننده‌های مورد استفاده می باشد. یکی از تکنیک‌های نسبتاً ساده‌ای که معمولاً مورد استفاده قرار می‌گیرد روش رگرسیون خطی چندگانه (MLR) است. این روش تنها زمانی می‌تواند به کار رود که تعداد مولکول‌های مورد بررسی بیشتر از تعداد توصیف‌کننده‌ها باشد و متغیرها با یکدیگر همبستگی زیادی نداشته باشند. MLR روشی است که برای مدل‌سازی ارتباط خطی بین متغیر وابسته و یک یا چند متغیر مستقل به کار می‌رود. متغیرهای مستقل، آنهایی هستند که سهمی در تابعی دارند که برای مدل‌سازی مناسب است. متغیرهای وابسته (متغیرهای پاسخ) نیز متغیرهایی هستند که یافتن وابستگی آماری میان آنها و تعدادی از متغیرهای مستقل موردنظر است و معمولاً از اندازه‌گیری‌های تجربی به‌دست می‌آیند.

برای n داده تجربی و p متغیر مستقل، مدل رگرسیون خطی چندگانه توسط معادله (۲-۱) نمایش داده می‌شود:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \qquad \text{(معادله۲-۱)}$$

۱. Stepwise

در این معادله β_0، β_1 و ... وβ_p ضرایب رگرسیون و x_1،x_2، ... وx_p متغیرهای مستقل (توصیف‌کننده‌های عددی) می باشند و Y_i متغیر وابسته مورد نظر (در این تحقیق پارامتر حلالیت یا شاخص بازداری) می‌باشد.

۲-۵-۱- شاخص‌های آماری[۱] [۲۶،۲۷]

شاخص‌های آماری برای ارزیابی کارایی مدل‌های رگرسیونی به کار گرفته می‌شوند. این شاخص‌ها، از دو شاخه آماری به نام‌های *بررسی صحت برازش*[۲] و *بررسی صحت پیش‌بینی*[۳] گرفته شده‌اند. در اولین شاخه، بیشتر به خواص برازشی مدل و در شاخه دوم (تکنیک‌های ارزیابی) بیشتر به قدرت پیش‌بینی مدل توجه می‌شود. شاخه صحت برازش، میزان برازش داده‌های سری آموزشی را در مدل مورد نظر می‌سنجد. در ادامه چندین شاخص آماری مهم فهرست شده‌اند. درجه آزادی خطا[۴](df_E) با فرمول (n-p') به دست میاید، که n تعداد نمونه‌ها و 'p تعداد پارامترهای مدل است. برای مثال برای یک مدل رگرسیون با p متغیر و عرض از مبدأ، درجه آزادی n-p-1 می‌باشد. پارامترهای دیگر درجه آزادی مدل[۵](df_m) و درجه آزادی کل[۶](df) هستند.

---

۱. Statistical indices

۲.Goodness of fit

۳.Goodness of prediction

۴.freedom degree of Error

۵.freedom degree of model

۶.Total freedom degree

۲-۵-۱-۱- شاخص‌های آماری مورد استفاده در برازش

- مجموع مربعات باقیمانده‌ها[۱] (RSS): به صورت مجموع مربعات تفاضل بین پاسخ‌های تجربی برای نمونه i ام (y_i) و پاسخ‌های محاسبه شده با مدل($\hat{y}_i$) برای نمونه i ام تعریف می‌شود.

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \text{(معادله ۲-۲)}$$

- مجموعه مربعات مدل[۲] (MSS): به صورت مجموع مربعات تفاضل بین پاسخ‌های محاسبه شده با مدل برای نمونه i ام و پاسخ‌های میانگین($\bar{y}$) تعریف می‌شود (معادله ۲-۳).

$$Mss = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad \text{(معادله ۲-۳)}$$

- مجموع مربعات کل[۳] (TSS): به صورت مجموع مربعات تفاضل بین مقدار تجربی و مقدار میانگین تعریف می‌شود (معادله ۲-۴).

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad \text{(معادله ۲-۴)}$$

ضریب تعیین[۴] (R^2): مقدار این آماره بین صفر و یک تغییر می‌کند و به دو شکل می‌توان آن را محاسبه کرد (معادله ۲-۵):

$$R^2 = \frac{MSS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad \text{(معادله ۲-۵)}$$

هنگامی که R^2 برابر با یک باشد، مدل و داده‌ها با هم برازش کامل دارند.

۱.Residual sum of squares

۲.Model sum of squares

۳.Total sum of squares

۴.Coefficient of determination

در این رابطه آماره دیگری به نام R نیز تعریف شده است که جذر R^2 بوده و ضریب همبستگی چند متغیره[1] نام دارد. R معیاری از ارتباط خطی بین پاسخ تجربی و پاسخ محاسبه شده توسط مدل است.

- آزمون نسبت F: یکی از معروف‌ترین آزمون‌های آماری است که به صورت زیر تعریف می‌شود:

$$F = \frac{MSS/df_M}{RSS/df_E}$$ (معادله ۲-۶)

مقدار به دست آمده با مقدار بحرانی برای F در همان درجه آزادی مقایسه می‌گردد. این آماره مقایسه‌ای بین واریانس مدل به دست آمده و واریانس باقیمانده‌هاست. مقادیر بالای F نشان‌دهنده مدل‌های قابل اعتماد هستند.

- ضریب همبستگی میان توصیف‌کننده‌ها یا همبستگی پیرسون[2] : مقیاسی از ارتباط خطی بین دو متغیر مستقل است. مقادیر ضریب همبستگی بین ۱- تا ۱+ تغییر می‌کند که علامت آن نشان دهنده جهت ارتباط خطی (مستقیم(+) یا عکس بودن(-) آن) و قدر مطلق آن بزرگی این همبستگی را نشان می‌دهد.

شاخه آماری که صحت را پیش‌بینی می‌کند، در واقع تکنیک‌های ارزیابی را مورد استفاده قرار می‌دهد. این تکنیک‌ها به عنوان معیاری برای انتخاب مدل نیز مورد استفاده قرار می‌گیرند.

- PRESS[3]: مجموع مربعات تفاضل بین پاسخ‌های تجربی برای نمونه i ام و پاسخ‌های تخمین‌زده شده است که به صورت زیر تعریف می‌شود:

$$PRESS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$ (معادله ۲-۷)

۱.Multiple Correlation Coeficient

۲.Pearson Correlation

۳.Predicted REsidual Sum of Squares

R^2_{cv} : ضریب تعیین به دست آمده از ارزیابی متقاطع[1] که به صورت زیر تعریف می‌شود (معادله ۲-۸):

$$R^2_{cv} = 1 - \frac{PRESS}{TSS}$$ (معادله ۲-۸)

۲-۶- تکنیک‌های اعتبارسنجی

این تکنیک‌ها برای ارزیابی اعتبار مدل‌های به دست آمده به‌کار می‌روند. تکنیک‌های اعتبارسنجی برای بررسی قدرت پیش‌بینی مدل‌ها استفاده می‌شوند، یعنی توانایی آنها را برای انجام پیش‌بینی‌های مطمئن برای موارد جدیدی که پاسخ آنها ناشناخته است می‌سنجند. یک شرط لازم برای اعتبار یک مدل رگرسیون آن است که ضریب تعیین (R^2) یا ضریب همبستگی چند متغیره (R) آن تا حد ممکن به یک نزدیک باشد و خطای استاندارد (SE) آن کوچک باشد. به هرحال این شرط (توانایی برازش)، زمانی که مدل‌ها برازش نزدیک‌تری (SE کوچک‌تر و R^2 بزرگ‌تر) را برای تعداد زیادتری از پارامترها و متغیرها در مدل به دست می‌دهند، برای اعتبارسنجی مدل کافی نیست؛ زیرا این پارامترها ارتباطی به توانایی مدل برای ایجاد پیش‌بینی‌های مطمئن بر روی داده‌های بعدی ندارند. مسائل دیگر برای اعتبارسنجی مدل‌ها زمانی پیش می‌آیند که لازم است مدل‌هایی با متغیرهای اندک، با استفاده از فرآیندهای بر پایه انتخاب متغیر به دست آیند. زمانی که یک مجموعه با تعداد زیادی از توصیف کننده‌ها برای انتخاب از آنها در دسترس باشد، مدل‌های به‌دست‌آمده دارای برازش خوب اما همراه با همبستگی شانسی[2] هستند، یعنی منحنی مقادیر نظری محاسبه شده از مدل بر حسب مقادیر تجربی در یک خط مستقیم واقع شده بدون این که توانایی پیش‌بینی داشته باشد. برای جلوگیری از مدل‌های با همبستگی شانسی، یک اعتبارسنجی

۱.Cross Validation (CV)

۲.Chance Correlation

متفاوت باید انجام شود. بهتر است یک سری ارزیابی مناسب از ترکیبات کـه در فراینـد آمـوزش و بهینـه سازی دخالت نداشته (سری ارزیابی بیرونی[1])، در دسترس باشد. تعدادی از تکنیک‌هـای اعتبارسـنجی در زیر آمده است.

۲-۶-۱- ارزیابی متقاطع

رایج‌ترین تکنیک اعتبارسنجی است که در آن در هر بار یک یا یک گروه کوچک از داده ها کنـار گذاشـته شده و سپس برای داده‌های باقی مانده، مدلی محاسبه می‌شود و پاسخ برای داده‌های کنارگذاشته شده از روی این مدل پیش‌بینی می‌شود.

۲-۶-۲- تقسیم کردن مجموعه به آموزش و ارزیابی

این تکنیک بر اساس تقسیم بندی مجموعه داده به یک سری آموزش[2] و یک سری ارزیابی[3] است. مدل از سری آموزش محاسبه شده و قدرت تعمیم توسط سری ارزیابی بررسـی مـی شـود. تقسـیم‌بنـدی داده‌ها به دو سری به صورت تصادفی انجام می‌گیرد.

۲-۶-۳- اعتبار سنجی بیرونی[4]

تکنیکی است که در آن یک سری تست[1] نیـز بـرای انجـام یـک بررسـی اضـافی بـر روی قـدرت پیش‌بینی مدل به دست آمده از سری آموزش و ارزیابی، در نظر گرفته می‌شود. سری تست نباید دخالتی در روند آموزش و انتخاب مدل داشته باشد.

۱.External Validation Set

۲.Training set

۳.Validation set

۴.External Validation

۲-۶-۴- آزمون Y- تصادفی[۲]

این تکنیک برای مطالعه همبستگی شانسی مدل طراحی شده است. اگر در مدلی متغیرهای مستقل به صورت تصادفی با متغیرهای پاسخ همبستگی داشته باشند، گفته می‌شود دارای همبستگی شانسی است. در این آزمون، مقادیر پاسخ (که در اینجا بردار Y نامیده می‌شود) به صورت تصادفی در محدوده پاسخ‌ها، تغییر داده می‌شود. سپس همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از یکی از شاخص‌های آماری (معمولاً R^2) مورد بررسی قرار می‌گیرد. اختلاف زیاد بین شاخص آماری به دست آمده از این روش و شاخص آماری به دست آمده از مدل اصلی، نشان‌دهنده عدم وجود همبستگی شانسی می‌باشد. این فرایند چندین بار تکرار می‌شود[۲۸].

۲-۷- نرم افزارهای مورد استفاده در این تحقیق

در این تحقیق از بسته های نرم افزار HyperChem 8 برای رسم ساختار ترکیبات مورد بررسی، Dragon 2/1 برای محاسبه توصیف‌کننده‌های مولکولی، SPSS 16 برای انجام آنالیزهای آماری رگرسیون های خطی چندگانه و MATLAB 2008 b به منظور برنامه‌نویسی شبکه عصبی مصنوعی استفاده شد، که در ادامه به معرفی آنها پرداخته خواهد شد. برای محاسبه برخی پارامترهای آماری و نیز رسم نمودارهای سه‌بعدی نیز نرم افزار Sigma Plot مورد استفاده قرار گرفت.

۲-۷-۱- نرم افزار Hyperchem

با این نرم افزار، ابتدا شکل مولکول به طور تقریبی بر روی صفحه نمایشگر کامپیوتر رسم و سپس توسط روش های کوانتومی– مکانیکی بهینه می‌شود. روش‌های کوانتومی- مکانیکی موجود عبارتند از

۱.Test set

۲.Y-Randomization

دینامیک مولکولی و روش‌های نیمه تجربی که خود از هـامیلتونین AM1، CNDO ، MNDO یـا INDO استفاده می‌کنند. با استفاده از این نرم افزار مـی‌تـوان اطلاعـات فراوانـی را نظیـر زوایـای پیونـدی، طـول پیوندها، زوایای پیچشی، بار اتم‌ها و انرژی تشکیل مولکول و غیره به دست آورد[۲۹].

۲-۷-۲- نرم افزار Dragon

از بسته نرم افزاری Dragon برای محاسبه توصیف‌کننده‌های مولکولی استفاده می‌شود که توسط گروه QSAR و کمومتریکس دانشگاه میلانو طراحی شده است. این توصیف‌کننده‌های مولکولی به منظـور ارزیابی روابط ساختار– فعالیت یا ساختار - ویژگی مولکولی به کار می‌رونـد. اولـین ویـرایش آن بـه سـال ۱۹۹۷ برمی‌گردد. به منظور تحقیقات پیشرفته‌تر در زمینه QSAR بـه روز رسـاندن و افـزایش گنجـایش توصیف‌کننده‌های مولکولی مرتباً در این نرم افزار ایجاد شد. نرم افزار Dragon قادر بـه محاسـبه بـیش از ۱۴۹۷ توصیف‌کننده می‌باشد که به ۱۸ دسته اصلی تقسیم می‌شوند که در جدول ۲-۱ آورده شده است.

جدول ۲-۱- توصیف کننده های محاسبه شده توسط نرم افزار Dragon

۱.توصیف‌کننده‌های زیر ساختاری	۱۰.توصیف‌کننده های هندسی
۲.توصیف‌کننده‌های توپولوژیکی	۱۱.توصیف‌کننده‌های RDF
۳.شمارنده‌های مولکولی مورس	۱۲.توصیف‌کننده‌های سه بعدی
۴.توصیف کننده‌های BCUT	۱۳.توصیف‌کننده های WHIM
۵.شاخص‌های بار	۱۴.توصیف‌کننده‌های GETAWAY
۶.خود ارتباطی‌های دو بعدی	۱۵.گروه‌های عاملی
۷.توصیف کننده‌های بار	۱۶.اجزای میان اتمی
۸.شاخص‌های آروماتیسیته	۱۷.توصیف‌کننده‌های تجربی
۹.پروفایل‌های مولکولی راندیک	۱۸.خصوصیات مولکولی

کاربر علاوه بر محاسبه ساده‌ترین نوع اتم‌ها، گروه‌های عاملی و شمارش اجزا می‌تواند تعداد زیادی توصیف‌کننده‌های توپولوژیکی و هندسی را نیز توسط این نرم افزار محاسبه کند. برای اجرای این نرم افزار

کاربر نیاز به فایل‌های ساختار مولکولی دارد که قبلا توسط سایر نرم افزارهای مدل سازی مولکولی مانند Hyperchem تأمین شده‌اند. اکثر فایل‌های مولکولی با فرمت‌های معمول، در این نرم افزار قابل قبول هستند. برای استفاده بهتر از محاسبات این نرم افزار باید بهینه‌سازی سه بعدی ساختارها با هیدروژن‌هایشان به کاررود. Dragon به عنوان یک نرم افزار QSAR طراحی نشده است، بلکه تنها توصیف‌کننده‌های مولکولی را نمایش می‌دهد و هیچ آنالیز QSAR انجام نمی‌دهد. با این حال، Dragon یک فایل خروجی کامل را که به سادگی توسط هر نرم افزار آنالیز همبستگی قابل کاربرد است، فراهم می‌سازد. اطلاعات کامل در مورد توصیف‌کننده‌های به کار رفته در این نرم افزار (تعاریف، فرمول‌ها، مراجع و جزئیات توصیف کننده‌ها) را می‌توان از کتاب مرجع توصیف‌کننده‌های مولکولی که ۳۳۰۰ مرجع برای کل توصیف‌کننده‌ها ارائه می‌دهد، به دست آورد[۳۰].

۲-۷-۳- نرم افزار SPSS

کلمه SPSS مخفف عبارت Statistical Package for Social Science (نرم افزار آماری برای علوم اجتماعی) می‌باشد. این بسته نرم‌افزاری، توانایی تحلیل کلی اطلاعات را دارا بوده و عمومی‌ترین نرم‌افزار آماری است. امروزه آمار، مجموعه‌ای از مفاهیم و روش‌هایی است که در هر زمینه پژوهشی، برای گردآوری و تحلیل اطلاعات مربوط به آن و انجام نتیجه‌گیری‌ها، در شرایطی که عدم قطعیت و تغییر وجود دارد، به‌کار می‌رود. SPSS دارای توانایی‌های بسیار بالایی است که به تعدادی از آنها اشاره می‌شود[۳۱] :

- تهیه خلاصه‌های آماری مانند گراف‌ها، جداول، آماره‌ها و ...
- محاسبه انواع توابع ریاضی مانند قدر مطلق، تابع علامت، لگاریتم، توابع مثلثاتی و...
- تهیه انواع جداول سفارشی مانند جداول فراوانی، فراوانی تجمعی، درصد فراوانی و...

- انواع توزیع‌های آماری شامل توزیع‌های گسسته و پیوسته
- تهیه انواع طرح‌های آماری
- انجام آنالیز واریانس یکطرفه، دوطرفه، چندطرفه و آنالیز کوواریانس
- تکنیک‌های تجزیه و تحلیل سری‌های زمانی
- ایجاد داده‌های تصادفی و پیوسته
- محاسبه انواع آماره‌های توصیفی
- انواع آزمون‌های مرتبط با مقایسه میانگین بین دو یا چند جامعه مستقل و وابسته
- قابلیت مبادله اطلاعات با نرم‌افزارهای دیگر
- برازش انواع مختلف رگرسیون

در این جا از نرم افزار SPSS برای حذف متغیرهای با همبستگی بالا، به‌دست آوردن بهترین مدل و ارزیابی آن استفاده شده است.

۲-۷-۴- نرم افزار MATLAB

MATLAB نرم افزاری با کارایی بالا برای محاسبات تخصصی است. اولین نسخه‌های آن در دانشگاه نیومکزیکو و استانفورد در سال های دهه ۱۹۷۰ میلادی جهت حل مسائل نظریه ماتریس‌ها، جبر خطی وآنالیز عددی به وجود آمد. به تدریج و با افزودن امکانات مختلف، MATLAB به عنوان یک نرم افزار قوی و در عین حال ساده برای کاربران تبدیل شد. این نرم افزار امکان حل مسائل تخصصی، به خصوص مسائلی که به صورت ماتریسی هستند را فراهم می‌کند. یکی از قابلیت‌های آن برنامه‌نویسی در محیطی ساده است که مسائل و راه‌حل‌ها با علائم ریاضی مشابهی تعریف می‌شوند. کاربردهای نوعی آن شامل ریاضیات و الگوریتم محاسباتی، شبیه‌سازی، مدل‌سازی و آنالیز اطلاعات می‌باشد. یکی از کاربردهای

مهم آن که مورد استقبال شیمیدانان قرار گرفته است، امکان استفاده از آن برای آنالیز آماری و مدل‌سازی داده های شیمیایی است. مدل‌سازی به روش شبکه عصبی مصنوعی از جمله کاربردهای مهم دیگر آن است که در تحقیق حاضر مورد استفاده قرار گرفته است. اطلاعات شیمیایی ترکیبات به صورت ماتریس به نرم افزار ارائه می شوند. در محیط MATLAB با استفاده از فرامین، آرایه ها و امکانات ویژه، مدلسازی غیر خطی فعالیت شیمیایی ترکیبات امکان پذیر است. سایر روش‌های کمومتریکس مانند آنالیز رگرسیون اجزای اصلی، روش کمترین مربعات جزئی، الگوریتم ژنتیکی و غیره نیز توسط این نرم افزار قابل اجرا هستند[۳۲].

۲-۸- شبکه عصبی مصنوعی

شبکه‌های عصبی زیر مجموعه‌ای از هوش مصنوعی هستند. این شبکه‌ها در واقع تقلیدی از سیستم‌های عصبی زیستی هستند و مانند آنها از عناصر عملیاتی ساده‌ای به نام نرون[۱] که به صورت موازی عمل میکنند، ساخته شده‌اند. پردازش موازی به معنای اجرای برنامه‌هایی است که در هر بار بیش از یک عمل را انجام می‌دهند. کامپیوترهای معمولی عمل‌ها را یکی پس از دیگری (به صورت ترتیبی) انجام می‌دهند.

پردازش موازی به دو دلیل مورد توجه هوش مصنوعی است. نخست این که گشودن تعدادی از مساله‌های هوش مصنوعی نیازمند استفاده طولانی مدت از کامپیوتر می‌باشد. اگر این زمان بتواند به وسیله راه‌حل‌های پردازش موازی چگالیده شود، مقدار زیادی در زمان اجرا صرفه‌جویی می‌شود. دوم اینکه برخی مسأله‌ها با شیوه‌های موازی بهتر حل می‌شوند. برای مثال بینایی فعالیتی است که در آن همه

۱.Neuron

تصویر باید به یکباره پردازش شود، چون تفسیر قسمتی از تصویر بـه تفسـیر قسـمت‌هـای دیگـر بسـتگی دارد[۳۳].

برای آشنایی با شبکه‌های عصبی مصنوعی، ابتدا باید سیستم عصبی زیستی را به اختصار مورد بررسی قرار داد.

۲-۸-۱- سیستم عصبی زیستی

از مغز به عنوان یک سیستم پردازش اطلاعات با ساختار موازی و کاملاً پیچیده که دو درصد وزن بدن را تشکیل می‌دهد و بیش از بیست درصد کـل اکسـیژن بـدن را مصـرف مـی‌کنـد بـرای بسـیاری از رفتارهای خودآگاه و ناخود آگاه همچون تفکر، نفس کشیدن، خواندن، دویدن و... استفاده می‌شود. مغـز از ۱۰۰ تریلیون نرون که با بیش از $۱۰^{۱۶}$ ارتباط به هم متصل شده‌اند تشکیل شده است. نرون‌ها ساده‌تـرین واحد ساختاری سیستم‌های عصبی هستند. بیشتر نرون‌ها از سه قسمت اساسی تشکیل شده‌اند:

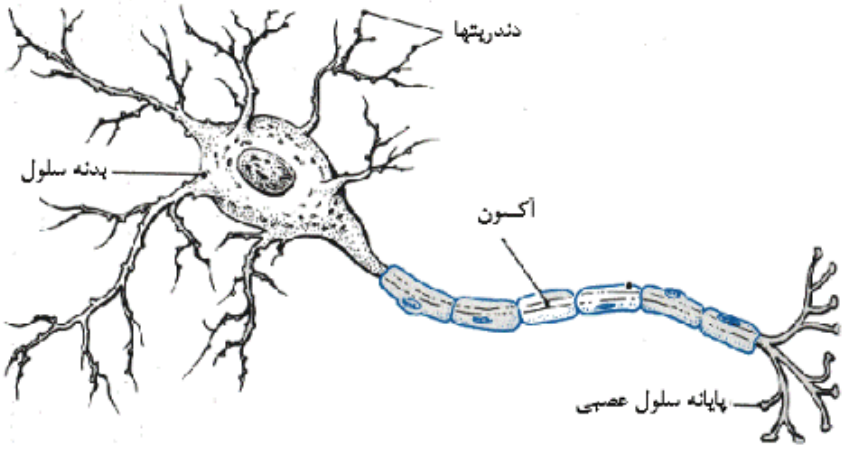
- بدنه‌ی سلول که شامل هسته و قسمت‌های اساسی دیگر می‌باشد.
- دندریت[۱]
- آکسون[۲]

دندریت‌ها به عنوان مناطق دریافت سیگنال‌های الکتریکی، شبکه‌هایی تشکیل یافتـه از فیبرهـای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بیشـمار مـی‌باشـند. بـه همـین علـت آنهـا را شبکه‌های دریافت‌کننده درخت مانند نیز می‌نامند. دندریت‌ها سیگنال‌های الکتریکی را بـه هسـته سـلول منتقل می‌کنند. بدنه سلول، انرژی لازم را برای فعالیت نرون فراهم نموده و بر روی سیگنال‌های دریـافتی عمل می‌کند.

۱.Dendrite

۲.Axon

آکسون‌ها بر عکس دندریت‌ها از سطحی هموارتر و تعداد شاخه‌های کمتری برخوردار هستند. آنها طول بیشتری دارند و سیگنال الکتروشیمیایی دریافتی از هسته سلول را به نرون‌های دیگر منتقل می‌کنند. هر آکسون توسط یک فاصله کوچک موسوم به سیناپس از نرون‌های مجاور جدا می‌شود. سیناپس‌ها واحدهای ساختاری کوچکی هستند که ارتباط بین نرون‌ها را برقرار می‌سازند. چگونگی برقراری این ارتباط به میزان مواد انتقال‌دهنده نرونی که در انتهای آکسون‌ها ذخیره شده‌اند بستگی دارد. وقتی که یک پتانسیل تحریک به انتهای یک آکسون می‌رسد، موجب آزاد شدن یک ماده شیمیایی به نام انتقال‌دهنده نرونی از انتهای آکسون می‌شود و پس از نفوذ در سیناپس‌ها، گیرنده‌های سلول‌های مجاور را فعال می‌کند [۳۴]. شکل ۲-۲ نشان‌دهنده ساختار کلی یک نرون می‌باشد.

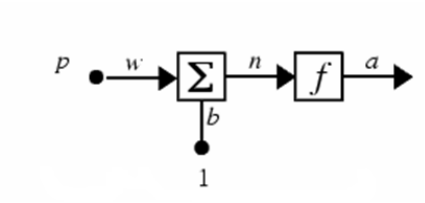


شکل ۲-۲ شمای سلول عصبی زیستی

۲-۸-۲- مدل ریاضی نرون

همانطور که گفته شد، در طبیعت ساختار شبکه‌های عصبی از طریق نحوه اتصال بین اجزا تعیین می‌شود. بنابراین می‌توان یک ساختار مصنوعی به تبعیت از شبکه‌های طبیعی ساخت و با تنظیم مقادیر

هر اتصال تحت عنوان وزن اتصال[۱] نحوه ارتباط بین اجزای شبکه را تعیین نمود. معمولاً برای نشـان دادن مدل ریاضی ساده‌ای از نرون شبکه عصبی مصنوعی از شکل زیر استفاده می‌شود:



شکل ۲-۳- مدل ریاضی نرون

در این شکل از علامت p برای نشان دادن یک سیگنال ورودی استفاده می‌شود. در واقع در این مدل یـک سیگنال ورودی پس از تقویت (یا تضعیف) شدن با وزن w، با مقدار n = pw به تابع انتقال[۲] نرون، f، اعمال می‌شود. ورودی بایاس[۳] یک مقدار ثابت، ۱، است. مقدار بایاس با n جمع می‌شود و درواقع تابع را به سمت چپ شیفت می‌دهد.

b و w دو پارامتر تنظیم شونده در نرون‌ها می‌باشند و ایده اصلی شبکه عصبی این است که با تغییر مقادیر b و w، شبکه یک رفتار یا تصمیم را اتخاذ کند. در جعبه ابزار مورد استفاده در MATLAB، بایاس در نظر گرفته شده اما استفاده از آن اختیاری می‌باشد[۳۵].

۲-۸-۲-۱- توابع انتقال

سه تابع انتقال رایج در شبکه‌های عصبی، که در این پایان‌نامه برای بهینه‌سازی شبکه به‌کار رفتـه است عبارتند از:

۱. Connection weight
۲.Transfer function
۳.bias

- تابع انتقال خطی

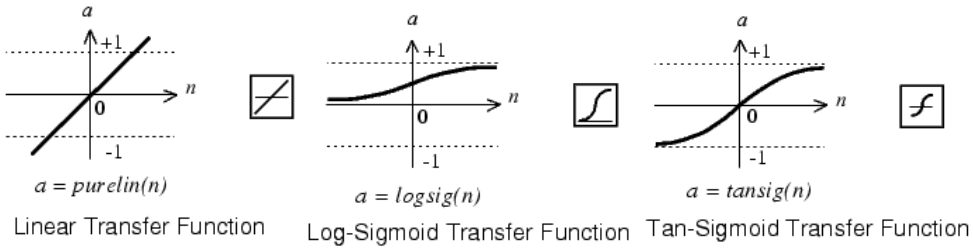
نرون‌هایی که از این تابع انتقال استفاده می‌کنند معمولاً برای تقریب خطی در فیلترهای خطی به کار می‌روند. این تابع همان مقدار ورودی را به عنوان خروجی برمی‌گرداند (شکل ۲-۴ الف).

- تابع انتقال لگاریتم زیگموئید (logsig)

از این تابع انتقال در شبکه‌های پس انتشار استفاده می‌شود. این تابع انتقال مقادیر ورودی را در محدوده ∞- و ∞+ دریافت کرده و خروجی بین ۱ و ۰ تولید می‌نماید (شکل ۲-۴ ب).

- تابع انتقال تانژانت هایپربولیک (tansig)

این تابع انتقال مقادیر ورودی را در محدوده ∞- و ∞+ دریافت کرده و خروجی بین ۱ و ۱- تولید می‌نماید (شکل ۲-۴ ج).



ج ب الف

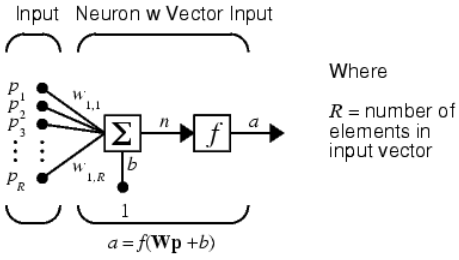
شکل ۲-۴- برخی از توابع انتقال مورد استفاده در شبکه‌های عصبی

۲-۸-۳- معماری شبکه‌های عصبی[۱]

اگر بردار ورودی p را با عناصر p_R،...،p_2،p_1 و بردار وزن‌ها، w، را بـا عناصـر $w_{1,R}$،.....، $w_{1,2}$،$w_{1,1}$ در نظر بگیریم، برای اعمال مقادیر وزن در مقادیر ورودی دو بردار p و w را در هم ضرب ماتریسی می‌کنـیم، بنابراین ورودی تابع انتقال F یعنی n به صورت زیر خواهد بود:

$$n = w_{1,1}p_1 + w_{1,2}p_2 + \ldots + w_{1,R}p_R + b$$ معادله ۲-۹

در نهایت b مقدار بایاس است که به حاصلضرب ماتریسی w در p اضافه شده اسـت. بـه ترکیـب وزن‌هـا، بایاس، تابع انتقال و عملیات ضرب و جمع انجام شده یک لایه از شبکه می‌گویند، شایان توجه اسـت کـه بردار ورودی در بیشتر نوشتارها، یک لایه محسوب نمی‌شود. اگر از تابع انتقال خاصی استفاده شود، نمـاد آن تابع در کادر مربوط به تابع قرار می‌گیرد. شکل ۲-۵ نمایش خلاصه‌ای از یک نرون تـک‌لایـه را نشـان می‌دهد.



شکل ۲-۵- نمایش یک نرون تک‌لایه

دو یا چند نرون می‌توانند در یک لایه با هم ترکیب شوند و یـک شـبکه مـی‌توانـد از یـک یـا چنـد لایـه اینچنینی تشکیل شود.

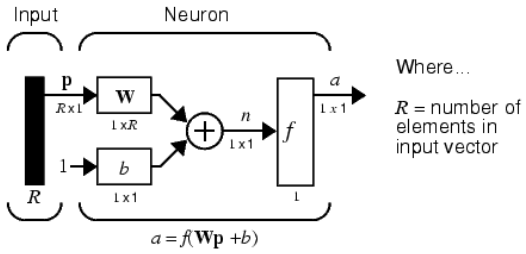
در این شبکه اعضای بردار ورودی p به همه نرون‌ها اعمـال مـی‌شـوند و پـس از ضـرب در بـردار وزن‌ها و جمع با بایاس به تابع انتقال اعمال شده و خروجی حاصل می‌گردد. خروجی شبکه بالا یک بـردار

۱.Network architecture

خواهد بود. در شبکه تک‌لایه بالا ماتریس وزن‌ها (w) یک ماتریس S×R خواهد بود. در این ماتریس $w_{n,m}$ نماینده وزن مربوط به ورودی m روی نرون n می‌باشد.

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,R} \\ & & & \\ w_{S,1} & w_{S,2} & \cdots & w_{S,R} \end{bmatrix}$$

یک شبکه تک‌لایه با S نرون و R ورودی به شکل خلاصه در شکل ۲-۶ نشان داده شده است.



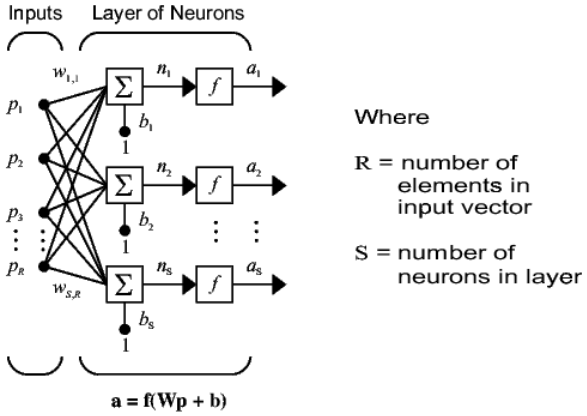
شکل ۲-۶- نمایش یک شبکه تک‌لایه با R ورودی و S نرون

۲-۸-۳-۱- ورودی‌ها و لایه‌ها

از آنجا که در ادامه، شبکه‌های چندلایه مورد بحث قرار خواهند گرفت، لازم است ماتریس‌های وزن مربوط به ورودی‌های نرون‌ها و ماتریس‌های وزن بین لایه‌ها از هم قابل تشخیص باشند. بنابراین مبدأ و مقصد ماتریس‌های وزن باید معلوم باشند. این دو نوع ماتریس را به اختصار با IW (وزن‌های ورودی[۱]) و LW (وزن‌های لایه‌ها[۲]) نشان می‌دهیم. بنابراین تصویر یک شبکه یک‌لایه با S نرون و R ورودی به صورت زیر در خواهد آمد (شکل ۲-۷).

۱. Input Weights

۲. Layer Weights



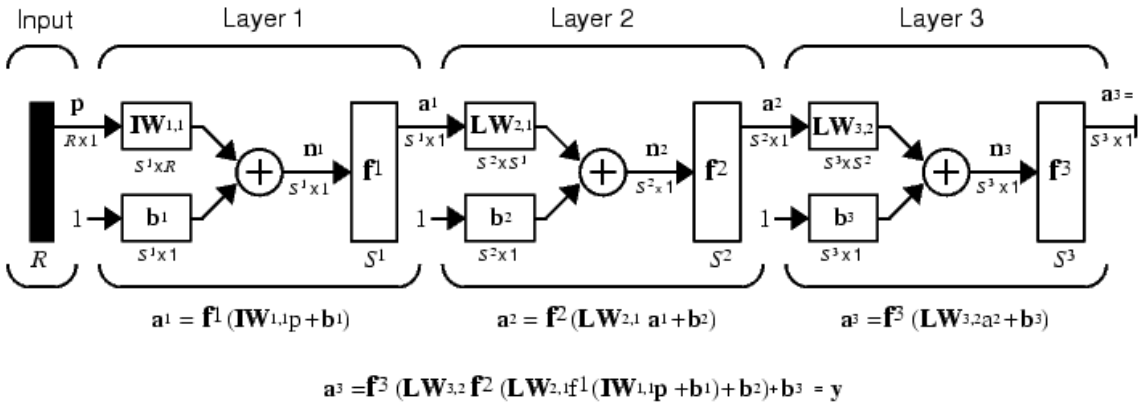
شکل ۲-۷- نمایش یک شبکه تک‌لایه با R ورودی و S نرون

در این شکل $IW^{1,1}$ نشان‌دهنده ماتریس وزن‌های ورودی از مبدأ لایه ۱ (عدد دوم) به مقصد لایه ۱ (عدد اول) می‌باشد. علاوه بر آن، S^1 نماینده تعداد نرون‌های لایه اول و n^1 نماینده تعداد خروجی‌های لایه اول می‌باشد.

۲-۸-۳-۲- شبکه‌های چند لایه

دو یا چند نرون می‌توانند در قالب یک لایه با یکدیگر ترکیب شوند. یک شبکه نیز می‌تواند به نوبه خود

از چند لایه تشکیل شود. هر لایه در شبکه دارای ماتریس وزن‌ها، بردار بایاس و خروجی مختص به خود می‌باشد. برای متمایز ساختن هر یک از این ماتریس‌های وزن و بردارهای بایاس و خروجی‌ها از عددی به صورت بالانویس استفاده می‌شود. شکل ۲-۸ یک شبکه سه‌لایه را همراه با اطلاعات تفکیک شده



شکل ۲-۸- نمایش شبکه چند لایه

بر حسب لایه‌های آن نشان می‌دهد. شبکه نشان داده شده در شکل ۲-۷، دارای تعداد R^1 ورودی و s^1 نرون در لایه اول و s^2 نرون در لایه دوم می‌باشد و در نهایت s^3 نرون در لایه سوم می‌باشد. همانطور که مشخص است یک شبکه می‌تواند دارای تعدادی متفاوت از نرون در لایه‌های مختلف خود باشد. همانطور که مشخص است، خروجی هر یک از لایه‌های میانی به عنوان ورودی لایه بعدی مورد استفاده قرار می‌گیرند. بنابراین لایه دوم از این شبکه را به تنهایی می‌توان یک شبکه تک‌لایه با S^1 (تعداد نرون‌های لایه اول) و S^2 نرون در نظر گرفت که دارای ماتریس وزن‌های w^2 با اندازه $S^2 \times S^1$ می‌باشد. همچنین ورودی لایه دوم a^1 (خروجی لایه اول) و خروجی آن a^2 می‌باشد. همین پارامترها را می‌توان به شکل مشابه برای لایه سوم تعریف نمود. لایه‌های یک شبکه چندلایه وظایف متفاوتی دارند. لایه‌ای که خروجی شبکه را ایجاد می‌کند (لایه آخر) تحت عنوان لایه خروجی شناخته می‌شود. بقیه لایه‌ها تحت عنوان لایه‌های مخفی نامگذاری شده‌اند.

شبکه‌های چندلایه بسیار قدرتمند هستند. به عنوان مثال یک شبکه دو لایه با لایه اول sigmoid و لایه دوم خطی می‌تواند هر تابع دلخواهی را با تعداد محدود ناپیوستگی تخمین بزند. این نوع شبکه دو لایه به شکل وسیع در شبکه‌های پس انتشار[۱] کاربرد دارد.

۲-۸-۳-۳- ساختار داده‌های[۲] مورد استفاده [۳۵]

در این بخش تأثیر فرمت ورودی‌ها در شبیه‌سازی شبکه مورد بررسی قرار می‌گیرد. دو گونه بردار ورودی قابل بحث اند:

- بردارهای ورودی همزمان[۳]: که ارائه داده‌ها به شبکه در یک زمان اتفاق می‌افتد، نه در توالی از زمان‌های مشخص
- بردارهای ورودی ترتیبی[۴]: برخلاف بردارهای همزمان، در نوع ترتیبی ترتیب بردارهای ورودی مهم است.

شبکه ایستا[۵]، شبکه‌ای است که در آن پسخورد[۶] و تأخیر[۷] وجود نداشته باشد. در اینگونه شبکه‌ها ورودی‌ها معمولاً همزمان خواهند بود.

شبکه‌های پویا[۸] شبکه‌هایی هستند که در آنها تأخیر و پسخورد وجود داشته باشد. بردارهای ورودی به شبکه پویا هم می‌توانند به صورت ترتیبی و هم به صورت همزمان ارائه شوند. در نرم‌افزار MATLAB،

۱.Back Propagation
۲.Data structures
۳.Concurrently inputs
۴.Sequential inputs
۵.Static
۶. Feedback
۷.Delay
۸.Dynamic

داده‌های ترتیبی در یک آرایه سلولی و داده‌های همزمان در بردار قرار می‌گیرند. در شبکه مورد استفاده برای این تحقیق، داده‌ها به صورت همزمان ارائه می‌شوند.

۲-۸-۴- آموزش شبکه عصبی

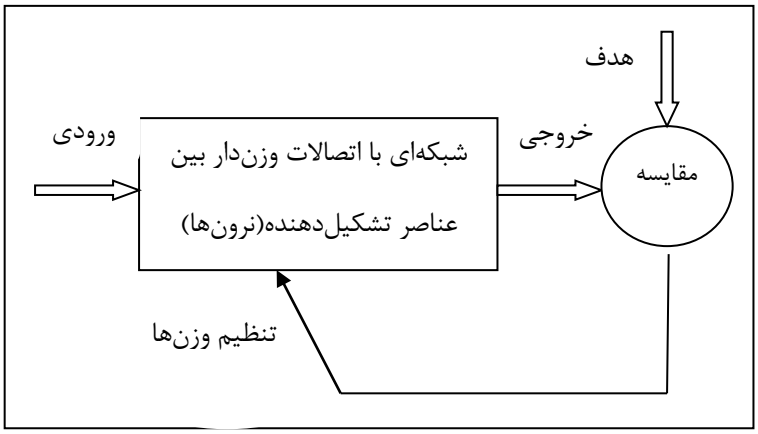
معمولاً شبکه‌های عصبی تنظیم می‌شوند یا آموزش می‌یابند تا اعمال یک ورودی خاص منجر به دریافت خروجی خاص موردنظر یا همان هدف شود. همانطور که در شکل ۲-۹ دیده می‌شود شبکه بر مبنای تطابق و همسنجی بین ورودی و هدف سازگار می‌شود تا اینکه خروجی شبکه و هدف بر هم منطبق گردند. عموماً تعداد زیادی از این زوج‌های ورودی و خروجی به‌کار گرفته می‌شوند تا در این روند که یادگیری تحت نظارت[۱] نامیده می‌شود، شبکه آموزش داده شود. در یادگیری تحت نظارت قاعده یادگیری با استفاده از مجموعه‌ای از مثال‌ها (مجموعه آموزشی) به دست می‌آید و شبکه را آموزش می‌دهد. زوج داده‌های زیر را در نظر بگیرید:

$\{p_1, t_1\}, \{p_2, t_2\}, \ldots \{p_Q, t_Q\}$

که p_Q یک ورودی شبکه و t_Q هدف مورد نظر متناظر با هر ورودی می‌باشد. زمانی که ورودی به شبکه اعمال می‌شود، خروجی آن با هدف مقایسه می‌شود. سپس قواعد یادگیری برای تنظیم وزن‌ها و بایاس‌ها به کار برده می‌شود تا خروجی شبکه را به هدف نزدیک نماید. اما در یادگیری نظارت نشده[۲] ورودی‌ها و بایاس‌ها تنها در مقابل ورودی شبکه اصلاح می‌شوند و در واقع هیچ هدفی وجود ندارد. این نوع الگوریتم‌ها غالباً برای دسته‌بندی سری داده‌ها مورد استفاده قرار می‌گیرند.

۱.Supervised Training

۲.Unsupervised training



شکل ۲-۹- شمای کلی عملکرد شبکه عصبی مورد بحث

عموماً برای آموزش شبکه‌های عصبی از قواعد یادگیری نظارت شده استفاده می‌شود اما می‌توان شبکه‌ها را با روش آموزش غیر نظارتی آموزش داد[۳۵].

۲-۸-۴-۱- روش‌های آموزش

در این بخش دو شیوه متفاوت آموزش شبکه مورد بحث قرار می‌گیرد.

- آموزش گام به گام[۱]: در این شیوه آموزش وزن‌ها و بایاس‌ها بعد از اعمال هر ورودی به شبکه به روز می‌شوند. این شیوه آموزش می‌تواند قابل استفاده در هر دو شبکه ایستا و پویا باشد اما به طور معمول در شبکه‌های پویا مورد استفاده می‌گیرد. در جعبه ابزار MATLAB آموزش گام به گام فقط با تابع adapt ممکن است.
- آموزش دسته‌ای[۲]: در این شیوه وزن‌ها و بایاس‌ها بعد از اعمال تمامی ورودی‌ها به شبکه یکبار به روز می‌شوند. این روش آموزش نیز در هر دو گونه شبکه پویا و ایستا قابل انجام می‌باشد. آموزش دسته‌ای با هر دو تابع train و adapt ممکن است، اما تابع train بهتر می‌باشد.

۱.Incremental training

۲.Batch training

در این تحقیق شیوه آموزش دسته‌ای در شبکه‌های پویا به کار گرفته شده و تابع مورد استفاده برای آموزش trainbr می‌باشد [۳۵].

۲-۸-۵- شبکه‌های پس انتشار

پس‌انتشار روشی است که در شبکه‌های عصبی با بیش از یک لایه نرونی یا به عبارتی با لایه پنهان استفاده می‌شود. برای هر الگوی ورودی خاص، خروجی واقعی با خروجی مورد انتظار (هدف) مقایسه می‌شود. سپس اختلاف بین این دو خروجی به همه اتصال‌هایی که اولین بار برای به دست آوردن خروجی استفاده شده بودند، بازپراکنده (باز پخش) می‌شود. اگر خروجی شبکه با هدف همانندی مناسبی داشته باشد، اتصال واحدهایی که در خروجی موثر بوده‌اند تقویت می‌شود. اگر چنین نباشد، اتصال بین واحدهای مورد نظر، از نظر قوت کاهش می‌یابد. بار دیگر که آن واحدهای خاص برانگیخته می‌شوند، نسبت به دفعه قبل تأثیر کمتری روی واحدهای پنهان و خروجی خواهند داشت. این روش یادگیری تقویت شده برخی از ویژگی‌های یادگیری انسانی را نشان می‌دهد، به خصوص اینکه چگونه بچه‌های کوچک به هنگام زبان گشودن قاعده‌های دستور زبان را بیش از حد تعمیم می‌دهند[۳۶].

این شبکه‌ها که از آنها، به اختصار تحت عنوان BP نیز یاد می‌شود، شبکه‌هایی چندلایه با قاعده یادگیری ویدرو - هوف[۱] می‌باشند. یک شبکه BP دارای بایاس، یک لایه زیگموئید و یک لایه خروجی خطی، توانایی تخمین زدن هر تابعی با نقاط ناپیوستگی محدود را داراست. BP استاندارد، یک الگوریتم با شیب نزولی می‌باشد که در آن وزن‌های شبکه در جهت خلاف تابع کارایی[۲] حرکت می‌کنند. در واقع لغت پس انتشار نیز به رفتار شبکه BP در محاسبه شیب در شبکه‌های غیر خطی چندلایه اشاره دارد.

۱.Widrow Hoff

۲.Performance Function

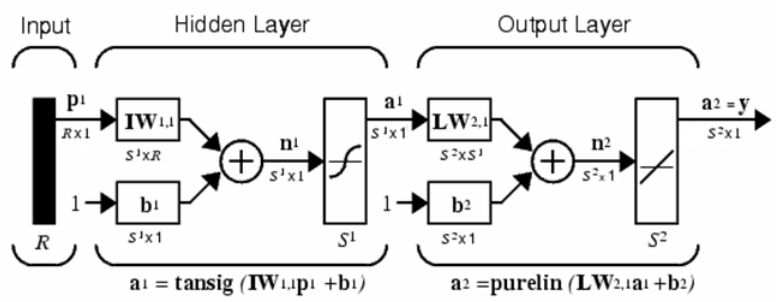
الگوریتم‌های مختلفی مانند روش‌های نیوتن[1] وجود دارند که بر مبنای این الگوریتم استاندارد می‌باشند. جعبه ابزار شبکه عصبی در نرم افزار MATLAB تعدادی از این الگوریتم‌های مختلف را ارائه می‌دهد. عموماً برای به‌کارگیری چنین شبکه‌ای چهار مرحله وجود دارد:

۱. فراهم کردن داده‌های آموزشی
۲. ایجاد شبکه بهینه
۳. آموزش شبکه
۴. شبیه‌سازی شبکه با داده‌های جدید

۲-۸-۵-۱-۱- شبکه‌های پیش‌خور[2]

شبکه‌های پیش‌خور اغلب یک یا چند لایه مخفی از نرون‌های زیگموئید می‌باشند و از یک لایه پایانی خطی استفاده می‌کنند. شبکه‌های چندلایه 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 خارج از محدوده ۱+ و۱- داشته باشد. در شکل ۲-۱۰، یک شبکه دو لایه tansig/purelin نشان داده شده است (شکل ۲-۱۰). این شبکه می‌تواند برای تقریب زدن هر تابع با تعداد محدود ناپیوستگی استفاده شود.

۱.Newton methods
۲.Feed forward



- شکل ۲-۱۰- شبکه دو لایه tansig/purelin

- ایجاد یک شبکه پیش‌خور: اولین مرحله در طراحی یک شبکه پیش‌خور، ایجاد شیئ شبکه می‌باشد و تابع newff نیز یک شبکه پیش‌خور ایجاد می‌کند. این تابع چهار ورودی دارد و شبکه تهیه شده را به عنوان خروجی برمی‌گرداند. اولین پارامتر ورودی یک ماتریس۲×R به عنوان مقادیر min و max هر ورودی از R ورودی می‌باشد. پارامتر سوم یک آرایه سلولی است که شامل توابع انتقال مورد استفاده در هر لایه است و پارامتر آخر نام تابع آموزشی مورد استفاده می‌باشد.
- مقداردهی آغازین به وزن‌ها: قبل از آموزش یک شبکه پیش‌خور، مقادیر ابتدایی وزن‌ها و بایاس‌ها باید تعیین گردند. تابع newff به صورت خودکار به وزن‌ها مقدار می‌دهد ولی ممکن است با دستور init آن را به صورت غیر خودکار نیز انجام داد.
- شبیه سازی: تابع sim یک شبکه را شبیه‌سازی می‌کند. این تابع شبکه و بردار ورودی را به عنوان ورودی دریافت کرده و خروجی شبکه را بر می‌گرداند.
- آموزش شبکه: پس از مقداردهی به وزن‌ها و بایاس‌ها نوبت به آموزش شبکه می‌رسد. شبکه می‌تواند برای تقریب توابع، تشخیص الگو، یا طبقه بندی الگوها مورد استفاده قرار گیرد. فرایند آموزش به یک سری مثال‌ها از رفتار مورد انتظار شبکه نیاز دارد که شامل ورودی شبکه (P) و هدف (T) می‌باشد. در طول فرایند آموزش وزن‌ها و بایاس‌ها تنظیم می‌شوند تا تابع کارایی شبکه،

حداقل شود. تابع کارایی پیش‌فرض برای شبکه‌های پیش‌خور، میانگین مجموع مربعات خطاها[۱](mse) می‌باشد.

$$mse = {}^{1}/_{N} \sum_{i=1}^{N} (e_i)^2 = {}^{1}/_{N} \sum_{i=1}^{N} (t - a_i)^2 \qquad \text{(معادله ۲-۱۰)}$$

در معادله فوق، t مقدار خروجی مورد انتظار برای نمونه ورودی i ام و a_i مقدار محاسبه شده توسط شبکه برای آن است[۳۷].

۲-۸-۵-۲- الگوریتم‌های مورد استفاده

در این قسمت نگاهی به الگوریتم‌های مختلف در شبکه‌های پیش‌خور خواهیم داشت. تمامی این توابع از شیب تابع کارایی برای تنظیم وزن‌ها و بایاس‌ها استفاده می‌کنند. این شیب با استفاده از تکنیک پس انتشار که قبلاً به آن اشاره شد تعیین می‌شود. محاسبه پس انتشار از قانون زنجیره‌ای در حساب دیفرانسیل مشتق می‌شود.

۲-۸-۵-۲-۱- الگوریتم پس انتشار

ساده‌ترین اجرای الگوریتم پس‌انتشار، وزن‌ها و بایاس‌های شبکه را در مسیری که در آن تابع کارایی با بیشترین سرعت با گرادیان منفی کاهش می‌یابد، به‌روز می‌نماید. یک تکرار از این الگوریتم می‌تواند به این صورت نوشته شود:

$$x_{k+1} = x_k - \alpha_k g_k \qquad \text{(معادله ۲-۱۱)}$$

که x_k یک بردار از وزن‌ها و بایاس‌های فعلی در k امین تکرار است، g_k شیب فعلی[۱] و α_k سرعت یادگیری[۲] است[۳۷].

۱. Least Mean Square Error

دو روش مختلف برای اجرای این الگوریتم پیاده‌سازی شده است، روش گام به گـام و روش دسـته‌ای کـه قبلاً به آنها اشاره شد.

۲-۸-۵-۲-۲- الگوریتم لونبرگ- مارکوارت[۳]

الگوریتم لونبرگ-مارکوارت به منظور رسیدن به سرعت آموزش در حد ثانیه بدون نیاز به محاسبه ماتریس هسیان[۴] طراحی شد.

وقتی تابع کارآیی شکل مجموع مربعات را دارد (که در شـبکه‌هـای پـیش‌خـور مرسـوم اسـت)، ماتریس هسیان به روش زیر قابل محاسبه است (معادله۲-۲۲):

$$H=J^TJ \qquad \text{(معادله ۲-۱۲)}$$

و گرادیان خطا نیز با معادله ۲-۲۳ محاسبه می‌شود:

$$g=J^Te \qquad \text{(معادله ۲-۱۳)}$$

که در آن J ماتریس ژاکوبین[۵] است که مشتقات اولیـه خطاهـای شـبکه را بـا در نظـر گـرفتن وزن هـا و بایاس‌ها در برمی گیرد. e نیز برداری از خطاهای شبکه است. ماتریس ژاکوبین از طریق تکنیک پس انتشار استاندارد که نسبت به محاسبه ماتریس هسیان خیلی کم تر پیچیده است، قابل محاسبه است.

الگوریتم لونبرگ-مارکوارت از تقریب زیر برای محاسبه ماتریس هسیان استفاده می کند:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T\mathbf{J} + \mu\mathbf{I}]^{-1}\mathbf{J}^T\mathbf{e} \qquad \text{(معادله ۲-۱۴)}$$

---

۱. Current Gradient
۲. Learning rate
۳.Levenberg -Marquardt
۴.Hessian Matrix
۵.Jacobian Matrix

زمانی که مقدار عددی μ صفر باشد، این تابع تبدیل به یک روش نیـوتن بـرای تقریـب مـاتریس هسـیان می‌شود. زمانی که مقدار عددی μ بزرگ باشد، این تابع تبدیل به روش شیب توأم با گام کوچک می‌شـود. روش نیوتن نسبت به روش شیب توأم، دقیق‌تر است. بنابراین μ بعد از هـر گـام موفـق (کـاهش در تـابع کارآیی) کاهش یافته و تنها وقتی که گام آزمایشی تابع کارآیی را افزایش دهد، افزایش داده می شـود. در این روش، تابع کارآیی همیشه در هر تکـراری از الگـوریتم کـاهش مـی‌یابـد. پارامترهـای آموزشـی بـرای الگوریتم لونبرگ- مارکوارت (trainlm) عبارتند از:

- دوره ها (epochs)
- نمایش(show)
- هدف(goal)
- زمان (time)
- گرادیان- مینیمم (min-grad)
- رد شدن-ماکزیمم (max-fail)
- mu
- کاهش- mu (mu-dec)
- افزایش-mu (mu-inc)
- ماکزیمم-mu (mu-max)
- کاهش-حافظه (mem-reduc)

پارامتر mu مقدار آغازین μ است. این مقدار در هر گام که تابع کارآیی کاهش یابد در mu-dec و هر زمان که افزایش یابد در mu-inc ضرب می‌شود. اگر mu از mu-max بزرگ‌تر شود، الگـوریتم متوقف می‌شود. پارامترmem-reduc مقدار حافظه مورد استفاده الگوریتم را کنترل می‌کنـد. ایـن

روش یکی از سریعترین روش‌های پیاده سازی شده در MATLAB می‌باشد و برای یک شبکه متوسط (با چند صد پارامتر موثر) دارای کارایی بسیار بالا می‌باشد[۳۷].

۲-۸-۵-۳- بهبود تعمیم

شبکه‌های پس‌انتشاری که به نحو مطلوب آموزش داده شده‌اند، تمایل دارند که در صورت ارائه ورودی‌هایی که هرگز آنها را ندیده‌اند، پاسخ های منطقی بدهند. این خاصیت تعمیم[۱] نامیده می شود. یکی از مشکلاتی که در طی آموزش شبکه عصبی رخ می دهد برازش اضافی[۲] است، یعنی خطای مربوط به سری آموزشی به یک مقدار خیلی کوچک رسانده می شود، اما زمانی که داده های جدید به شبکه ارائه می شوند، خطا بزرگ می‌شود دو راه حل برای افزایش عمومیت شبکه در جعبه ابزار شبکه‌های عصبی در نظر گرفته شده است: تنظیم[۳] و توقف زود هنگام(در ابتدا)[۴][۳۷].

۲-۸-۵-۳-۱- توقف در ابتدا

در این تکنیک داده‌های در دسترس به سه زیر مجموعه تقسیم می شوند. اولین زیر مجموعه سری آموزشی است، که برای محاسبه گرادیان و به روز کردن وزن‌ها و بایاس‌های شبکه استفاده می‌شود. دومین زیر مجموعه سری اعتبار یا ارزیابی است. خطای مربوط به سری ارزیابی در طی فرآیند آموزش نظارت می‌شود. خطای ارزیابی به طور طبیعی در طول فاز اولیه آموزش کاهش می‌یابد، همان گونه که خطای سری آموزشی کاهش می‌یابد. اما زمانی که شبکه شروع به برازش اضافی داده‌ها بکند، خطای سری ارزیابی شروع به بالا رفتن می‌کند. زمانی که خطای سری ارزیابی برای یک تعداد معین از دوره‌ها افزایش

۱.Generalization
۲.Overfitting
۳.Regularization
۴.Early Stopping

می‌یابد، آموزش متوقف شده و وزن‌ها و بایاس‌ها در مینیمم خطای اعتبار انطباق داده می‌شوند. زیر مجموعه سوم سری تست یا آزمایش است که در طول فرایند آموزش کاربرد ندارد و از آن برای مقایسه مدل‌های متفاوت استفاده می‌شود. توابع مختلفی برای تقسیم داده‌ها به سری‌های آموزش، اعتبار و تست وجود دارند. در این پروژه تمام مجموعه داده‌ها از طریق شاخص‌هایی که در نرم افزار MATLAB موجود است، به سه سری آموزش، اعتبار و تست تقسیم شد.

۲-۸-۵-۳-۲- تنظیم

روش دیگر برای بهبود تعمیم، تنظیم نامیده می‌شود که با اصلاح تابع کارآیی متداول (mse) انجام می‌گیرد. برای تنظیم دو روش وجود دارد: اصلاح توابع کارایی[۱] و تنظیم خودکار[۲].

۲-۸-۵-۳-۳- تنظیم خودکار

تعیین پارامترهای کارایی به صورت خودکار بسیار مطلوب می‌باشد. یک راه دستیابی به این فرایند چهارچوب کاری بائزین[۳] می‌باشد. در این چهارچوب وزن‌ها و بایاس‌های شبکه مقادیر تصادفی با یک توزیع خاص فرض می‌شوند. پارامترهای تنظیم با یک واریانس نامعلوم از این توزیع‌ها مرتبط هستند و می‌توان آنها را با استفاده از تکنیک‌های آماری تخمین زد. تنظیم بائزین درتابع آموزشی trainbr پیاده‌سازی شده است[۳۷].

trainbr یک تابع آموزش برای شبکه است که وزن‌ها و بایاس را بر اساس بهبود الگوریتم لونبرگ-مارکوارت به روز رسانی می‌کند. این تابع ترکیبی خطی از مربع خطاها و نیز وزن‌ها را به حداقل

۱. Performance Function Modification

۲.Automated Regularization

۳.Bayezian Regularization

رسانده و سپس ترکیب صحیحی از آنها را برای ایجاد یک شبکه با قابلیت تعمیم پـذیری خـوب مشـخص می‌کند. یک ویژگی این الگوریتم این است که یک اندازه‌گیری از تعداد وزن‌ها و بایاس‌های موثر در شبکه ارائه می‌دهد. به این ترتیب که در پایان آموزش می‌توان تشخیص داد که چند پـارامتر شـبکه در آمـوزش موثر بوده اند. با افزایش پارامترهای ورودی شبکه، این عدد تقریباً ثابت باقی می‌ماند، به این ترتیب مزیت این الگوریتم نسبت به trainlm قدرت تعمیم‌پذیری بسیار بالای آن می‌باشد.

افزودن تنظیم بایزین به الگوریتم لـونبرگ-مـارکوارت تغییراتـی در آن بـه وجـود مـی‌آورد. ایـن الگوریتم به صورت زیر محاسبه می‌شود:

۱. محاسبه ژاکوبین
۲. محاسبه گرادیان خطا
۳. محاسبه هسیان با استفاده از حاصلضرب ضربدری ژاکوبین
۴. محاسبه تابع ارزش به صورت زیر:

$$C = \beta * E_d + \alpha * E_w$$ (معادله ۲-۱۵)

در معادله بالا E_d جمع مربعات خطا و E_w جمع مربعات وزن‌ها می‌باشد.

۵. حل معادله $(H+\mu I)\ \sigma = g$ برای پیدا کردن σ. σ بردار به روزرسانی وزن‌ها می‌باشد.
۶. به روز رسانی وزن‌های شبکه w با استفاده از σ
۷. محاسبه مجدد تابع ارزش با استفاده از وزن‌های به روز رسانده شده
۸. اگر مقدار ارزش کاهش نیافته بود :

- از وزن‌های جدید صرف نظر می‌شود، مقدار μ کاهش داده می‌شود و به مرحله ۵ برمی‌گردیم.

۹. در غیر این صورت μ کاهش داده می‌شود.

۱۰. پارامترهای دیگر بائزین با استفاده از فرمول‌های مک‌کی و پولند به روزرسانی می‌شوند:

- $\mu= w- (\alpha*tr(H^{-1}))$ (معادله ۲-۱۶)
- $\beta= (N -\mu)/2.0*E_d$ (معادله ۲-۱۷)
- $\alpha=(w/2.0)*E_w+tr (H^{-1})$ [به روز رسانی با فرمول اصلاح شده پولند] (معادله ۲-۱۸)
- $\alpha=\mu/(2.0*E_w)$ [به روز رسانی با فرمول اصلی مک کین] (معادله ۲-۱۹)

در فرمول های بالا W تعداد پارامترهای شبکه (تعداد وزن‌ها و بایاس‌ها)، N تعداد ورودی‌های سری آموزش، μ فاکتور تعدیل لونبرگ-مارکوارت و $tr(H^{-1})$ اثر ماتریس معکوس هسیان است.

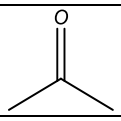
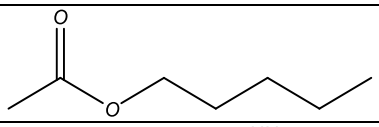
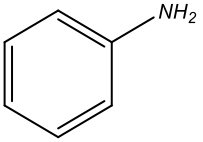
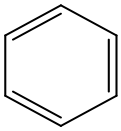
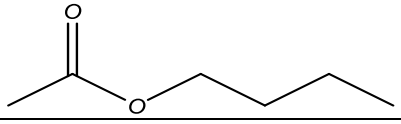
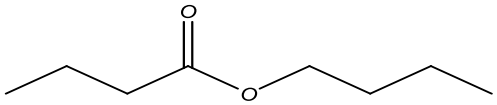
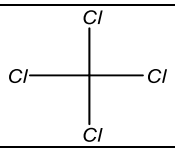
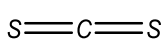
فصل سوم

مدل‌سازی QSPR پارامتر حلالیت حلال‌ها با استفاده از روش رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی(ANN)

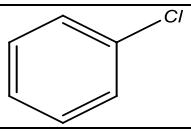
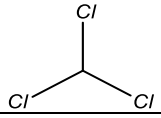
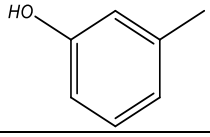
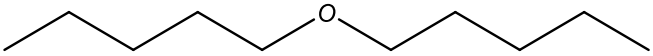
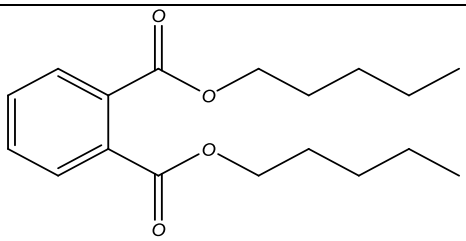
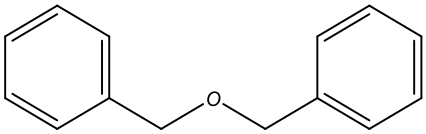
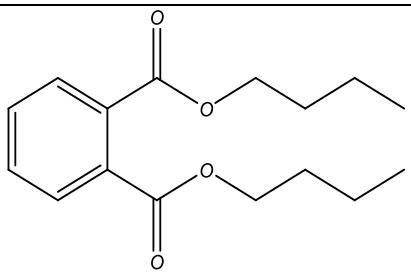
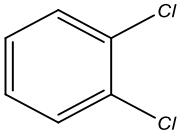
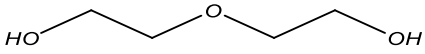
۳-۱- سری داده‌ها

در این قسمت از تحقیق، پارامتر حلالیت ۷۰ ترکیب (حلال پلی‌مر) که دارای گروه‌های عاملی گوناگونی هستند به عنوان سری داده‌ها مورد استفاده قرار گرفت[۳]. نام و ساختار و پارامتر حلالیت این ترکیبات در جدول ۳-۱ آمده است.

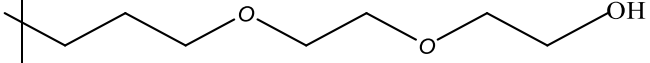
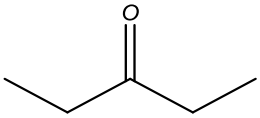
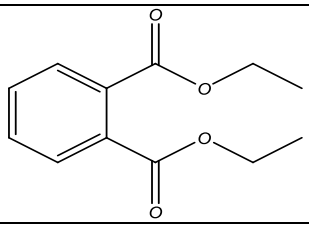
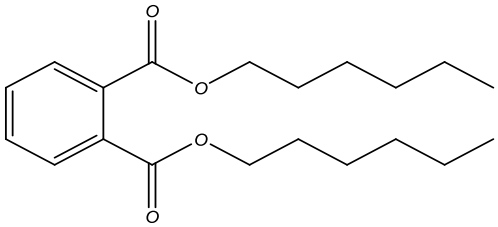
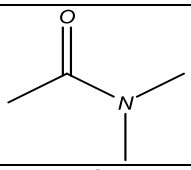
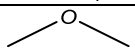
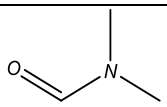
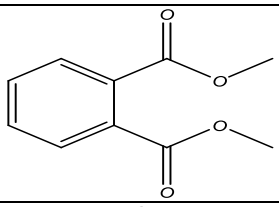
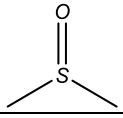
جدول۳-۱- سری داده‌های به کار رفته در مدل‌سازی

شماره ترکیب	حلال	ساختار حلال	پارامتر حلالیت
۱	Acetone		۹/۹
۲	Amyl acetate		۸/۵
۳	Aniline		۱۰/۳
۴	Benzene		۹/۲
۵	Butyl acetate		۸/۳
۶	Butyl butyrate		۸/۱
۷	Carbon tetrachloride		۸/۶
۸	Carbon disulfide		۱۰

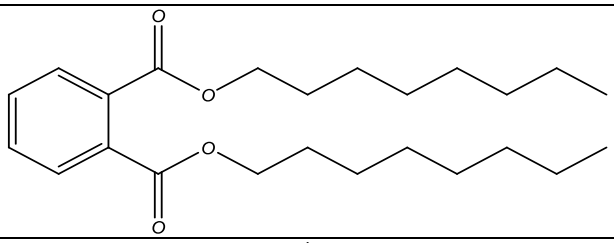
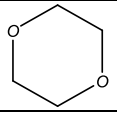
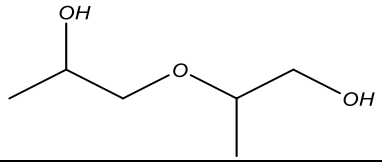
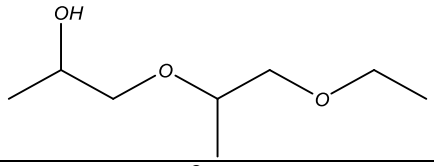
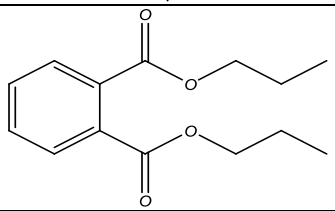
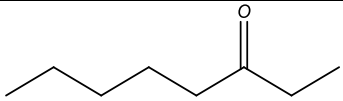
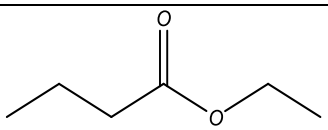
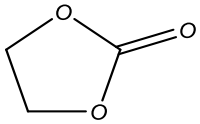
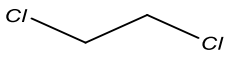
ادامه جدول۳-۱

شماره ترکیب	حلال	ساختار حلال	پارامتر حلالیت
۹	Chlorobenzene		۹/۵
۱۰	Chloroform		۹/۳
۱۱	m-cresol		۱۰/۲
۱۲	Diamyl ether		۷/۳
۱۳	Diamyl phthalate		۹/۱
۱۴	Dibenzyl ether		۹/۴
۱۵	Dibutyl phthalate		۹/۳
۱۶	1,2-Dichlorobenzene		۱۰/۰
۱۷	Diethylene glycol		۱۲/۱

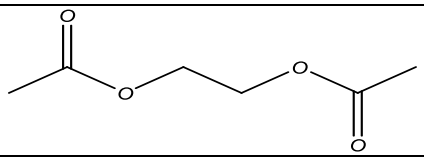
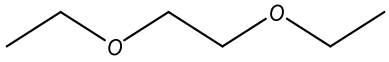
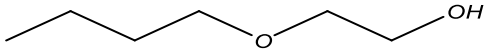
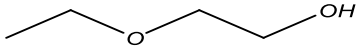
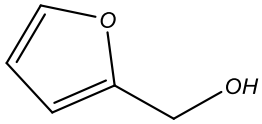
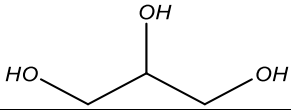
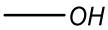
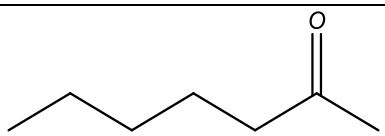
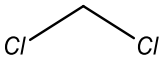
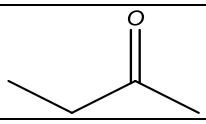
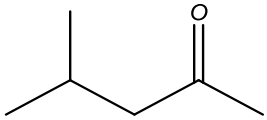
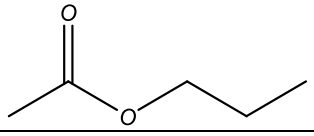
ادامه جدول۳-۱

شماره ترکیب	حلال	ساختار حلال	پارامتر حلالیت
۱۸	Dyethylene glycol monobutyl ether		۹/۵
۱۹	Diethyl ketone		۸/۸
۲۰	Diethyl phthalate		۱۰
۲۱	Di-n-hexyl phthalate		۸/۹
۲۲	N,N-Dimethylacetamide		۱۰/۸
۲۳	Dimethyl ether		۸/۸
۲۴	N,N-Dimethyl formamide		۱۲/۱
۲۵	Dimethyl phthalate		۱۰/۷
۲۶	Dimethyl sulfoxide		۱۲/۰

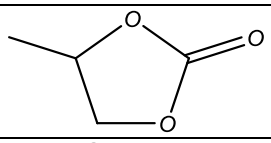
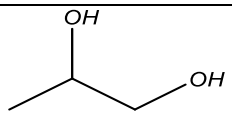
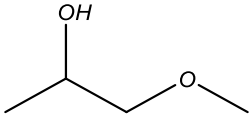
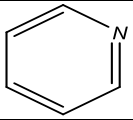
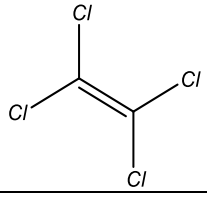
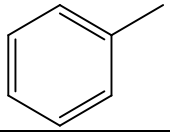
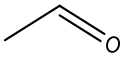
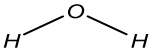
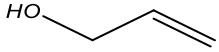
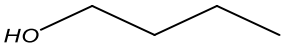
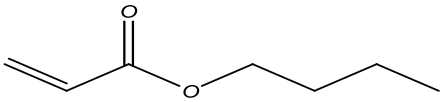
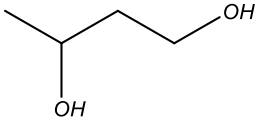
ادامه جدول۳-۱

شماره ترکیب	حلال	ساختار حلال	پارامتر حلالیت
۲۷	Dioctyl phthalate		۷/۹
۲۸	1,4-Dioxane		۱۰
۲۹	Di(propylene glycol)		۱۰
۳۰	Dipropylene glycol monoethyl ether		۹/۳
۳۱	Dipropyl phthalate		۹/۷
۳۲	Ethyl amyl ketone		۸/۲
۳۳	Ethyl butyrate		۸/۵
۳۴	Ethylene carbonate		۱۴/۷
۳۵	Ethylene dichloride		۹/۸

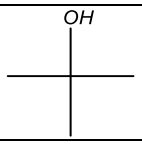
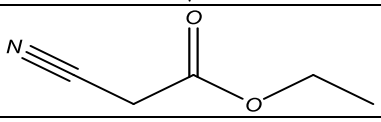
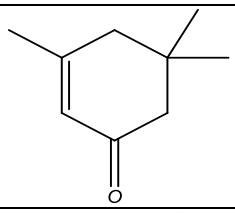
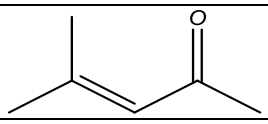
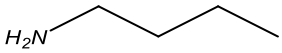
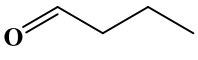
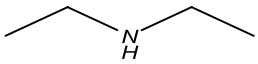
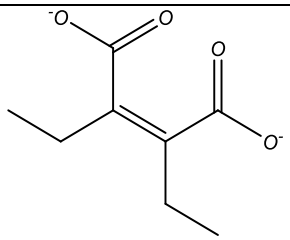
ادامه جدول۳-۱

شماره ترکیب	حلال	ساختار حلال	پارامتر حلالیت
۳۶	Ethylene glycol diacetate		۱۰
۳۷	Ethylene glycol diethyl ether		۸/۳
۳۸	Ethylene glycol monobutyl ether		۹/۵
۳۹	Ethylene glycol monoethyl ether		۱۰/۵
۴۰	Furfuryl alcohol		۱۲/۵
۴۱	Glycerol anhydrous		۱۶/۵
۴۲	Hexane		۷/۳
۴۳	Methanol		۱۴/۵
۴۴	Methyl amyl ketone		۸/۵
۴۵	Methylen chloride		۹/۷
۴۶	Methyl ethyl ketone		۹/۳
۴۷	Methyl isobutyl ketone		۸/۴
۴۸	Propyl acetate		۸/۸

ادامه جدول۳-۱

شماره ترکیب	حلال	ساختار حلال	پارامتر حلالیت
۴۹	1,2-Propylene carbonate		۱۳/۳
۵۰	Propylene glycol		۱۲/۶
۵۱	Propylene glycol methyl ether		۱۰/۱
۵۲	Pyridine		۱۰/۷
۵۳	Tetrachloro ethylene		۹/۳
۵۴	Toluene		۸/۹
۵۵	Acetaldehyde		۱۰/۳
۵۶	Water		۲۳/۴
۵۷	Allyl alcohol		۱۱/۸
۵۸	Butyl alcohol		۱۱/۴
۵۹	N-butyl acrylate		۸/۸
۶۰	1,3-Butane diole		۱۱/۶

ادامه جدول۳-۱

شماره ترکیب	حلال	ساختار حلال	پارامتر حلالیت
۶۱	Tert-butyl alcohol		۱۰/۶
۶۲	Ethyl cyanoacetate		۱۱/۰
۶۳	Ethylene dichloride		۹/۸
۶۴	Isophorone		۹/۱
۶۵	Mesityl oxide		۹
۶۶	n-butyl amine		۸/۷
۶۷	Butyraldehyde		۹
۶۸	Diethylamine		۸
۶۹	diethylmaleate		۹/۹
۷۰	Ethylene glycol		۱۴/۶

۳-۲- رسم و بهینه‌سازی ساختار هندسی و به دست آوردن توصیف‌کننده‌ها

در این مرحله از مطالعه، ابتدا ساختار مولکول‌ها به وسیله نرم افزار HyperChem رسم و ساختار هندسی آنها با احتساب اتم‌های هیدروژن و با استفاده از روش AM1 بهینه شد. بهینه‌سازی تا زمانی ادامه یافت که گرادیان انرژی به ۰/۰۰۱ کیلو کالری بر مول برسد. سپس ساختارهای بهینه شده به عنوان ورودی به نرم‌افزار Dragon وارد شدند. محاسبه توصیف‌کننده‌ها توسط این نرم‌افزار صورت گرفت و ۱۴۸۱ توصیف‌کننده از هیجده گروه مختلف برای هر ترکیب به دست آمد.

۳-۳- انتخاب بهترین توصیف‌کننده‌ها و به دست آوردن مدل نهایی به روش رگرسیون خطی چندگانه (MLR)

توصیف‌کننده‌های به دست آمده از مرحله قبل، به عنوان متغیر مستقل، و پارامتر حلالیت به عنوان متغیر وابسته وارد نرم افزار SPSS شدند. سپس به منظور انتخاب توصیف‌کننده‌های مهم و جلوگیری از طولانی شدن فرایند برازش و مدل‌سازی، روش‌های کاهش متغیر به کار گرفته شد. برای نیل به این هدف، تعدادی از توصیف‌کننده‌ها که دارای بیش از ۹۰٪ مقادیر یکسان و یا تنها دارای ۱۰٪ مقادیر غیر صفر بودند حذف شدند. سپس، از میان توصیف‌کننده‌هایی که علاوه بر همبستگی زیاد با پارامتر وابسته با پارامتر مستقل دیگری همبستگی قابل توجه (بیش از ۰/۹) داشتند، پارامتری که همبستگی بیشتری با پارامتر وابسته مورد نظر داشت نگه داشته و دیگری حذف شد. برای این کار برنامه‌ای در نرم‌افزار MATLAB نوشته و اجرا شد. در نهایت با استفاده از روش انتخاب متغیر مرحله‌ای تعدادی از بهترین توصیف‌کننده‌ها که بیشترین همبستگی را با پارامتر وابسته مورد نظر (پارامتر حلالیت) داشتند، انتخاب شدند. داده ها با نسبت ۲۰: ۲۰: ۶۰ و به سه گروه تقسیم شدند. ۴۲ داده در سری آموزش و ۱۴ داده در هرکدام از سری‌های ارزیابی و تست قرار گرفت. سپس بر اساس این توصیف‌کننده‌ها و به‌کارگیری

روش رگرسیون خطی چندگانه (MLR)، مدل‌سازی برای سری آموزش صورت گرفت. چهار مدل نهایی همراه با توصیف‌کننده‌های وارد شده در مدل و همچنین پارامترهای آماری مربوط در جدول ۳-۲ آورده شده است.

جدول۳-۲- مقایسه آماره‌های چند مدل نهایی به دست آمده از روش MLR

شماره مدل	توصیف کننده‌ها	R	SE	F
۱	Ms, Mor29m, E1v	۰/۹۳۷	۰/۹۹۹۱۷	۹۰/۴۰۸
۲	Ms, Mor29m, ATS1e, E1v	۰/۹۵۶	۰/۸۴۲۹۴	۹۹/۳۶۸
۳	Ms, Mor29m, ATS1e, E1v,X1Av	۰/۹۶۸	۰/۷۳۰۰۹	۱۰۸/۶۳۲
۴	Ms, Mor29m, ATS1e, E1v,X1Av, HATS1u	۰/۹۶۸	۰/۷۴۰۲۳	۸۸/۰۶۷

در جدول بالا آماره R ضریب همبستگی خطی بین پارامتر حلالیت پیش‌بینی شده و مقدار تجربی، آماره SE خطای استاندارد رگرسیون و آماره F مربوط به آزمون فیشر است که در فصل دوم مورد بحث قرار گرفت. این پارامترها به کمک نرم‌افزار SPSS محاسبه شدند.

با بررسی مدل‌های به دست آمده، مدل سوم به خاطر داشتن آماره‌های F و R بزرگتر و خطای استاندارد کمتر انتخاب شد. جدول۳-۳ ضریب رگرسیون، گروه و نام کامل توصیف‌کننده‌ها و اثر متوسط آنها در مدل منتخب را نشان می‌دهد. اثر متوسط یک متغیر مستقل با استفاده از فرمول ۳-۱ به دست می‌آید:

$$\text{MeanEffect} = \frac{\beta_x \times \sum x_n}{\sum SP}$$ معادله (۳-۱)

که در معادله یاد شده، β_x ضریب متغیر مستقل x در مدل MLR، x_n مقدار متغیر مستقل x مورد نظر برای ترکیب n ام و SP نیز پارامتر حلالیت تجربی می‌باشد.

جدول۳-۳- توصیف‌کننده‌های وارد شده در مدل MLR منتخب

شماره	علامت	ضریب رگرسیون	اثر متوسط	گروه	نام کامل
۱	Ms	۱/۴۳۸	۰/۳۹۹۲	Constitutional descriptors	Mean electerotopoligical state
۲	Mor29m	۹/۲۲۲	-۰/۰۴۷۴	3D-MoRSE descriptors	3D-MoRSE –signal 29/wheighted by atomic masses
۳	ATS1e	۱۸/۷۹۲	۱/۹۲۴۵	2D autocorrelations descriptors	Broto-moreau autocorrelation of a topological structure- lag 1/weighted by atomic Sanderson electronegativities
۴	E1v	-۸/۴۵۳۵	-۰/۳۷۶۴	WHIM descriptors	1st component accessibility directional WHIM index
۵	X1Av	۳/۷۲۶۶	۰/۱۵۸۹	topological descriptors	Average valence connectivity index- chi 1

جدول ۳-۴ ماتریس همبستگی توصیف کننده‌های وارد شده در مدل منتخب را نشان می‌دهد. همانطور که ملاحظه می‌شود مقدار همبستگی بین پارامترها قابل قبول (کمتر از ۰/۹) می‌باشد.

جدول۳-۴- ماتریس همبستگی توصیف‌کننده‌های به کار رفته در مدل نهایی

توصیف کننده	Mor29m	ATS1e	E1v	Ms	X1Av
Mor29m	۱				
ATS1e	-۰/۴۸۵	۱			
E1v	-۰/۱۰۱	۰/۲۲۴	۱		
Ms	-۰/۰۴۰	۰/۷۰۷	-۰/۰۰۱	۱	
X1Av	-۰/۳۴۴	-۰/۲۰۳	۰/۲۴۸	-۰/۴۲۴	۱

به این ترتیب معادله خطی نهایی به صورت زیر به دست آمد (معادله۳-۲):

معادله ۳-۲ $y = -10.735 + 1.438\ \mathrm{Ms} + 9.291\ \mathrm{Mor29m} + 18.792\ \mathrm{ATS1e} - 8.4535\ \mathrm{E1v} + 3.7266\ \mathrm{X1Av}$

مقادیر عددی توصیف‌کننده‌های وارد شده در مدل MLR نهایی نیز در جدول ۳-۵ آورده شده است.

جدول۳-۵- مقادیر عددی توصیف کننده‌های وارد شده در مدل MLR نهایی

شماره ترکیب	X1Av	Ms	E1v	ATS1e	Mor29m
۱	۰/۴۰۱	۳/۱۷۰	۰/۳۸۰	۰/۹۹۹	-۰/۰۷۴۰
۲	۰/۴۲۶	۲/۴۶۰	۰/۴۸۰	۱/۰۰۹	-۰/۰۵۴
۳	۰/۳۱۴	۲/۲۴۰	۰/۲۸۵	۱/۰۰۶	۰/۰۵۳
۴	۰/۳۳۳	۲/۰۰۰	۰/۲۷۱	۱/۹۷۲	۰/۰۶۹
۵	۰/۴۱۵	۲/۵۸۰	۰/۴۸۳	۱/۰۱۷	-۰/۰۶۳
۶	۰/۴۴۱	۲/۳۷۰	۰/۵۱۳	۱/۰۰۴	-۰/۱۰۱
۷	۰/۵۶۷	۲/۵۴۰	۰/۳۸۰	۱/۲۶۵	۰/۸۳۰
۸	۰/۶۱۲	۳/۲۸۰	۰/۷۰۴	۱/۰۷۷	۰/۰۰۲
۹	۰/۳۵۴	۲/۲۵۰	۰/۴۱۷	۰/۹۹۹	۰/۰۳۸
۱۰	۰/۶۵۵	۳/۴۲۰	۰/۵۵۲	۱/۱۸۵	۰/۴۴۷
۱۱	۰/۳۱۸	۲/۴۲۰	۰/۳۰۱	۱/۰۱۲	۰/۰۹۶
۱۲	۰/۴۹۹	۱/۷۷۰	۰/۵۲۵	۰/۹۸۲	-۰/۰۵۹
۱۳	۰/۳۷	۲/۳۵۰	۰/۴۱۴	۱/۰۱۱	۰/۰۳۳
۱۴	۰/۳۱۹	۱/۹۹۰	۰/۵۱۲	۰/۹۶۶	۰/۱۴۶
۱۵	۰/۳۵۷	۲/۴۳۰	۰/۳۷۲	۱/۰۱۸	۰/۰۵۱
۱۶	۰/۳۷	۲/۴۴۰	۰/۴۸۶	۱/۰۲۶	-۰/۰۷۰
۱۷	۰/۳۶۸	۳/۰۷۰	۰/۴۱۹	۱/۰۸۷	۰/۰۵۷
۱۸	۰/۴۱۸	۲/۳۲۰	۰/۵۰۳	۱/۰۳۴	۰/۰۱۸
۱۹	۰/۴۶۵	۲/۶۱۰	۰/۴۲۲	۰/۹۸۵	-۰/۰۸۷
۲۰	۰/۳۲۱	۲/۶۷۰	۰/۳۳۰	۱/۰۴	-۰/۱۲۱
۲۱	۰/۳۸۱	۲/۲۸۰	۰/۴۲۰	۱/۰۰۶	-۰/۱۲۴
۲۲	۰/۳۶۴	۲/۷۸۰	۰/۳۸۵	۱/۰۲۳	۰/۰۴۳
۲۳	۰/۴۰۸	۲/۵۰	۰/۵۲۱	۱/۰۴۱	۰/۰۵۹
۲۴	۰/۳۴۷	۳/۰۰۰	۰/۴۰۰	۱/۰۳۹	۰/۰۸۰
۲۵	۰/۲۸۳	۲/۸۳۰	۰/۴۰۷	۱/۰۵۹	۰/۰۴۰

ادامه جدول۳-۵

شماره ترکیب	X1Av	Ms	E1v	ATS1e	Mor29m
۲۶	۰/۹۸۳	۳/۰۶۰	۰/۳۶۰	۱/۰۲۸	-۰/۲۰۳
۲۷	۰/۳۹۸	۲/۱۷۰	۰/۴۴۱	۰/۹۹۸	-۰/۰۵۶
۲۸	۰/۳۵۹	۲/۱۷۰	۰/۴۶۳	۱/۰۶۲	-۰/۰۲۴
۲۹	۰/۴۰۱	۲/۷۲۰	۰/۴۵۳	۱/۰۵۳	-۰/۰۱۰
۳۰	۰/۳۹۹	۲/۴۰	۰/۴۷۱	۱/۰۴۳	۰/۰۰۹
۳۱	۰/۳۴۱	۲/۵۴۰	۰/۴۶۷	۱/۰۲۷	-۰/۱۲۹
۳۲	۰/۴۶۵	۲/۲۸۰	۰/۴۲۶	۰/۹۷۶	-۰/۰۸۰
۳۳	۰/۴۲۴	۲/۵۸۰	۰/۴۱۶	۱/۰۱۷	-۰/۱۱۲
۳۴	۰/۲۸۲	۳/۱۱۰	۰/۳۶۳	۱/۱۴۳	۰/۱۹۲
۳۵	۰/۷۰۱	۲/۸۱۰	۰/۷۴۱	۱/۰۴۴	-۰/۱۵۱
۳۶	۰/۳۲۲	۳/۱۳۰	۰/۵۵۲	۱/۰۷۵	-۰/۰۲۱
۳۷	۰/۴۳۸	۲/۱۳۰	۰/۵۱۴	۱/۰۲۶	۰/۰۳۱
۳۸	۰/۴۴۳	۲/۳۸۰	۰/۴۷۰	۱/۰۲۵	-۰/۰۰۷
۳۹	۰/۴۲۰	۲/۶۷۰	۰/۴۵۱	۱/۰۵۰	۰/۰۳۲
۴۰	۰/۲۹۵	۲/۶۷۰	۰/۳۷۵	۱/۰۷۴	۰/۱۵۹
۴۱	۰/۳۴۱	۳/۷۲۰	۰/۳۳۳	۱/۱۱۴	۰/۰۲۸
۴۲	۰/۵۸۳	۱/۶۷۰	۰/۴۶۳	۰/۹۵۹	۰/۰۴۸
۴۳	۰/۴۴۷	۴/۰۰۰	۰/۳۵۳	۱/۰۸۴	۰/۰۴۰
۴۴	۰/۴۶۶	۲/۳۳۰	۰/۴۶۴	۰/۹۷۸	-۰/۰۴۲
۴۵	۰/۸۰۲	۳/۲۴۰	۰/۷۹۶	۱/۱۰۵	-۰/۱۶۶
۴۶	۰/۴۴۱	۲/۸۳۰	۰/۴۳۱	۰/۹۹۰	-۰/۰۹۰
۴۷	۰/۴۳۷	۲/۵۰	۰/۴۰۶	۰/۹۸۱	-۰/۰۸۰
۴۸	۰/۴۰۱	۲/۷۴۰	۰/۵۱۹	۱/۰۲۷	-۰/۰۲۲
۴۹	۰/۳۰۳	۲/۹۳۰	۰/۴۰۲	۱/۱۰۱	۰/۱۲۳
۵۰	۰/۳۹۰	۲/۳۷۰	۰/۳۴۷	۱/۰۷۰	۰/۰۱۹
۵۱	۰/۳۸۸	۲/۷۲۰	۰/۴۴۲	۱/۰۵۰	۰/۰۳۶
۵۲	۰/۳۰۸	۲/۱۷۰	۰/۲۸۵	۱/۰۰۴	۰/۰۶۱
۵۳	۰/۵۰۴	۳/۳۰۰	۰/۷۹۵	۱/۲۱۲	-۰/۳۲۶
۵۴	۰/۳۴۴	۱/۹۵۰	۰/۳۳۶	۰/۹۷۰	۰/۰۶۲

ادامه جدول۳-۵

شماره ترکیب	X1Av	Ms	E1v	ATS1e	Mor29m
۵۵	۰/۴۰۷	۳/۶۷۰	۰/۵۵۷	۱/۰۱۸	-۰/۰۳۴
۵۶	۰/۰۰۰	۱۰/۰۰۰	۰/۴۳۵	۱/۲۵۶	۰/۰۰۸
۵۷	۰/۳۷۸	۳/۱۳۰	۰/۳۳۷	۱/۰۳۴	۰/۰۱۷
۵۸	۰/۵۰۶	۲/۵۰۰	۰/۳۸۲	۱/۰۰۶	-۰/۰۳۳
۵۹	۰/۴۳۳	۲/۴۶۰	۰/۴۹	۱/۰۰۹	-۰/۰۷۸
۶۰	۰/۴۱۲	۳/۰۶۰	۰/۳۵۹	۱/۰۴۸	-۰/۰۲۴
۶۱	۰/۴۳۱	۲/۶۵۰	۰/۳۰۸	۱/۰۰۶	-۰/۰۴۶
۶۲	۰/۳۳۴	۳/۲۱۰	۰/۵۴۹	۱/۰۵۴	۰/۰۱۹
۶۳	۰/۷۰۱	۲/۸۱۰	۰/۷۴۱	۱/۰۴۴	-۰/۰۰۶
۶۴	۰/۳۷۰	۲/۲۶۰	۰/۳۳۳	۰/۹۸۱	-۰/۰۲۸
۶۵	۰/۳۸۰	۲/۶۲۰	۰/۴۱۶	۰/۹۸۶	-۰/۰۷۲
۶۶	۰/۵۲۹	۲/۱۰۰	۰/۴۱۶	۰/۹۹۰	-۰/۰۱۸
۶۷	۰/۴۶۳	۲/۸۰۰	۰/۴۵۳	۰/۹۹۰	-۰/۰۶۱
۶۸	۰/۵۳۰	۱/۹۰۰	۰/۴۸۷	۰/۹۹۱	-۰/۰۵
۶۹	۰/۳۳۸	۲/۹۴۰	۰/۴۳۵	۱/۰۵۷	-۰/۰۲۱
۷۰	۰/۳۷۷	۳/۷۵۰	۰/۳۰۰	۱/۱۰۵	-۰/۰۳۳

با قرار دادن مقادیر عددی توصیف‌کننده‌های مربوط به هر سری در معادله فوق، می‌توان مقادیر پیش‌بینی شده را با مقادیر تجربی مقایسه کرد. در جداول ۳-۶، ۳-۷ و ۳-۸ مقادیر محاسبه شده، مقادیر تجربی، مقدار خطای مطلق و درصد خطای نسبی به ترتیب برای سری آموزش، ارزیابی و تست آورده شده‌اند. مقدار خطای مطلق و درصد خطای نسبی به ترتیب با استفاده از معادلات ۳-۳ و ۳-۴ به دست آمده‌اند. در معادلات زیر E خطای مطلق، x_i مقدار محاسبه شده برای پارامتر مورد نظر، x_t مقدار تجربی پارامتر و E_r درصد خطای نسبی می‌باشد.

$$E = x_i - x_t$$ (معادله ۳-۳)

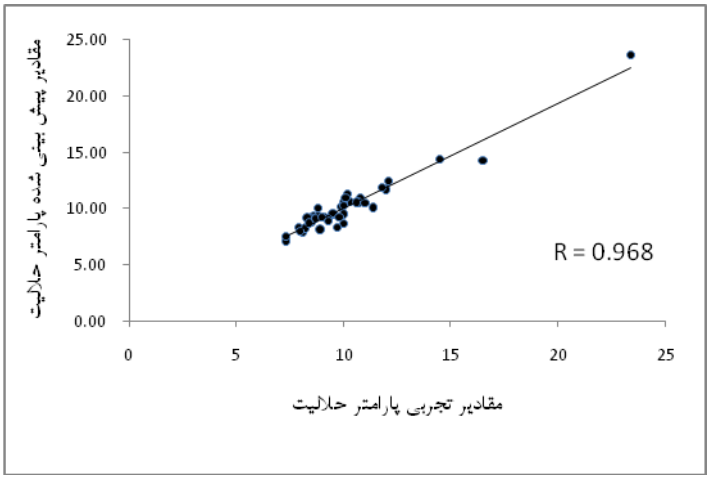
$$E_r = \frac{E}{x_t} \times 100$$ (معادله ۳-۴)

جدول۳-۶- مقایسه مقادیر تجربی و محاسبه شده پارامتر حلالیت با روش MLR برای سری آموزش

شماره ترکیب	مقدار واقعی	مقدار محاسبه شده	خطای مطلق	درصد خطای نسبی
۱	۹/۹	۱۰/۱۹	۰/۲۹	۲/۹۳
۲	۸/۵	۸/۷۹	۰/۲۹	۳/۴۱
۳	۱۰/۳	۱۰/۶۵	۰/۳۵	۳/۳۰
۶	۸/۱	۷/۹۱	-۰/۱۹	-۲/۳۴
۷	۸/۶	۹/۳۲	۰/۷۲	۸/۳۷
۸	۱۰	۱۰/۵۷	۰/۵۷	۵/۷۰
۱۱	۱۰/۲	۱۱/۳۰	۱/۱۰	۱۰/۷۸
۱۲	۷/۳	۷/۱۴	-۰/۱۶	-۲/۱۹
۱۳	۹/۱	۹/۲۲	۰/۱۲	۱/۳۲
۱۶	۱۰	۸/۶۸	-۱/۳۲	-۱۳/۲۰
۱۷	۱۲/۱	۱۲/۴۷	۰/۳۷	۳/۰۶
۱۸	۹/۵	۹/۵۱	۰/۰۱	۰/۱۰
۲۱	۸/۹	۸/۱۷	-۰/۷۳	-۸/۲۰
۲۲	۱۰/۸	۱۰/۹۹	۰/۱۹	۱/۷۶
۲۳	۸/۸	۱۰/۰۹	۱/۲۹	۱۴/۶۶
۲۶	۱۲	۱۱/۷۲	-۰/۲۸	-۲/۳۳
۲۷	۷/۹	۸/۳۸	۰/۴۸	۶/۰۸
۲۸	۱۰	۹/۵۴	-۰/۴۶	-۴/۵
۳۱	۹/۷	۸/۳۴	-۱/۳۶	-۱۴/۰۲
۳۲	۸/۲	۸/۲۷	۰/۰۷	۰/۸۵
۳۳	۸/۵	۹/۱۱	۰/۶۱	۷/۱۸
۳۶	۱۰	۱۰/۳۱	۰/۳۱	۳/۱۰
۳۷	۸/۳	۹/۱۸	۰/۸۸	۱۰/۶۰
۳۸	۹/۵	۹/۵۶	۰/۰۶	۰/۶۳
۴۱	۱۶/۵	۱۴/۲۷	-۲/۲۳	-۱۳/۵۲
۴۲	۷/۳	۷/۵۰	۰/۲۰	۲/۷۴
۴۳	۱۴/۵	۱۴/۴۴	-۰/۰۶	-۰/۴۱
۴۶	۹/۳	۹/۱۰	۰/۲۰	۲/۱۵
۴۷	۸/۴	۸/۷۵	۰/۳۵	۴/۱۷

ادامه جدول۳-۶

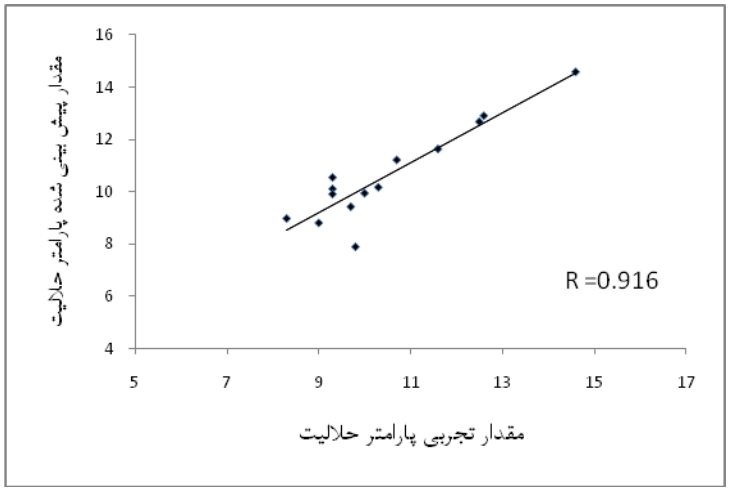
شماره ترکیب	مقدار واقعی	مقدار محاسبه شده	خطای مطلق	درصد خطای نسبی
۴۸	۸/۸	۹/۴۱	۰/۶۱	۶/۹۳
۵۱	۱۰/۱	۱۰/۹۵	۰/۸۵	۸/۴۲
۵۲	۱۰/۷	۱۰/۵۶	-۰/۱۴	-۱/۳۱
۵۳	۹/۳	۸/۹۲	-۰/۳۸	-۴/۰۹
۵۶	۲۳/۴	۲۳/۶۵	۰/۲۵	-۰/۵۶
۵۷	۱۱/۸	۱۱/۹۲	۰/۱۲	۱/۰۲
۵۸	۱۱/۴	۱۰/۱۲	-۱/۲۸	-۱۱/۲۳
۶۱	۱۰/۶	۱۰/۵۶	-۰/۰۴	-۰/۳۸
۶۲	۱۱	۱۰/۴۷	-۰/۵۳	-۴/۸۲
۶۳	۹/۸	۹/۲۲	-۰/۵۸	-۵/۹۲
۶۶	۸/۷	۹/۱۸	۰/۴۸	۵/۵۲
۶۷	۹	۹/۲۳	۰/۲۳	۲/۵۶
۶۸	۸	۸/۰۱	۰/۰۱	۰/۱۲



شکل۳-۱- رسم مقادیر محاسبه شده توسط روش MLR بر حسب مقادیر تجربی برای سری آموزش

برای سری ارزیابیMLRجدول۳-۷- مقایسه مقادیر تجربی و پیش‌بینی شده پارامتر حلالیت با روش

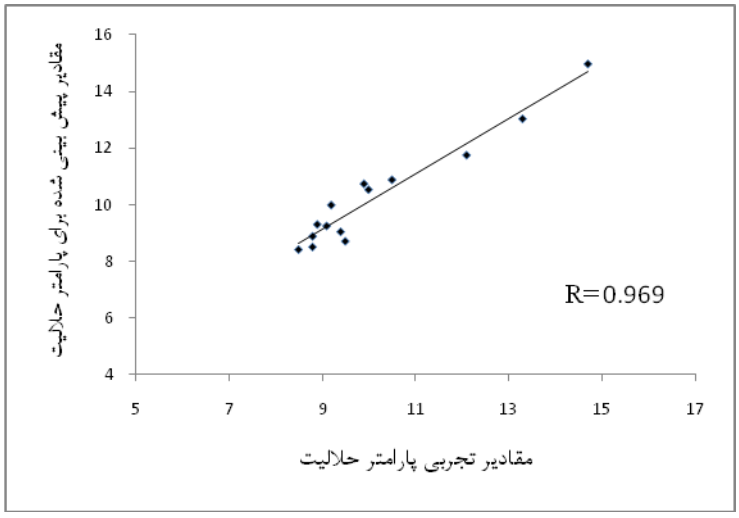
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۵	۸/۳	۸/۹۷	۰/۶۷	۸/۰۷
۱۰	۹/۳	۱۰/۱۰	۰/۸۰	۸/۶۰
۱۵	۹/۳	۱۰/۵۵	۱/۲۵	۱۳/۴۴
۲۰	۱۰	۹/۹۴	-۰/۰۶	-۰/۶۰
۲۵	۱۰/۷	۱۱/۲۲	۰/۵۲	۴/۸۶
۳۰	۹/۳	۹/۹۰	۰/۶۰	۶/۴۵
۳۵	۹/۸	۷/۸۸	-۱/۹۲	-۱۹/۵۹
۴۰	۱۲/۵	۱۲/۶۸	۰/۱۸	۱/۴۴
۴۵	۹/۷	۹/۴۲	-۰/۲۸	-۲/۸۹
۵۰	۱۲/۶	۱۲/۹۲	۰/۳۲	۲/۵۴
۵۵	۱۰/۳	۱۰/۱۷	-۰/۱۳	-۱/۲۶
۶۰	۱۱/۶	۱۱/۶۴	۰/۰۴	۰/۳۴
۶۵	۹	۸/۸۰	-۰/۲۰	-۲/۲۲
۷۰	۱۴/۶	۱۴/۶۰	۰/۰	۰/۰



شکل۳-۲- رسم مقادیر پیش‌بینی شده توسط روش MLR بر حسب مقادیر تجربی برای سری ارزیابی

جدول۳-۸- مقایسه مقادیر تجربی و پیش‌بینی شده پارامتر حلالیت با روش MLR برای سری تست

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۴	۹/۲	۹/۹۹	۰/۷۹	۸/۵۹
۹	۹/۵	۸/۷۲	-۰/۷۸	-۸/۲۱
۱۴	۹/۴	۹/۰۵	-۰/۳۵	-۳/۷۲
۱۹	۸/۸	۸/۸۹	۰/۰۹	۱/۲۳
۲۴	۱۲/۱	۱۱/۷۶	-۰/۳۴	-۲/۸۱
۲۹	۱۰	۱۰/۵۴	۰/۵۴	۵/۴۰
۳۴	۱۴/۷	۱۴/۹۷	۰/۲۷	۱/۸۴
۳۹	۱۰/۵	۱۰/۸۸	۰/۳۸	۳/۶۲
۴۴	۸/۵	۸/۴۲	-۰/۰۸	-۰/۹۴
۴۹	۱۳/۳	۱۳/۰۴	-۰/۲۶	-۱/۹۵
۵۴	۸/۹	۹/۳۱	۰/۴۱	۴/۶۱
۵۹	۸/۸	۸/۵۲	-۰/۲۸	-۳/۱۸
۶۴	۹/۱	۹/۲۶	۰/۱۶	۱/۷۶
۶۹	۹/۹	۱۰/۷۴	۰/۸۵	۸/۵۸



شکل۳-۳- رسم مقادیر پیش‌بینی شده توسط روش MLR بر حسب مقادیر تجربی برای سری تست

همانطور که شکل‌های ۳-۱ تا ۳-۳ مشاهده می‌شود، ضریب همبستگی برای نمودار مقادیر پیش-بینی شده توسط روش MLR در مقابل مقادیر تجربی برای سری آموزش، ارزیابی و تست به ترتیب ۰/۹۶۸، ۰/۹۱۶و ۰/۹۶۸ می‌باشد که نشان‌دهنده توانایی روش MLR در پیش‌بینی پارامتر حلالیت ترکیبات مورد نظر است.

۳-۳-۱- ارزیابی مدل به روش حذف مرحله‌ای تک تک داده‌ها (LOO)

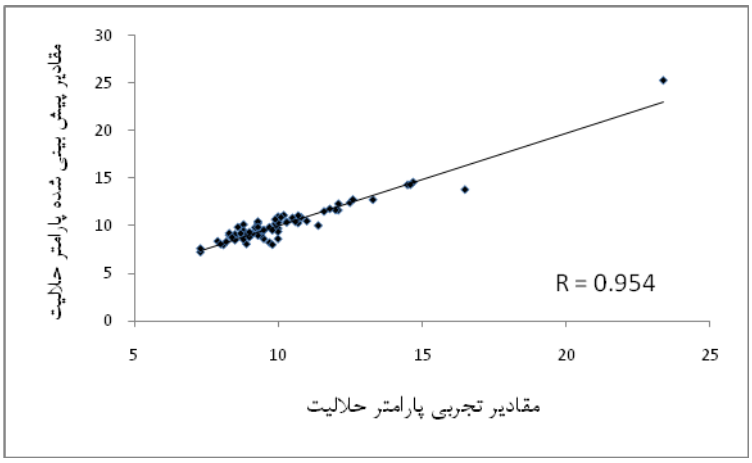
برای ارزیابی مدل به دست آمده علاوه بر استفاده از سری تست، روش حذف مرحله‌ای تک تک داده‌ها نیز به‌کار گرفته شد. در این روش هر بار یکی از ملکول‌ها از سری داده‌ها حذف و سپس مدل‌سازی با ۶۹ ملکول باقیمانده صورت گرفت. این مدل برای پیش‌بینی پارامتر حلالیت ملکول حذف شده به کار گرفته شد. مقادیر پیش‌بینی‌شده با مقادیر تجربی مقایسه و نمودار مربوطه نیز رسم شد (جدول ۳-۹ و شکل ۳-۴). نزدیک بودن این نتایج به نتایج به دست آمده از مدل منتخب نشان دهنده صحت مدل می‌باشد.

جدول۳-۹- ارزیابی مدل خطی به روش حذف مرحله‌ای تک تک داده‌ها

شماره ترکیب	مقدار واقعی	مقدار پیش-بینی شده	خطای مطلق	درصد خطای نسبی
۱	۹/۹	۱۰/۱۸	۰/۲۸	۲/۸۳
۲	۸/۵	۸/۸۰	۰/۳۰	۳/۵۳
۳	۱۰/۳	۱۰/۳۸	۰/۰۸	۰/۷۸
۴	۹/۲	۹/۷۹	۰/۵۹	۶/۴۱
۵	۸/۳	۸/۹۸	۰/۶۸	۸/۱۹
۶	۸/۱	۷/۹۶	-۰/۱۴	-۱/۷۳
۷	۸/۶	۹/۸۳	۱/۲۳	۱۴/۳۰
۸	۱۰/۰	۱۰/۹۵	۰/۹۵	۹/۵۰
۹	۹/۵	۸/۵۹	-۰/۹۰	-۹/۴۷
۱۰	۹/۳	۱۰/۲۵	۰/۹۵	۱۰/۲۲
۱۱	۱۰/۲	۱۱/۰۹	۰/۸۹	۸/۷۲
۱۲	۷/۳	۷/۱۹	-۰/۱۱	-۱/۵۱
۱۳	۹/۱	۹/۱۲	۰/۰۲	۰/۲۲
۱۴	۹/۴	۸/۹۸	-۰/۴۲	-۴/۴۷
۱۵	۹/۳	۱۰/۴۱	۱/۱۱	۱۱/۹۳
۱۶	۱۰/۰	۸/۵۹	-۱/۴۱	-۱۴/۱۰
۱۷	۱۲/۱	۱۲/۳۰	۰/۲۰	۱/۶۵
۱۸	۹/۵	۹/۴۶	-۰/۰۳	-۰/۳۲
۱۹	۸/۸	۸/۹۱	۰/۱۱	۱/۲۵
۲۰	۱۰/۰	۹/۷۰	-۰/۳۰	-۳/۰۰
۲۱	۸/۹	۸/۰۵	-۰/۸۵	-۹/۵۵
۲۲	۱۰/۸	۱۰/۸۷	۰/۰۷	۰/۶۵
۲۳	۸/۸	۱۰/۱۱	۱/۳۱	۱۴/۸۹
۲۴	۱۲/۱	۱۱/۶۲	-۰/۴۸	-۳/۹۷
۲۵	۱۰/۷	۱۱/۰۵	۰/۳۵	۳/۲۷
۲۶	۱۲/۰	۱۱/۶۴	-۰/۳۵	-۲/۹۲
۲۷	۷/۹	۸/۳۴	۰/۴۴	۵/۵۷
۲۸	۱۰/۰	۹/۳۵	-۰/۶۴	-۶/۴۰
۲۹	۱۰/۰	۱۰/۴۶	۰/۴۶	۴/۶۰
۳۰	۹/۳	۹/۸۳	۰/۵۳	۵/۷۰
۳۱	۹/۷	۸/۲۰	-۱/۵۰	-۱۵/۴۶
۳۲	۸/۲	۸/۲۸	۰/۰۸	۰/۹۸
۳۳	۸/۵	۹/۰۷	۰/۵۷	۶/۷۰
۳۴	۱۴/۷	۱۴/۵۹	-۰/۱۱	-۰/۷۵
۳۵	۹/۸	۷/۹۸	-۱/۸۲	-۱۸/۵۷
۳۶	۱۰/۰	۱۰/۳۱	۰/۳۱	۳/۱۰

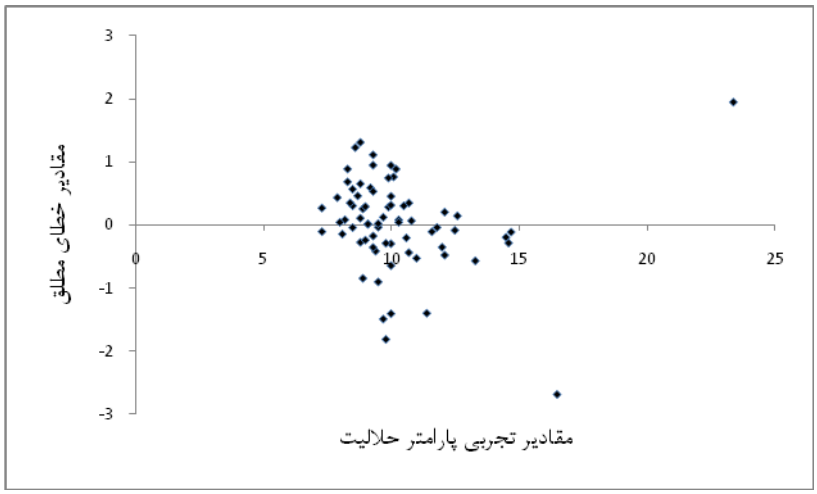
ادامه جدول۳-۹

شماره ترکیب	مقدار واقعی	مقدار پیش-بینی شده	خطای مطلق	درصد خطای نسبی
۳۷	۸/۳	۹/۱۹	۰/۸۹	۱۰/۷۲
۳۸	۹/۵	۹/۵۲	۰/۰۲	۰/۲۱
۳۹	۱۰/۵	۱۰/۸۰	۰/۳۰	۲/۸۶
۴۰	۱۲/۵	۱۲/۴۱	-۰/۰۹	-۰/۷۲
۴۱	۱۶/۵	۱۳/۸۱	-۲/۶۹	-۱۶/۳۰
۴۲	۷/۳	۷/۵۷	۰/۲۷	۳/۷۰
۴۳	۱۴/۵	۱۴/۳۰	-۰/۲۰	-۱/۳۸
۴۴	۸/۵	۸/۴۶	-۰/۰۴	-۰/۴۷
۴۵	۹/۷	۹/۸۲	۰/۱۲	۱/۲۴
۴۶	۹/۳	۹/۱۲	-۰/۱۷	-۱/۸۳
۴۷	۸/۴	۸/۷۵	۰/۳۵	۴/۱۷
۴۸	۸/۸	۹/۴۵	۰/۶۵	۷/۳۹
۴۹	۱۳/۳	۱۲/۷۳	-۰/۵۷	-۴/۲۸
۵۰	۱۲/۶	۱۲/۷۴	۰/۱۴	۱/۱۱
۵۱	۱۰/۱	۱۰/۸۶	۰/۷۶	۷/۵۲
۵۲	۱۰/۷	۱۰/۲۶	-۰/۴۴	-۴/۱۱
۵۳	۹/۳	۸/۹۴	-۰/۳۶	-۳/۸۷
۵۴	۸/۹	۹/۱۶	۰/۲۶	۲/۹۲
۵۵	۱۰/۳	۱۰/۳۴	۰/۰۴	۱۰/۳۹
۵۶	۲۳/۴	۲۵/۳۵	۱/۹۵	۸/۳۳
۵۷	۱۱/۸	۱۱/۷۶	-۰/۰۴	-۰/۳۴
۵۸	۱۱/۴	۱۰/۰۰	-۱/۴۰	-۱۲/۲۸
۵۹	۸/۸	۸/۵۲	-۰/۲۸	-۳/۱۸
۶۰	۱۱/۶	۱۱/۴۹	-۰/۱۱	-۰/۹۵
۶۱	۱۰/۶	۱۰/۳۹	-۰/۲۱	-۱/۹۸
۶۲	۱۱/۰	۱۰/۴۷	-۰/۵۳	-۴/۸۲
۶۳	۹/۸	۹/۵۱	-۰/۲۹	-۲/۹۶
۶۴	۹/۱	۹/۱۱	۰/۰۱	۰/۱۱
۶۵	۹/۰	۸/۷۶	-۰/۲۴	-۲/۶۷
۶۶	۸/۷	۹/۱۶	۰/۴۶	۵/۲۹
۶۷	۹/۰	۹/۲۹	۰/۲۹	۳/۲۲
۶۸	۸/۰	۸/۰۴	۰/۰۴	۰/۵۰
۶۹	۹/۹	۱۰/۶۵	۰/۷۵	۷/۵۸
۷۰	۱۴/۶	۱۴/۳۱	-۰/۲۹	-۱/۹۹



شکل۳-۴- رسم منحنی مقادیر پیش‌بینی شده در ارزیابی مدل MLR با روش LOO در مقابل مقادیر تجربی پارامتر حلالیت

همچنین نمودار مقادیر خطای مطلق در ارزیابی مدل MLR بر حسب مقادیر تجربی پارامتر حلالیت نیز در شکل ۳-۵ رسم شد. تقارن نسبی داده‌ها حول خط صفر نشان‌دهنده عدم وجود خطای معین می‌باشد.



شکل۳-۵- رسم مقادیر خطای مطلق در ارزیابی مدل MLR با روش LOO در مقابل مقادیر تجربی پارامتر حلالیت

۳-۴- مدل‌سازی با استفاده از شبکه عصبی مصنوعی(ANN)

از آنجاییکه ممکن است مدل خطی بهترین مدل برای پیش‌بینی پارامتر حلالیت نباشد، شبکه عصبی مصنوعی برای مدل‌سازی غیرخطی مورد استفاده قرار گرفت. به این منظور برنامه نویسی و اجرای یک شبکه دولایه در محیط نرم افزار MATLAB نسخه 2008b صورت گرفت. در این روش تابع کارایی شبکه، میانگین مجذور خطای استاندارد (mse) می‌باشد. همانند روش MLR، داده‌ها به صورت تصادفی و با نسبت ۶۰:۲۰:۲۰ به سه سری آموزش، ارزیابی و تست تقسیم شدند. سری آموزش برای آموزش شبکه و سری ارزیابی برای بررسی قدرت تعمیم‌پذیری مدل به دست آمده از سری آموزش به کار گرفته شد. سری تست نیز به عنوان داده‌های خارجی برای ارزیابی نهایی شبکه در نظر گرفته شد (بخش۲-۶-۳). برای به دست آوردن بهترین معادله و کمترین خطا، پارامترهای شبکه از جمله تابع آموزش، تابع انتقال، تعداد گره‌ها در لایه مخفی، پارامتر mu و تعداد دورهای آموزش بهینه‌سازی شدند. قابل ذکر است که به دلیل کم بودن تعداد توصیف کننده‌های وارد شده به مدل MLR منتخب، همان تعداد توصیف‌کننده به عنوان ورودی‌های شبکه نیز در نظر گرفته شدند.

۳-۴-۱- بهینه‌سازی تابع آموزش، تابع انتقال و تعداد نرون‌های لایه مخفی

تعداد گره‌ها از ۲ تا ۱۰ برای سه تابع انتقال tansig، purelin و logsig با هر دو تابع آموزش trainlm و trainbr تغییر داده شد (جدول ۳-۱۰). نقطه‌ای که کمترین خطا را برای سری ارزیابی داشت به عنوان نقطه بهینه انتخاب گردید. همانطور که در این جدول ملاحظه می‌شود، شبکه‌هایی با تابع آموزش trainbr و تابع انتقال logsig با تعداد گره ۴ و ۵ نتایج بهتری نشان می‌دهند. با وجود اینکه مقدار mse برای تعداد گره‌های ۴ و ۵ یکسان می‌باشد، اما به دلیل اینکه افزایش گره‌ها باعث پیچیده شدن شبکه می‌گردد، تعداد ۴ نرون در لایه مخفی به عنوان تعداد بهینه نرون‌های شبکه انتخاب شد.

جدول ۳-۱۰- مقادیر خطای شبکه با تغییر تعداد نرون‌ها

تابع آموزش	تابع انتقال لایه مخفی	تعداد نرون‌ها	mse سری ارزیابی
trainlm	tansig	۲	۰/۴۶۵۶
		۳	۰/۵۷۰۶
		۴	۱۳/۳۲۰۱
		۵	۸۴/۴۶۵۵
		۶	۸۰/۷۰۲۵
		۷	۲۰/۶۱۹۳
		۸	۳۶۷/۰۲۹۲
		۹	۱۲۹/۴۴۵۳
		۱۰	۴۳/۴۹۲۷
	logsig	۲	۰/۴۶۵۹
		۳	۴/۵۰۱۴
		۴	۲/۶۰۴۷
		۵	۲۹۲/۱۴۵۷
		۶	۳۶/۱۲۵۱
		۷	۷/۹۸۵۵
		۸	۳/۹۱۲۴
		۹	۱۴/۰۹۲۴
		۱۰	۶/۲۷۴۳
	purelin	۲	۰/۵۰۰۲
		۳	۰/۵۰۰۲
		۴	۰/۵۰۰۲
		۵	۰/۵۰۰۲
		۶	۰/۵۰۰۲
		۷	۰/۵۰۰۲
		۸	۰/۵۰۰۲
		۹	۰/۵۰۰۲
		۱۰	۰/۵۰۰۲
trainbr	tansig	۲	۰/۳۷۹۷
		۳	۰/۳۷۸۶
		۴	۰/۳۷۸۳

ادامه جدول ۳-۱۰

mse سری ارزیابی	تعداد نرون‌ها	تابع انتقال لایه مخفی	تابع آموزش
۰/۳۷۷۹	۵	tansig	trainbr
۰/۵۸۹۵	۶		
۰/۳۷۷۷	۷		
۰/۳۷۷۶	۸		
۰/۳۷۷۵	۹		
۰/۳۷۷۵	۱۰		
۰/۳۸۳۴	۲	logsig	
۰/۳۷۷۷	۳		
۰/۳۷۷۱	۴		
۰/۳۷۷۱	۵		
۰/۳۷۷۵	۶		
۰/۳۷۷۸	۷		
۰/۳۷۸۱	۸		
۰/۵۸۸۷	۹		
۰/۵۸۹۰	۱۰		
۰/۴۹۹۷	۲	purelin	
۰/۵۰۰۳	۳		
۰/۴۹۹۷	۴		
۰/۴۹۹۷	۵		
۰/۵۰۷۴	۶		
۰/۴۹۹۷	۷		
۰/۵۰۰۸	۸		
۰/۴۹۹۷	۹		
۰/۴۹۹۷	۱۰		

۳-۴-۲- بهینه‌سازی پارامتر mu

در این مرحله تابع آموزش trainbr، تابع انتقال logsig و تعداد ۴ نرون در لایه مخفی بـه عنـوان تعداد بهینه نرون‌های شبکه قرار داده شد و برای بهینه سازی mu مقدار این پارامتر بین ۰/۰۰۰۱ تا ۰/۱

تغییر داده شد و سپس برای هر مورد مقدار mse سری‌های ارزیابی محاسبه گردید. نقطه ای که کمترین خطا را برای سری آموزش داشت به عنوان مقدار بهینه انتخاب گردید. در این مورد مقدار mu بهینه ۰/۰۰۱ به دست آمد (جدول۳-۱۱).

جدول۳-۱۱- مقادیر خطای شبکه با تغییر mu

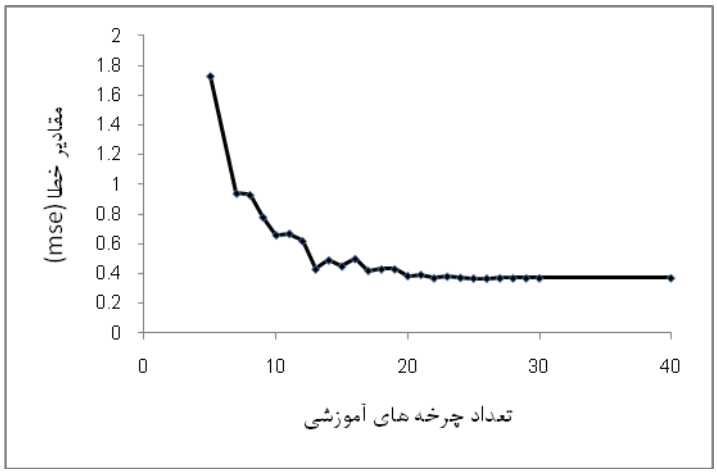
mse سری ارزیابی	mu	mse سری ارزیابی	mu
۰/۳۷۷۷	۰/۰۰۶	۰/۳۷۷۳	۰/۰۰۰۱
۰/۳۷۷۹	۰/۰۰۷	۰/۳۷۷۶	۰/۰۰۰۲
۰/۳۷۷۲	۰/۰۰۸	۰/۳۷۶۴	۰/۰۰۰۳
۰/۳۷۷۳	۰/۰۰۹	۰/۳۷۷۲	۰/۰۰۰۴
۰/۳۷۲۳	۰/۰۱	۰/۳۷۷۰	۰/۰۰۰۵
۰/۳۷۶۸	۰/۰۲	۰/۳۷۷۲	۰/۰۰۰۶
۰/۳۷۲۳	۰/۰۳	۰/۳۷۷۱	۰/۰۰۰۷
۰/۳۷۲۳	۰/۰۴	۰/۳۷۷۱	۰/۰۰۰۸
۰/۳۷۲۳	۰/۰۵	۰/۳۷۲۳	۰/۰۰۰۹
۰/۳۷۲۳	۰/۰۶	۰/۳۷۲۳	۰/۰۰۱
۱/۰۳۴۱	۰/۰۷	۰/۳۷۶۵	۰/۰۰۲
۰/۳۷۲۳	۰/۰۸	۰/۳۷۶۶	۰/۰۰۳
۰/۳۷۲۳	۰/۰۹	۰/۳۷۷۴	۰/۰۰۴
۰/۳۷۲۳	۰/۱	۰/۳۷۷۹	۰/۰۰۵

۳-۴-۳- بهینه سازی تعداد چرخه‌های آموزشی

تعداد چرخه‌های آموزشی پارامتری است که برای جلوگیری از آموزش بیش از حد شبکه بهینه‌سازی می‌شود. برای این منظور، ابتدا مقدار بهینه تعداد نرون‌ها (۴) و مقدار بهینه مومنتم (۰/۰۰۱) در شبکه مورد نظر با تابع آموزش trainbr و تابع انتقال logsig قرار داده شد و سپس منحنی تعداد چرخه‌ها بر حسب خطای مجذور میانگین شبکه رسم گردید (جدول۳-۱۲ و شکل۳-۶).

جدول ۳-۱۲- مقادیر خطای شبکه با تغییر تعداد چرخه‌های آموزشی

تعداد چرخه‌های آموزشی	mse سری ارزیابی	تعداد چرخه‌های آموزشی	mse سری ارزیابی
۵	۱/۷۳۰	۱۹	۰/۴۳۰
۷	۰/۹۴۰	۲۰	۰/۳۸۴
۸	۰/۹۳۰	۲۱	۰/۳۹۳
۹	۰/۷۸۰	۲۲	۰/۳۷۰
۱۰	۰/۶۶۰	۲۳	۰/۳۸۱
۱۱	۰/۶۷۰	۲۴	۰/۳۷۳
۱۲	۰/۶۲۰	۲۵	۰/۳۶۹
۱۳	۰/۴۳۰	۲۶	**۰/۳۶۸**
۱۴	۰/۴۹۰	۲۷	۰/۳۷۰
۱۵	۰/۴۵۰	۲۸	۰/۳۶۹
۱۶	۰/۵۰۰	۲۹	۰/۳۷۰
۱۷	۰/۴۲۰	۳۰	۰/۳۷۰
۱۸	۰/۴۳۰	۴۰	۰/۳۷۱



شکل ۳-۶- منحنی مقدار مربع میانگین خطاها در برابر تغییر تعداد چرخه‌های آموزشی

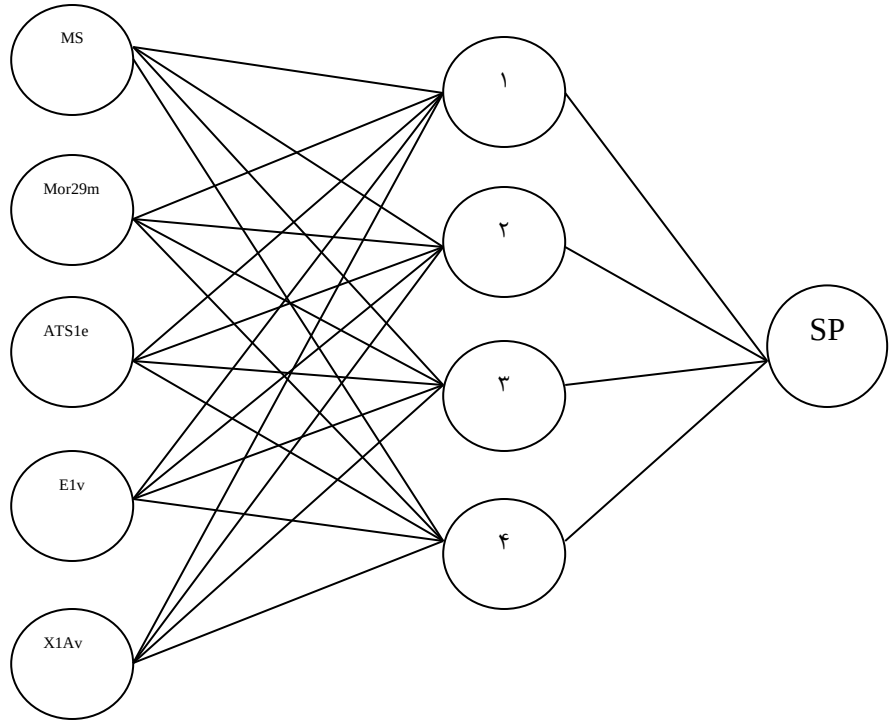
با توجه به نتایج سری ارزیابی، تعداد بهینه چرخه‌های آموزشی ۲۶ انتخاب گردید.

مشخصات شبکه بهینه در جدول ۳-۱۳ آورده شده است.

جدول۳-۱۳- مشخصات شبکه بهینه

تابع آموزش	trainbr
تابع انتقال لایه پنهان	logsig
تابع انتقال لایه خروجی	purelin
تابع کارآیی	mse
تابع مقدار اولیه دهی وزن و بایاس	rands
بهبود تعمیم	Early Stopping
تعداد نرون های لایه پنهان	۴
پارامترmu	۰/۰۰۱
تعداد چرخه‌های آموزشی	۲۶

ساختار شبکه عصبی به دست آمده پس از بهینه‌سازی فاکتورهای بحث شده در بالا به طور شماتیک در شکل۳-۷ نشان داده شده است.



شکل ۳-۷- تصویر شماتیک ساختار شبکه عصبی مصنوعی به دست آمده پس از بهینه سازی

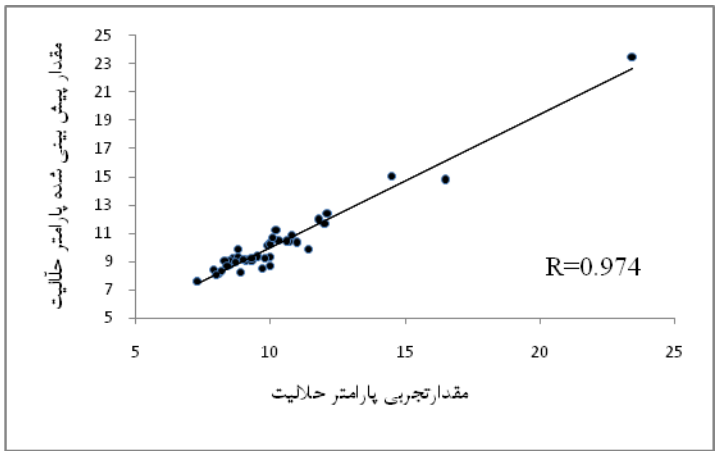
با استفاده از شبکه بهینه مقادیر پارامتر حلالیت برای سری ارزیابی و تست نیز محاسبه شد. در جداول ۳-۱۴، ۳-۱۵ و ۳-۱۶ به ترتیب مقادیر پیش‌بینی شده برای سری آموزش، سری ارزیابی و سری تست آورده شده است. نمودارهای مقادیر پیش‌بینی شده در مقابل مقادیر تجربی پارامتر حلالیت برای سری آموزش، ارزیابی و تست نیز به ترتیب در شکل‌های ۳-۸، ۳-۹ و ۳-۱۰ رسم شد. همانطور که در شکل‌های ۳-۸ تا ۳-۱۰ مشاهده می‌شود، ضریب همبستگی برای نمودار مقادیر پیش‌بینی شده توسط روش ANN در مقابل مقادیر تجربی برای سری آموزش، ارزیابی و تست به ترتیب ۰/۹۷۴، ۰/۹۵۰ و ۰/۹۷۵ می‌باشد که نشان‌دهنده توانایی روش ANN در پیش‌بینی پارامتر حلالیت ترکیبات مورد نظر است.

جدول ۳-۱۴- مقایسه مقادیر تجربی و پیش‌بینی شده پارامتر حلالیت برای سری آموزش با روش ANN

شماره ترکیب	مقدار واقعی پارامتر حلالیت	مقدار پیش‌بینی شده پارامتر حلالیت	خطای مطلق	درصد خطای نسبی
۱	۹/۹	۱۰/۱۶	۰/۲۶	۲/۶۳
۲	۸/۵	۸/۷۸	۰/۲۸	۳/۲۹
۳	۱۰/۳	۱۰/۵۰	۰/۲۰	۱/۹۴
۶	۸/۱	۸/۱۵	۰/۰۵	۰/۶۲
۷	۸/۶	۹/۲۳	۰/۶۳	۷/۳۲
۸	۱۰	۱۰/۳۲	۰/۳۲	۳/۲۰
۱۱	۱۰/۲	۱۱/۲۱	۱/۰۱	۹/۹
۱۲	۷/۳	۷/۵۷	۰/۲۷	۳/۷۰
۱۳	۹/۱	۹/۱۰	۰	۰
۱۶	۱۰	۸/۷۲	-۱/۲۸	-۱۲/۸۰
۱۷	۱۲/۱	۱۲/۴۲	۰/۳۲	۲/۶۴
۱۸	۹/۵	۹/۳۲	-۰/۱۸	-۱/۹۱
۲۱	۸/۹	۸/۲۵	-۰/۶۵	-۷/۳۰
۲۲	۱۰/۸	۱۰/۸۵	۰/۰۵	۰/۴۶
۲۳	۸/۸	۹/۸۵	۱/۰۵	۱۱/۹۳
۲۶	۱۲	۱۱/۷۰	-۰/۳۰	-۲/۵
۲۷	۷/۹	۸/۴۰	۰/۵۰	۶/۳۳
۲۸	۱۰	۹/۳۲	-۰/۶۸	-۶/۸

ادامه جدول ۳-۱۴

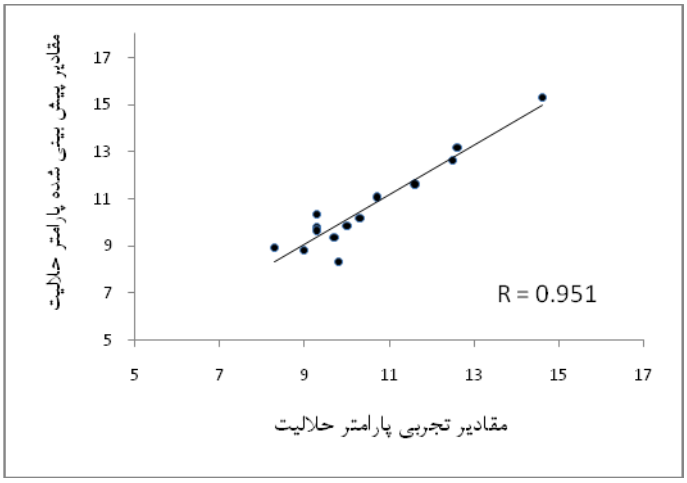
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۳۱	۹/۷	۸/۴۹	-۱/۲۱	-۱۲/۴۷
۳۲	۸/۲	۸/۳۰	۰/۱۰	۱/۲۲
۳۳	۸/۵	۹/۰۲	۰/۵۲	۶/۱۲
۳۶	۱۰	۱۰/۲۰	۰/۲۰	۲/۰۰
۳۷	۸/۳	۹/۰۳	۰/۷۳	۸/۸۰
۳۸	۹/۵	۹/۳۶	-۰/۱۴	-۱/۴۷
۴۱	۱۶/۵	۱۴/۸۳	-۱/۶۷	-۱۰/۱۲
۴۲	۷/۳	۷/۶۲	۰/۳۲	۴/۳۸
۴۳	۱۴/۵	۱۵/۰۸	۰/۵۸	۴/۰۰
۴۶	۹/۳	۹/۰۷	-۰/۱۳	-۱/۴
۴۷	۸/۴	۸/۷۲	۰/۳۲	۳/۸۱
۴۸	۸/۸	۹/۳۴	۰/۵۴	۶/۱۳
۵۱	۱۰/۱	۱۰/۷۲	۰/۶۲	۶/۱۴
۵۲	۱۰/۷	۱۰/۴۰	-۰/۳۰	-۲/۸۰
۵۳	۹/۳	۹/۲۳	-۰/۰۷	-۰/۷۵
۵۶	۲۳/۴	۲۳/۴۷	۰/۰۷	۰/۳۰
۵۷	۱۱/۸	۱۲/۰۰	۰/۲۰	۱/۶۹
۵۸	۱۱/۴	۹/۸۹	-۱/۵۱	-۱۳/۲۴
۶۱	۱۰/۶	۱۰/۴۷	-۰/۱۳	-۱/۲۳
۶۲	۱۱	۱۰/۳۶	-۰/۶۴	-۵/۸۲
۶۳	۹/۸	۹/۲۱	-۰/۵۹	-۶/۰۲
۶۶	۸/۷	۸/۹۴	۰/۲۴	۲/۷۶
۶۷	۹	۹/۱۶	۰/۱۶	۱/۷۷
۶۸	۸	۸/۰۶	۰/۰۶	۰/۷۵



شکل۳-۸- رسم مقادیر پیش‌بینی شده پارامتر حلالیت توسط روش ANN بر حسب مقادیر تجربی برای سری آموزش

جدول ۳-۱۵- مقایسه مقادیر تجربی و پیش‌بینی شده برای سری ارزیابی با روش ANN

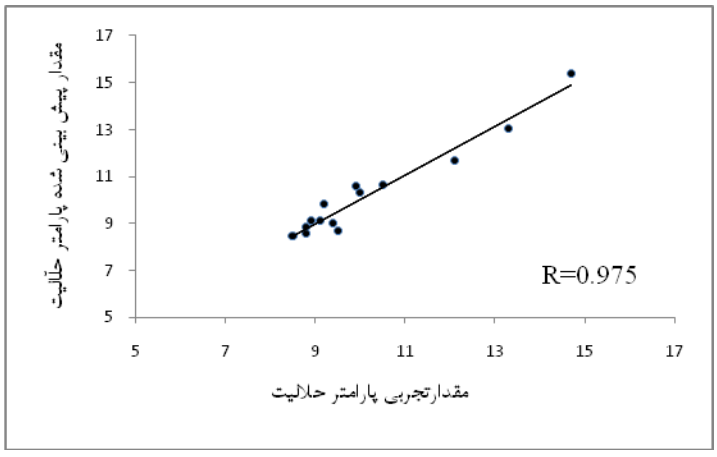
شماره ترکیب	مقدار تجربی پارامتر حلالیت	مقدار پیش‌بینی شده پارامتر حلالیت	خطای مطلق	درصد خطای نسبی
۵	۸/۳	۸/۹۴	۰/۶۴	۷/۷۱
۱۰	۹/۳	۹/۸۲	۰/۵۲	۵/۵۹
۱۵	۹/۳	۱۰/۳۴	۱/۰۴	۱۱/۱۸
۲۰	۱۰	۹/۸۴	-۰/۱۶	-۱/۶۰
۲۵	۱۰/۷	۱۱/۰۸	۰/۳۸	۳/۵۵
۳۰	۹/۳	۹/۶۷	۰/۳۷	۳/۹۸
۳۵	۹/۸	۸/۳۵	-۱/۲۵	-۱۲/۷۶
۴۰	۱۲/۵	۱۲/۶۵	۰/۱۵	۱/۲۰
۴۵	۹/۷	۹/۳۶	-۰/۳۴	-۳/۵۰
۵۰	۱۲/۶	۱۳/۱۶	۰/۵۶	۴/۴۴
۵۵	۱۰/۳	۱۰/۱۷	-۰/۱۳	-۱/۲۶
۶۰	۱۱/۶	۱۱/۶۳	۰/۰۳	۰/۲۶
۶۵	۹	۸/۸۱	-۰/۱۹	-۲/۱۱
۷۰	۱۴/۶	۱۵/۲۸	۰/۶۸	۴/۶۶



شکل۳-۹- رسم مقادیر پیش‌بینی شده پارامتر حلالیت توسط روش ANN بر حسب مقادیر تجربی برای سری ارزیابی

جدول ۳-۱۶- مقایسه مقادیر مورد انتظار و پیش‌بینی شده سری تست با روش ANN

شماره ترکیب	مقدار تجربی پارامتر حلالیت	مقدار پیش‌بینی شده پارامتر حلالیت	خطای مطلق	درصد خطای نسبی
۴	۹/۲	۹/۸۲	۰/۶۲	۶/۷۴
۹	۹/۵	۸/۶۹	-۰/۸۱	-۸/۵۳
۱۴	۹/۴	۹/۰۱	-۰/۳۹	-۴/۱۵
۱۹	۸/۸	۸/۸۴	۰/۰۴	۰/۴۵
۲۴	۱۲/۱	۱۱/۶۹	-۰/۴۱	-۳/۳۹
۲۹	۱۰	۱۰/۳۱	۰/۳۱	۳/۱۰
۳۴	۱۴/۷	۱۵/۳۹	۰/۶۹	۴/۶۹
۳۹	۱۰/۵	۱۰/۶۳	۰/۱۳	۱/۲۴
۴۴	۸/۵	۸/۴۵	۰/۰۵	۰/۵۹
۴۹	۱۳/۳	۱۳/۰۶	-۰/۲۴	-۱/۸۰
۵۴	۸/۹	۹/۱۴	۰/۲۴	۲/۷۰
۵۹	۸/۸	۸/۵۷	-۰/۲۳	-۲/۶۱
۶۴	۹/۱	۹/۱۳	۰/۰۳	۰/۳۳
۶۹	۹/۹	۱۰/۵۹	۰/۶۹	۶/۹۷



شکل۳-۱۰- رسم مقادیر پیش‌بینی شده پارامتر حلالیت توسط روش ANN بر حسب مقادیر تجربی برای سری تست

۳-۴-۴- ارزیابی مدل ANN

۳-۴-۴-۱- ارزیابی شبکه به روش حذف مرحله‌ای تک تک داده‌ها

برای ارزیابی مدل، علاوه بر به‌کارگیری سری تست، همانطور که قبلاً گفته شد روش حذف مرحله‌ای تک تک داده‌ها نیز مورد استفاده قرار گرفت (بخش ۳-۳-۱) و نتایج آن با نتایج محاسبه شده توسط مدل غیرخطی برای سری داده‌ها مقایسه گردید. در جدول ۳-۱۷ مقدار واقعی، پیش‌بینی شده، خطای مطلق و درصد خطای نسبی برای سری داده‌ها مشاهده می‌شود. در شکل ۳-۱۱ نیز منحنی مقادیر پیش‌بینی شده پارامتر حلالیت در ارزیابی مدل ANN بر حسب مقادیر تجربی برای سری داده‌ها رسم شده است.

جدول۳-۱۷- ارزیابی مدل غیر خطی به روش حذف مرحله‌ای تک تک داده‌ها

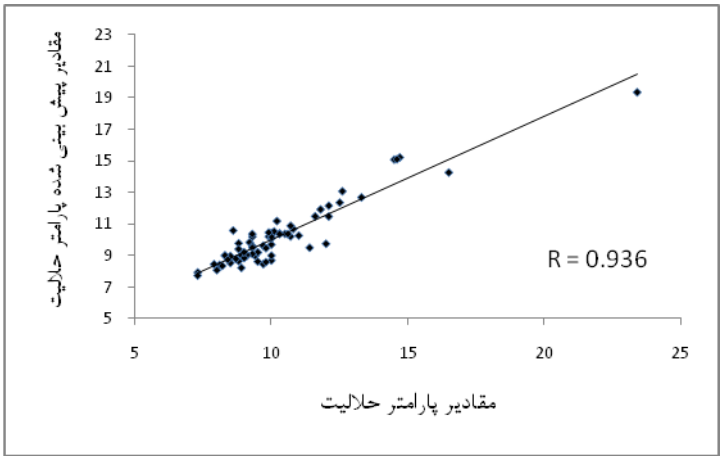
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۱	۹/۹	۱۰/۲۱	۰/۳۱	۳/۱۳
۲	۸/۵	۸/۸۳	۰/۳۳	۱/۱۶
۳	۱۰/۳	۱۰/۴۲	۰/۱۲	۷/۰۶
۴	۹/۲	۹/۸۵	۰/۶۵	۸/۴۳
۵	۸/۳	۹/۰۰	۰/۷۰	
۶	۸/۱	۸/۴۲	۰/۳۲	۳/۹۵
۷	۸/۶	۱۰/۶۰	۲/۰۰	۲۳/۲۶
۸	۱۰/۰	۱۰/۴۵	۰/۴۵	۴/۵۰
۹	۹/۵	۸/۶۳	-۰/۸۷	-۱۲
۱۰	۹/۳	۹/۶۲	۰/۳۲	۳/۴۴
۱۱	۱۰/۲	۱۱/۲۰	۱/۰۰	۹/۸۰
۱۲	۷/۳	۷/۹۴	۰/۶۴	۸/۷۷
۱۳	۹/۱	۹/۰۰	-۰/۱۰	-۱/۱۰
۱۴	۹/۴	۹/۰۰	-۰/۴۰	-۴/۲۶
۱۵	۹/۳	۱۰/۲۱	۰/۹۱	۹/۷۸
۱۶	۱۰/۰	۸/۷۰	-۱/۳۰	-۱۳
۱۷	۱۲/۱	۱۲/۱۷	۰/۰۷	۰/۵۸
۱۸	۹/۵	۹/۲۰	-۰/۳۰	-۳/۱۶
۱۹	۸/۸	۸/۸۶	۰/۰۶	۰/۶۸
۲۰	۱۰/۰	۹/۷۱	-۰/۲۹	-۲/۹۰
۲۱	۸/۹	۸/۲۳	-۰/۶۷	-۷/۵۳
۲۲	۱۰/۸	۱۰/۷۰	-۰/۱	-۰/۹۲
۲۳	۸/۸	۹/۷۹	۰/۹۹	۱۱/۲۵
۲۴	۱۲/۱	۱۱/۴۹	-۰/۶۱	-۵/۰۴۱
۲۵	۱۰/۷	۱۰/۹۰	۰/۲۰	۱/۸۷
۲۶	۱۲/۰	۹/۷۶	-۲/۲۴	-۱۸/۶۷
۲۷	۷/۹	۸/۴۶	۰/۵۶	۷/۰۹
۲۸	۱۰/۰	۹/۰۱	-۰/۹۹	-۹/۹
۲۹	۱۰/۰	۱۰/۱۱	۰/۱۱	۱/۱

ادامه جدول۳-۱۷

شماره ترکیب	مقدار واقعی پارامتر حلالیت	مقدار پیش‌بینی شده پارامتر حلالیت	خطای مطلق	درصد خطای نسبی
۳۰	۹/۳	۹/۵۱	۰/۲۱	۲/۲۶
۳۱	۹/۷	۸/۴۷	-۱/۲۳	-۱۲/۶۸
۳۲	۸/۲	۸/۳۶	۰/۱۶	۱/۹۵
۳۳	۸/۵	۸/۹۸	۰/۴۸	۵/۶۵
۳۴	۱۴/۷	۱۵/۲۴	۰/۵۴	۳/۶۷
۳۵	۹/۸	۸/۶۰	-۱/۲۰	-۱۲/۲۴
۳۶	۱۰/۰	۱۰/۲۲	۰/۲۲	۰/۲۲
۳۷	۸/۳	۹/۰۲	۰/۷۲	۸/۶۷
۳۸	۹/۵	۹/۲۲	-۰/۲۸	-۲/۹۵
۳۹	۱۰/۵	۱۰/۳۹	-۰/۱۱	-۱/۰۵
۴۰	۱۲/۵	۱۲/۳۷	-۰/۱۳	-۱/۰۴
۴۱	۱۶/۵	۱۴/۲۸	-۲/۲۲	-۱۳/۴۵
۴۲	۷/۳	۷/۷۴	۰/۴۴	۶/۰۳
۴۳	۱۴/۵	۱۵/۱۰	۰/۶۰	۴/۱۴
۴۴	۸/۵	۸/۵۴	۰/۰۴	۰/۴۷
۴۵	۹/۷	۹/۶۳	-۰/۰۷	-۰/۷۲
۴۶	۹/۳	۹/۱۱	-۰/۱۹	-۲/۰۴
۴۷	۸/۴	۸/۷۵	۰/۳۵	۴/۱۷
۴۸	۸/۸	۹/۴۲	۰/۶۲	۷/۰۴
۴۹	۱۳/۳	۱۲/۷۰	-۰/۶۰	-۴/۵۱
۵۰	۱۲/۶	۱۳/۰۹	۰/۴۹	۳/۸۹
۵۱	۱۰/۱	۱۰/۵۴	۰/۴۴	۴/۳۶
۵۲	۱۰/۷	۱۰/۲۲	-۰/۴۸	-۴/۴۸
۵۳	۹/۳	۱۰/۳۸	۱/۰۸	۱۱/۶۱
۵۴	۸/۹	۹/۰۳	۰/۱۳	۱/۴۶
۵۵	۱۰/۳	۱۰/۳۸	۰/۰۸	۰/۷۸

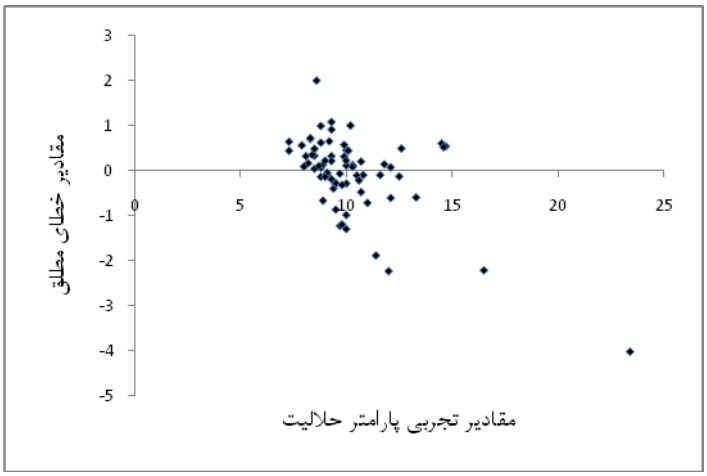
ادامه جدول۳-۱۷

شماره ترکیب	مقدار واقعی پارامتر حلالیت	مقدار پیش‌بینی شده پارامتر حلالیت	خطای مطلق	درصد خطای نسبی
۵۶	۲۳/۴	۱۹/۳۷	-۴/۰۳	-۱۷/۲۲
۵۷	۱۱/۸	۱۱/۹۴	۰/۱۴	۱/۱۹
۵۸	۱۱/۴	۹/۵۱	-۱/۸۹	-۱۶/۵۸
۵۹	۸/۸	۸/۶۶	-۰/۱۴	۱/۵۹
۶۰	۱۱/۶	۱۱/۵۰	-۰/۰۱	-۰/۸۶
۶۱	۱۰/۶	۱۰/۳۸	-۰/۲۲	-۲/۰۸
۶۲	۱۱/۰	۱۰/۲۸	-۰/۷۲	-۶/۵۴
۶۳	۹/۸	۹/۴۸	-۰/۳۲	-۳/۲۶
۶۴	۹/۱	۹/۰۵	-۰/۰۵	-۰/۵۵
۶۵	۹/۰	۸/۸۶	-۰/۱۴	-۱/۵۶
۶۶	۸/۷	۸/۸۰	۰/۱۰	۱/۱۵
۶۷	۹/۰	۹/۲۲	۰/۲۲	۲/۴۴
۶۸	۸/۰	۸/۰۹	۰/۰۹	۱/۱۲
۶۹	۹/۹	۱۰/۴۷	۰/۵۷	۵/۷۶
۷۰	۱۴/۶	۱۵/۱۲	۰/۵۲	۳/۵۶



شکل ۳-۱۱- رسم مقادیر پیش‌بینی شده پارامتر حلالیت در ارزیابی مدل ANN بر حسب مقادیر تجربی برای سری داده‌ها

همچنین نمودار مقادیر خطای مطلق محاسبه شده در ارزیابی مدل ANN بر حسب مقادیر تجربی پارامتر حلالیت نیز در شکل ۳-۱۲ رسم شد. تقارن نسبی داده‌ها حول خط صفر نشان‌دهنده عدم وجود خطای معین می‌باشد.



شکل۳-۱۲- رسم منحنی مقادیر خطای مطلق در ارزیابی مدل ANN به روش LOO در مقابل مقادیر تجربی پارامتر حلالیت

۳-۴-۴-۲- ارزیابی شبکه به روش Y-Randomization

در این روش، با استفاده از برنامه‌ای که در نرم افزار MATLAB نوشته و اجرا شد، مقادیر پاسخ به صورت تصادفی در محدوده پاسخ‌ها تغییر داده شد. سپس همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از شاخص R^2 مورد بررسی قرار گرفت. این کار ده بار تکرار شد. در جدول ۳-۱۸ مقادیر R^2 برای سری ارزیابی و تست آورده شده است. مقادیر پایین R^2 نشان‌دهنده عدم وجود همبستگی شانسی در مدل به دست آمده از شبکه هستند.

جدول۳-۱۸- مقادیر R^2 برای سری ارزیابی و تست در روش Y-Randomization

R^2 برای سری تست	R^2 برای سری ارزیابی	شماره تکرار آزمون
۰/۰۹	۰/۳۳	۱
۰/۰۰	۰/۰۱	۲
۰/۰۱	۰/۰۰	۳
۰/۰۱	۰/۰۰	۴
۰/۱۲	۰/۱۶	۵
۰/۰۹	۰/۰۳	۶
۰/۰۰	۰/۰۱	۷
۰/۰۱	۰/۰۰	۸
۰/۰۱	۰/۰۰	۹
۰/۰۷	۰/۲۴	۱۰

۳-۵- بررسی نتایج و نتیجه‌گیری

در ادامه به بررسی و مقایسه نتایج به دست آمده در این فصل می‌پردازیم.

۳-۵-۱- مقایسه روش های رگرسیون خطی چندگانه و شبکه عصبی مصنوعی

برخی از پارامترهای آماری حاصل از دو روش MLR و ANN در جدول ۳-۱۹ آورده شده اند. مقادیر F و SE با استفاده از نرم‌افزار سیگما پلات به دست آمدند.

جدول ۳-۱۹- شاخص‌های آماری دو روش ANN و MLR در پیش‌بینی پارامتر حلالیت

	R_t	R_{val}	R_{tr}	F_t	F_{val}	F_{tr}	SE_t	SE_{val}	SE_{tr}	R_{cv}
روشMLR	۰/۹۶۹	۰/۹۱۶	۰/۹۶۸	۱۸۲/۱۵۴	۶۴/۲۷۸	۶۰۳/۵۱۴	۰/۴۹۴	۰/۷۵۱	۰/۶۷۱	۰/۹۵۴
روش ANN	۰/۹۷۶	۰/۹۵۱	۰/۹۷۴	۲۴۱/۶۹۵	۱۱۲/۹۰۶	۷۵۴/۳۲۹	۰/۴۵۰	۰/۶۲۵	۰/۶۰۴	۰/۹۳۶

در جدول فوق R_{tr}،R_{val}، R_t به ترتیب مقادیر R برای سری آموزش، ارزیابی و تست و F_{tr}، F_{val}، F_t به ترتیب مقادیر F برای سری آموزش، ارزیابی و تست هستند. SE_{tr}،SE_{val} ، SE_t نیز به ترتیب خطای استاندارد برای سری آموزش، ارزیابی و تست هستند. آنچه از این جدول استنباط می‌شود، با توجه به ضرایب همبستگی بزرگتر و خطای استاندارد کوچکتر سری آموزش، ارزیابی و تست برای مدل شبکه عصبی مصنوعی، برتری این روش بر روش رگرسیون خطی چندگانه در پیش‌بینی پارامتر حلالیت می‌باشد. نتایج به دست آمده برای سری تست، نسبت به مرجع شماره ۹ صحت بیشتری داشته و همینطور تعداد توصیف کننده های به دست آمده کمتر است.

۳-۵-۲- بررسی ارتباط توصیف‌کننده‌های ساختاری وارد شده در مدل با پارامتر حلالیت حلال‌ها

در این قسمت با توجه به توصیف‌کننده‌های وارد شده در مدل‌های MLR و ANN یک بررسی اجمالی روی اثرات مختلف موجود در پارامتر حلالیت ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده شامل پنج توصیف کننده است که هرکدام بیانگر خصوصیات متفاوتی از ملکول‌های مورد بررسی است. در ابتدا لازم است قبل از بررسی توصیف‌کننده‌ها، مختصری در مورد نظریه گراف شرح داده شود. گراف ملکولی را با علامت $G = G(V, E)$ تعریف می‌کنند که در این رابطه V به عنوان رئوس که نشان‌دهنده تعداد اتم‌های موجود در ساختار مولکول هستند، و E به عنوان یال‌ها که نشان‌دهنده پیوندهای متصل‌کننده این رئوس (اتم‌ها) به یکدیگر هستند، معرفی می‌شوند.

تعداد یال‌های متصل به یک رأس را درجه آن رأس می‌نامند. یک گراف مولکولی که بدون درنظرگرفتن هیچ یک از اتم‌های هیدروژن به دست بیاید، گراف مولکولی تهی‌شده از هیدروژن[۱] نامیده می‌شود. با استفاده از نظریه گراف و با در نظر گرفتن موقعیت رئوس مولکول نسبت به یکدیگر می‌توان ساختار یک مولکول را به صورت یک ماتریس نشان داد. این ماتریس با افزایش تعداد اتم‌ها در مولکول (رئوس) و نیز با بزرگتر شدن مولکول و ورود حلقه‌ها به ساختار آن و همچنین پرشاخه‌تر شدن مولکول پیچیده‌تر می‌شود[۲۰].

۱. Hydrogen depleted

۳-۵-۲-۱- متوسط حالت الکتروتوپولوژیکی (Ms)

این توصیف‌کننده از گروه توصیف‌کننده‌های ساختاری می‌باشد. توصیف‌کننده‌های ساختاری ساده‌ترین و رایج‌ترین توصیف‌کننده‌ها هستند که منعکس‌کننده ساختار تشکیل‌دهنده یک ترکیب می‌باشند. تعداد اتم‌ها، تعداد پیوندها، تعداد حلقه‌ها و وزن مولکولی از توصیف کننده‌های زیرمجموعه این گروه هستند. این گروه از توصیف کننده‌ها به تفاوت‌های کنفورماسیونی حساس نیستند و بنابراین بین ایزومرهای فضایی یک ترکیب تفاوت قائل نمی‌شوند.

حالت الکتروتوپولوژیکی اتم i ام در یک مولکول که با S_i نشان داده می‌شود، اطلاعاتی راجع به حالت توپولوژیکی و الکترونی اتم در مولکول می‌دهد و به صورت زیر تعریف می‌شود (معادله ۳-۵):

$$S_i = I_i + \Delta I_i = I_i + \sum_{j=1}^{A} \frac{I_i - I_j}{(d_{ij}+1)^k}$$ (معادله ۳-۵)

در معادله فوق، I_i حالت الکتروتوپولوژیکی ذاتی اتم i ام و ΔI_i اثر میدان روی اتم i ام است که به صورت انحراف حالت الکتروتوپولوژیکی اتم i ام توسط سایر اتم‌های ملکول محاسبه میگردد. d_{ij} فاصله الکتروتوپولیژیکی بین i امین اتم و j امین اتم است و A نیز تعداد اتم‌ها در ملکول است. K نیز پارامتری است که اثر اتم‌های مجاور یا دور را برای مطالعات خاص اصلاح می‌کند و معمولاً برابر با ۲ در نظر گرفته می‌شود. حالت الکتروتوپولیژیکی i امین اتم به صورت زیر محاسبه می‌شود:

$$I_i = \frac{\left(2/L_i\right)^2 . \delta_i^V + 1}{\delta_i}$$ (معادله ۳-۶)

در معادله فوق، L_i عدد کوانتومی اصلی (۲برای اتم‌های C، N، O و F و ۳ برای Si ,S و Cl و...)، δ_i^V تعداد الکترون‌های لایه ظرفیت و δ_i تعداد الکترون‌های سیگمای i امین اتم در گراف ملکولی تهی شده از هیدروژن است. حالت الکتروتوپولیژیکی ذاتی یک اتم را می‌توان به صورت ساده نسبت جفت الکترون‌های تنها و σ به تعداد پیوندهای π در گراف ملکولی برای اتم مورد نظر در نظر گرفت. بنابراین حالت الکتروتوپولوژیکی ذاتی یک اتم منعکس‌کننده امکان پخش شدن اثر الکترون‌های غیر سیگما در اطراف اتم مورد نظر است. هرچه پخش

شدن اثر الکترون کمتر باشد، الکترون‌های لایه ظرفیت برای برهمکنش‌های بین ملکولی بیشتر در دسترس هستند.

از تعریف حالت الکتروتوپولوژیکی ذاتی و اثرات میدان می‌توان استنباط کرد که مقادیر مثبت و بزرگ S_i به اتم‌هایی با الکترونگاتیویته بالا و یا اتم‌های انتهایی مربوط می‌شوند. مقادیر کوچک یا منفی S_i به اتم‌هایی مربوط می‌شوند که فقط دارای الکترون‌های σ هستند و یا داخل ملکول محبوس شده و یا در نزدیکی الکترون‌هایی با الکترونگاتیویته بالاتر قرارگرفته‌اند. بنابراین، حالت الکتروتوپولوژیکی منعکس کننده میزان دردسترس بودن الکترون‌های اتم می‌باشد و می‌توان آن را به صورت معیاری از احتمال برهمکنش با سایر ملکول‌ها نیز تفسیر کرد. انتظار داریم اثر این توصیف‌کننده بر روی مدل مثبت باشد، یعنی با افزایش میانگین حالت الکتروتوپولوژیکی برای حلال مورد نظر، برهم کنش با سایر ترکیبات افزایش یافته و بنابراین پارامتر حلالیت نیز افزایش یابد. با توجه به مقدار مثبت اثر متوسط این توصیف کننده (۰/۳۹۹۲) در مدل، فرضیه فوق تایید می‌شود (جدول۳-۳)[۳۸].

۳-۵-۲-۲- توصیف‌کننده E1V

این توصیف‌کننده در گروه توصیف‌کننده‌های WHIM[۱] قرار گرفته است. توصیف‌کننده‌های WHIM طوری طراحی شده اند که اطلاعات ملکولی سه بعدی مناسبی را درباره‌ی اندازه ملکول، شکل و تقارن آن و همچنین نحوه توزیع اتم‌ها در ملکول، نسبت به یک چهارچوب ثابت به عنوان مرجع، فراهم کنند.

توضیح کامل درباره نحوه به دست آوردن این توصیف‌کننده‌ها از بحث ما خارج است، فقط باید به این نکته اشاره کنیم که برای به دست آوردن این توصیف‌کننده‌ها از چند طرح متفاوت برای وزن‌دهی ماتریس کوواریانس[۲] استفاده می‌شود (معادله۳-۷).

$$S_{jk} = \frac{\sum_{i=1}^{A} w_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^{A} w_i} \qquad \text{(معادله ۳-۷)}$$

۱.Wheighted Holistic Invariant Molecular descriptors

۲.Covariance matrix

در معادله فوق، s_{jk} کواریانس وزن‌دار شده بین j امین و k امین کوئوردینه اتمی است، A تعداد اتم‌ها، w_i وزن i امین اتم، q_{ij} و q_{ik} به ترتیب نماینده‌ی j امین و k امین کوئوردینه‌ی i امین اتم و $\bar{q}$ مقدار میانگین می‌باشد. شش طرح متفاوت برای وزن‌دهی پیشنهاد شده است؛ یکی از این طرح‌ها استفاده از حجم‌های وان‌در والس[1](v) مربوط به هر اتم است که به صورت ضریبی از حجم وان‌دروالس اتم کربن به عنوان مرجع، به دست می‌آید(جدول ۳-۲۰).

حجم وان در والس یک مولکول ارتباط نزدیکی با قابلیت در دسترس بودن آن برای ایجاد برهمکنش با ملکول‌های دیگر دارد.

جدول ۳-۲۰- مقادیر اصلی و مقیاس بندی شده حجم وان در والس برای چند اتم

	v	W/V(C)
H	۶/۷۰۹	۰/۲۹۹
C	۲۲/۴۴۹	۱/۰۰۰
N	۱۵/۵۹۹	۰/۶۹۵
O	۱۱/۴۹۴	۰/۵۱۲
Cl	۲۳/۲۲۸	۱/۰۳۵

افزایش مقدار عددی این توصیف‌کننده باعث کاهش مقدار پارامتر حلالیت می‌گردد[۳۹]. اثر متوسط این توصیف‌کننده روی پارامتر حلالیت منفی است (۰/۳۷۶۴-).

۳-۵-۲-۳- توصیف کننده Mor29m

این توصیف کننده جزء توصیف کننده‌های 3D-MoRSE است. ایده توصیف کننده‌های 3D-MoRSE، به دست آوردن اطلاعات از مختصات اتمی سه بعدی با استفاده از تبدیل است که در مطالعات پراش الکترونی برای فراهم آوردن منحنی‌های نظری تفرق کاربرد دارد. یک تابع تفرق تعمیم یافته به نام تبدیل ملکولی را

۱.Van der Waals volumes

می‌توان به عنوان تابع پایه برای به دست آوردن ارتباط آنالیزی ویژه اشعه x و پراش الکترونی از روی یک ساختار ملکولی مشخص مورد استفاده قرار داد.

این توصیف کننده با جرم اتمی (m) وزن‌دار شده است، و اثر آن روی پارامتر حلالیت کوچک ومنفی (-۰/۰۰۴۷) است[۴۰].

۳-۵-۲-۴- توصیف کننده X1Av

این توصیف کننده جزء گروه توصیف کننده‌های توپولوژیکی است. شاخص‌های اتصال [۱]جزء رایج‌ترین توصیف کننده‌های توپولوژیکی هستند و با توجه به درجه رأس اتم‌ها در گراف ملکولی تهی شده از هیدروژن محاسبه می‌شوند. شاخص اتصال راندیک اولین شاخص اتصال بود که توسط راندیک پیشنهاد شد و شاخص شاخه‌ای شدن نیز نامیده شده و با معادله ۳-۶ تعریف میگردد:

$$\chi_R = \sum_b \left(\delta_i . \delta_j\right)_b^{-1/2} \qquad \text{(معادله ۳-۸)}$$

b از پیوندهای B بین i-j در ملکول عبور می‌کند؛ δ_i و δ_j درجات رأس اتم‌های واقع در پیوند مورد نظر می‌باشند. این پارامتر ارتباط نزدیکی به عنوان معیاری از میزان شاخه‌دار بودن مولکول پیشنهاد شده است.

هر عبارت از جمع بالا $\left(\delta_i . \delta_j\right)_b^{-1/2}$ اتصال یالی نامیده شده و می‌تواند برای تعیین درجه پیوند یال‌ها مورد استفاده قرار بگیرد. درجه پیوند را می‌توان برای محاسبه در دسترس بودن یک پیوند از نظر امکان برهمکنش با سایر پیوندها به کار گرفت، چون معکوس درجه‌ی رأس جزئی از تعداد کل الکترون‌های سیگمای غیر هیدروژن است که پیوندهای مختلف از آن سهم دارند.

تفسیر شاخص بالا با تأکید روی امکان برهمکنش بین دو ملکول صورت می‌گیرد و منعکس کننده تأثیر میزان در دسترس بودن هر ملکول برای سایر ملکول‌های قرار گرفته در محیط اطرافش، روی خواص جمعی ملکول

۱.Connectivity indices

است. بنابراین شاخص اتصال راندیک(χ_R) را می‌توان به صورت سهم یک ملکول در برهمکنش‌های بین ملکـولی در نظر گرفت.

اگر در معادله ۳-۶ درجه رأس را با درجه رأس ظرفیت δ^v جایگزین کنیم، شاخص‌های اتصال ظرفیت به دسـت می‌آیند.

درجه رأس ظرفیت δ^v با فرمول زیر به دست می‌آید:

$$\delta_i^v = z_i^v - h_i = \delta_i + \pi_i + n_i - h_i$$ (معادله ۳-۹)

در معادله فوق z_i^v تعداد الکترون‌های ظرفیتی (الکترون‌های π ، الکترون‌های σ و جفـت الکتـرون‌هـای تنهـا) در i امین اتم است و h_i تعداد اتم‌های هیدروژن مربوط به i امین اتم است. این تعریف برای اتم‌های قرار گرفتـه در دومین تراز کوانتومی اصلی(C,N,O,F) صدق می‌کند. کایر[۱] و هال[۲] پیشنهاد کردند که برای اتم‌های قرار گرفته در ترازهای کوانتومی اصلی بالاتر (P,S,Cl,I)، هم الکترون‌های ظرفیتی و هم غیر ظرفیتی در نظر گرفته شوند:

$$\delta_i^v = (z_i^v - h_i)/(z_i - z_i^v - 1)$$ (معادله ۳-۱۰)

در معادله فوق، z_i تعداد کل الکترون‌های i امین اتم، یعنی عدداتمی آن است.

از آنجایی که X1Av توصیفی از میزان در دسترس بودن ملکول‌ها برای برهمکنش با سایر ملکول‌هـا مـی‌باشـد، انتظار داریم با افزایش مقدار آن، پارامتر حلالیت نیز افزایش بیابد. همانطور که در جدول ۳-۳ ملاحظه می‌شود، اثر متوسط این توصیف کننده مثبت (۰/۱۵۸۹) می‌باشد[۴۱].

۳-۵-۲-۵- توصیف کننده ATS1e

این توصیف کننده زیر مجموعه توصیف کننده های خودهمبستگی است. این دسته از توصیف کننده‌ها توصیف کننده‌های ملکولی هستند که بر اساس تابع خودهمبستگی (AC_1)محاسبه می‌شوند.

۱.Kier

۲.Hall

تابع خودهمبستگی به صورت زیر تعریف می‌شود:

$$AC_l = \int_a^b f(x).f(x+l).dx \qquad \text{(معادله ۳-۱۱)}$$

که $f(x)$ هر تابعی از متغیر x و l، lag است که نشان‌دهنده‌ی یک بازه از x می‌باشد. a و b بازه‌ی کلی مورد مطالعه‌ی تابع را تعریف می‌کنند. تابع $f(x)$ معمولاً یا یک تابع وابسته به زمان (مانند یک سیگنال الکتریکی وابسته به زمان) و یا یک تابع وابسته به فضا (مانند چگالی جمعیت در فضا) است.

این توصیف کننده با الکترونگاتیویته ساندرسون[۱] وزن‌دار شده است. هرچه الکترونگاتیویته اتم‌های تشکیل‌دهنده مولکول حلال بیشتر باشد، مقدار عددی این توصیف کننده نیز افزایش پیدا می‌کند. انتظار می‌رود افزایش الکترونگاتیویته باعث افزایش برهمکنش مولکول حلال با محیط اطراف و در نتیجه افزایش پارامتر حلالیت شود. اثر متوسط بزرگ و مثبت (۱/۹۲۴۵) این توصیف کننده نیز این مسئله را تأیید می‌کند[۴۳،۴۲] .

۳-۶- آینده‌نگری

تعداد حلال‌هایی که در صنایع شیمیایی مانند داروسازی و پلی‌مر و رنگ مورد استفاده قرار می‌گیرد بسیار گسترده است. می‌توان با توجه به نوع گروه‌های عاملی حلال، یک طبقه بندی انجام داد و مدل دقیقتری برای پیش‌بینی پارامتر حلالیت هر دسته ارائه داد.

به جای شبکه عصبی نیز می‌توان از روش‌های غیر خطی دیگری چون SVM و LS-SVM[۲] و به جای روش MLR نیز می‌توان از روش‌های خطی دیگری چون PLS بهره جست و نتایج را با هم مورد مقایسه قرار داد.

۱.Sanderson electronegativity

۲.Least Square supported Vector Machine

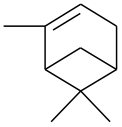
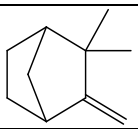
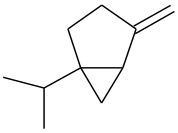
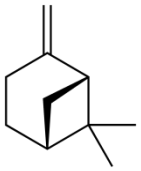
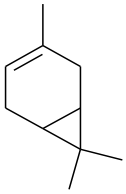
فصل چهارم

مدل‌سازی QSPR شاخص بازداری (RI) روغن‌های ضروری با استفاده از روش رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی(ANN)

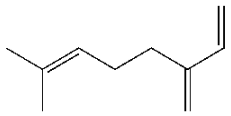
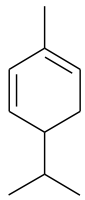
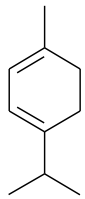
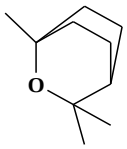
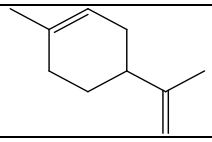
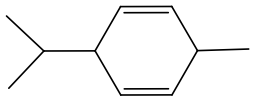
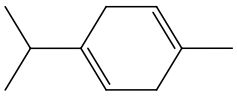
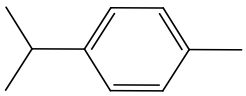
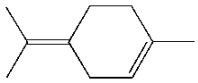
۴-۱- سری داده‌ها

در این قسمت از تحقیق شاخص بازداری (RI) ۷۰ ترکیب موجود در روغن‌های ضروری با گروه‌های عاملی گوناگون به عنوان سری داده‌ها مورد استفاده قرار گرفت (جدول ۴-۱)[۶]. نام و ساختار و شاخص بازداری این ترکیبات در جدول ۴-۱ آمده است.

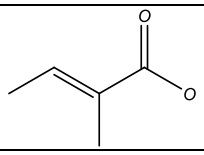
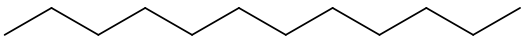
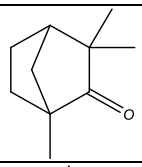
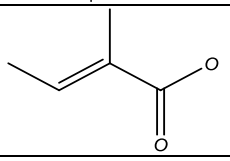
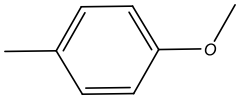
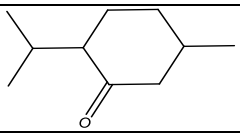
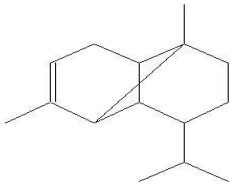
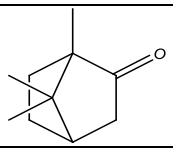
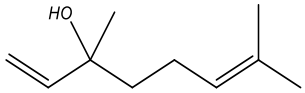
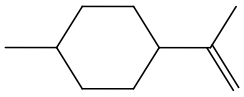
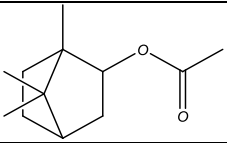
جدول۴-۱- سری داده‌های به کار رفته در مدل‌سازی

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۱	α-Thujene		۱۰۳۲
۲	α-Pinene		۱۰۳۸
۳	Camphene		۱۰۶۰
۴	Sabinene		۱۰۹۲
۵	β-Pinene		۱۱۰۲
۶	Δ^3-Carene		۱۱۳۱

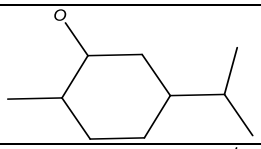
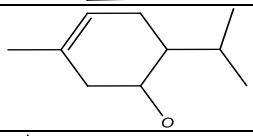
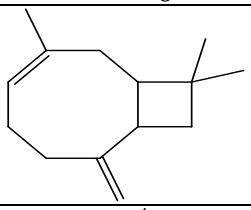
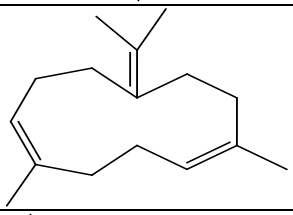
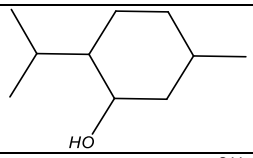
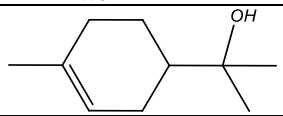
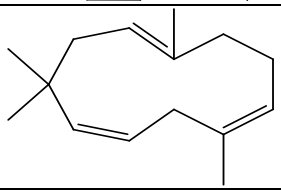
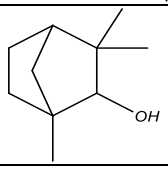
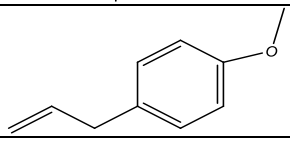
ادامه جدول۴-۱

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۷	Myrcene		۱۱۴۸
۸	α-Phellandrene		۱۱۶۱
۹	α-Terpinene		۱۱۶۳
۱۰	1-8-Cineole		۱۱۷۹
۱۱	Limonene		۱۱۸۳
۱۲	β-Phellandrene		۱۱۸۷
۱۳	g-Terpinene		۱۲۳۱
۱۴	p-Cymene		۱۲۴۷
۱۵	α-Terpinolene		۱۲۶۰

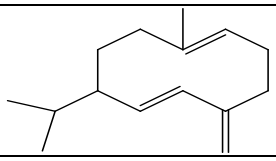
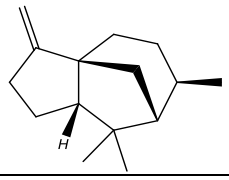
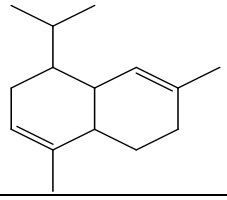
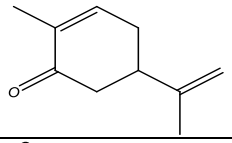
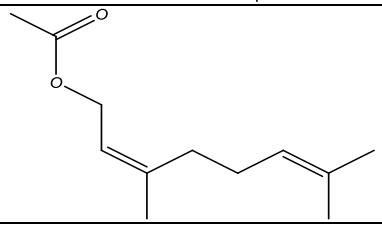
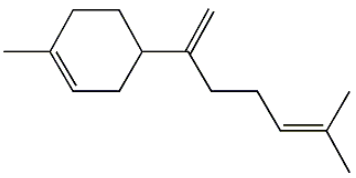
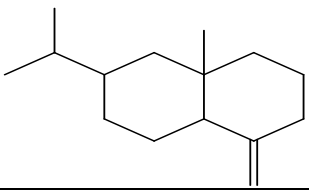
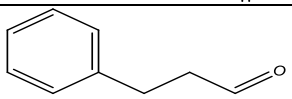
ادامه جدول۴-۱

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۱۶	(Z)-2-methyl-2-butene acid		۱۲۷۲
۱۷	Dodecane		۱۳۲۹
۱۸	Fenchone		۱۳۶۵
۱۹	(E)-2-methyl-2-butene acid		۱۳۷۳
۲۰	4-methyl anisole		۱۳۹۲
۲۱	Isomenthone		۱۴۵۲
۲۲	α-Copaene		۱۴۶۶
۲۳	Camphor		۱۴۸۲
۲۴	Linalool		۱۵۰۷
۲۵	Menth-8-ene		۱۵۳۰
۲۶	Bornyl acetate		۱۵۵۰

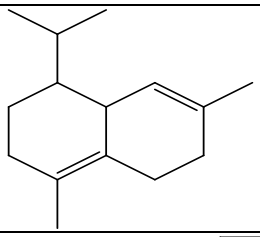
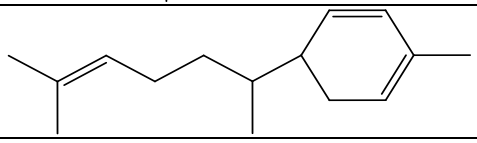
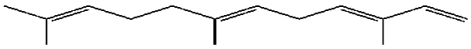
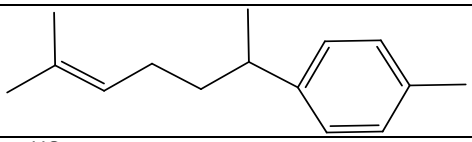
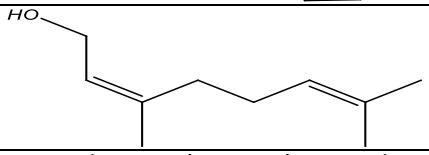
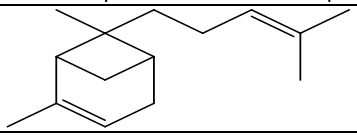
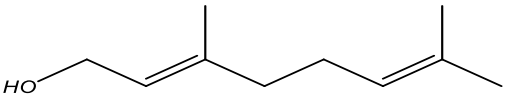
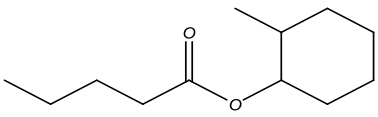
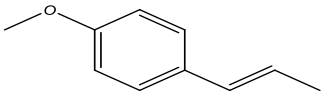
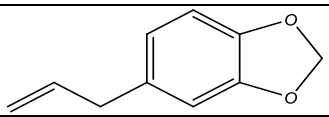
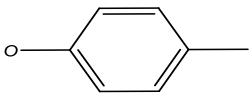
ادامه جدول۴-۱

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۲۷	Neoisomenthol		۱۵۵۶
۲۸	Terpinen-4-ol		۱۵۶۱
۲۹	Caryophyllene		۱۵۸۵
۳۰	Germacrene B		۱۶۰۶
۳۱	Neomenthol		۱۶۱۲
۳۲	α-Terpineol		۱۶۲۴
۳۳	α-Humulene		۱۶۳۸
۳۴	Fenchol		۱۶۴۶
۳۵	Estragole		۱۶۵۵

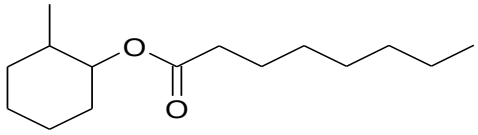
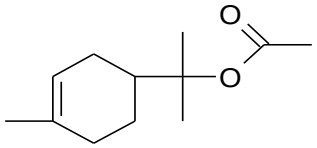
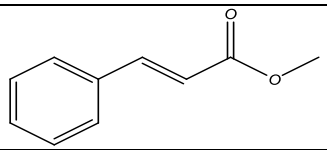
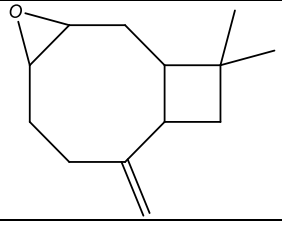
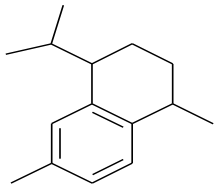
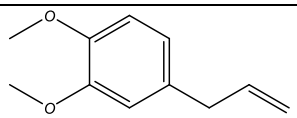
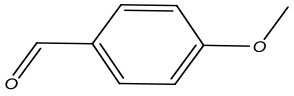
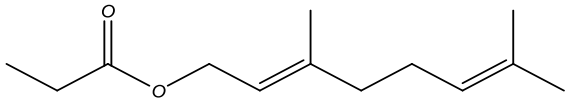
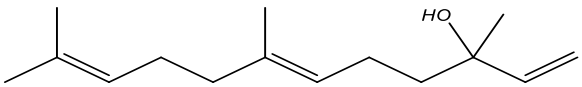
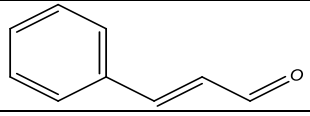
ادامه جدول۴-۱

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۳۶	Germacrene D		۱۶۶۹
۳۷	α-Cedrene		۱۶۷۴
۳۸	α-Muurolene		۱۶۸۳
۳۹	Carvone		۱۶۸۴
۴۰	Neryl acetate		۱۶۹۲
۴۱	β-Bisabolene		۱۶۹۴
۴۲	β-Selinene		۱۶۹۵
۴۳	Benzenepropanal		۱۷۰۹

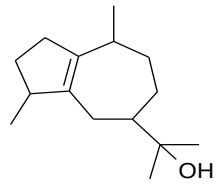
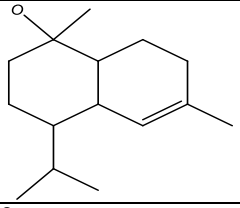
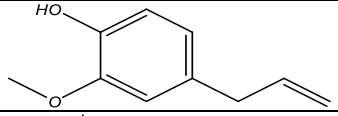
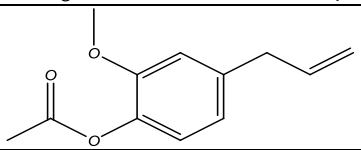
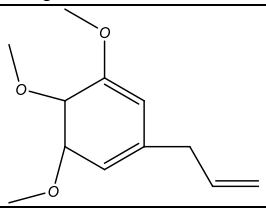
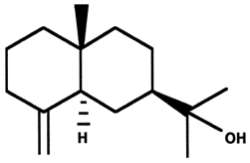
ادامه جدول۴-۱

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۴۴	Δ-Cadinene		۱۷۱۶
۴۵	α-Zingiberene		۱۷۲۴
۴۶	α-Farnesene		۱۷۲۵
۴۷	ar-Curcumene		۱۷۴۷
۴۸	Nerol		۱۷۶۲
۴۹	α-Bergamotene		۱۷۷۹
۵۰	Geraniol		۱۷۸۷
۵۱	2-Methylcyclohexyl pentanoate		۱۷۹۸
۵۲	trans-Anethole		۱۸۰۹
۵۳	Safrole		۱۸۰۹
۵۴	Cresole		۱۸۱۲

ادامه جدول۴-۱

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۱۸۵۶	2-Methylcyclohexyl octanoate		۱۸۵۶
۱۸۸۰	α-Terpinyl acetate		۱۸۸۰
۵۷	Methyl cinnamate		۱۹۰۰
۵۸	Caryophyllene oxide		۱۹۱۷
۵۹	cis-Calamenene		۱۹۲۴
۶۰	Methyl eugenol		۱۹۴۷
۶۱	Anisaldehyde		۱۹۴۷
۶۲	Geranyl propanoate		۱۹۵۶
۶۳	Nerolidol		۱۹۹۶
۶۴	trans-Cinnam aldehyde		۱۹۹۷

ادامه جدول۴-۱

شماره ترکیب	نام ترکیب	ساختار ترکیب	شاخص بازداری
۶۵	Guaiol		۲۰۸۳
۶۶	α-Cadinol		۲۰۸۵
۶۷	Eugenol		۲۰۹۸
۶۸	Eugenyl acetate		۲۱۰۷
۶۹	Elemicine		۲۱۶۵
۷۰	β-Eudesmol		۲۱۷۶

۴-۲- رسم و بهینه‌سازی ساختار هندسی و به دست آوردن توصیف‌کننده‌ها

با توجه به موضوع مورد بررسی (بررسی ارتباط کمّی ساختار- ویژگی) ابتدا ساختار مولکول‌ها در نرم افزار HyperChem رسم و ساختار هندسی آنها با احتساب اتم‌های هیدروژن و با استفاده از روش AM1 بهینه شد. بهینه‌سازی تا زمانی ادامه یافت که گرادیان انرژی به ۰/۰۰۱ کیلو کالری بر مول برسد. سپس ساختارهای بهینه شده توسط این نرم افزار به عنوان ورودی به نرم افزار Dragon وارد شدند. محاسبه توصیف‌کننده‌ها توسط این نرم‌افزار صورت گرفت و ۱۴۸۱ توصیف‌کننده از هیجده گروه مختلف برای هر ترکیب محاسبه گردید.

۴-۳- انتخاب بهترین توصیف‌کننده‌ها و به دست آوردن مدل جهت پیش‌بینی RI به روش رگرسیون خطی چندگانه

توصیف‌کننده‌های به دست آمده از مرحله قبل به عنوان متغیر مستقل و شاخص بازداری ترکیبات به عنوان متغیر وابسته وارد نرم افزار SPSS شدند. سپس همانطور که در فصل قبل (بخش ۳-۳) توضیح داده شد و با به‌کارگیری تکنیک‌های کاهش متغیر و روش مرحله‌ای برای انتخاب متغیر، بهترین توصیف‌کننده‌ها انتخاب شدند. سپس بر اساس این توصیف‌کننده‌ها و به‌کارگیری روش رگرسیون خطی چندگانه (MLR)، مدل‌سازی صورت گرفت. شش مدل نهایی همراه با توصیف‌کننده‌های وارد شده در مدل و همچنین پارامترهای آماری مربوطه در جدول ۴-۲ آورده شده است.

جدول۴-۲- مقایسه آماره‌های چند مدل نهایی به دست آمده از روش MLR

شماره مدل	توصیف کننده‌ها	R	SE	F
۲	Ss, MATS2e	۰/۷۷۸	۱۵۱/۲۹	۱۵۱/۰۱۸
۳	Ss, MATS2e, BIC5	۰/۹۱۲	۱۳۰/۳۵۸	۱۳۰/۰۱۸
۳	Ss, MATS2e, BIC5, nOH	۰/۹۳۴	۱۱۴/۴۷۸	۱۱۱/۳۸۵
۴	Ss, MATS2e, BIC5, nOH, H2m	۰/۹۵۰	۱۰۰/۸۸۰۱	۱۱۸/۶۸۹
۵	Ss, MATS2e, BIC5, nOH, H2m, R8e	۰/۹۵۶	۹۵/۴۷۴	۱۱۱/۸۳۶
۶	Ss, MATS2e, BIC5, nOH, H2m, R8e, RDF050m	۰/۹۶۲	۸۹/۲۰۰	۱۱۱/۲۷۲
۷	Ss, MATS2e, BIC5, nOH, H2m, R8e, RDF050m, IC2	۰/۹۶۹	۸۲/۰۰۰	۱۱۶/۷۵۷
۸	Ss, MATS2e, BIC5, nOH, H2m, R8e, RDF050m, IC2, Mor11m	**۰/۹۷۳**	**۷۷/۳۴۱**	**۱۱۷/۶۱۴**

همانطور که در جدول ۴-۲ مشاهده می‌شود، مدل ششم با در نظر گرفتن هر سه پارامترهای آماری R، SE و F بر سایر مدل‌ها برتری داشت و به عنوان مدل نهایی انتخاب شد. توصیف‌کننده‌های وارد شده در این مدل به همراه ضریب رگرسیون و اثر متوسط آنها در جدول ۴-۳ آورده شده است.

جدول۴-۳- توصیف کننده‌های وارد شده در مدل MLR

شماره	علامت	ضریب رگرسیون	اثر متوسط	گروه	نام کامل
۱	Ss	۲۲/۶۸۸	۰/۳۵۱۷۴	constitutional descriptors	sum of Kier-Hall electrotopological states
۲	MATS2e	-۷۹۹/۲۶	-۰/۰۳۴۲۴	2D autocorrelation descriptors	Moran autocorrelation of lag 1/weighted by atomic sanderson electronegativity
۳	BIC5	۲۲۱۵/۱	۱/۱۲۸۷	topological descriptors	Bond information content (neighborhood symmetry of 5-order)
۴	nOH	۱۸۱/۶	۰/۰۲۱۷۶۸	constitutional descriptors	Numbor of OH
۵	H2m	۹۹۲/۱۹	۰/۲۲۹۰۵	GETAWAY descriptors	H autocorrelation of lag 2/weighted by atomic masses
۶	R8e	-۸۴۷/۹۴	-۰/۰۵۶۷۱۸	GETAWAY descriptors	R autocorrelation of lag 5/weighted by atomic sanderson electronegativity
۷	RDF050m	۲۸/۷۲۴	۰/۰۵۰۳۸۹	RDF descriptors	Radial Distribution Function -0-5 weighted by atomic masses
۸	IC2	-۱۴۰/۷۳	-۰/۲۸۱۲۸	Molecular descriptors	Information content index (neighborhood symmetry of 2-order)
۹	Mor11m	-۸۳/۷۵	-۰/۰۱۶۵۶۷	3D-MorSe descriptors	3D-MorSe-signal 05/wheited by atomic masses

پارامترهای وارد شونده به معادله خطی باید حتی الامکان نسبت به یکدیگر مستقل باشند. جدول ۴-۴ ماتریس همبستگی توصیف کننده‌های وارد شده در مدل منتخب را نشان می‌دهد. همانطور که ملاحظه می‌شود مقدار همبستگی بین پارامترها قابل قبول (کمتر از ۰/۹) است.

جدول۴-۴- ماتریس همبستگی توصیف‌کننده‌های به‌کار رفته در مدل نهایی

توصیف کننده	Ss	MATS2e	BIC5	nOH	H2m	R8e	RDF050	IC2	Mor11m
Ss	۱								
MATS2e	۰/۱۷۹	۱							
BIC5	۰/۵۶۶	۰/۱۹۰	۱						
nOH	۰/۱۰۴	۰/۲۳۵	۰/۱۰۳	۱					
H2m	۰/۸۷	۰/۲۰۱	۰/۵۷۹	۰/۰۰۷	۱				
R8e	۰/۵۱۶	-۰/۱۲	۰/۳۰۳	۰/۲۴۶	۰/۵۱۳	۱			
RDF050m	۰/۶۷۰	-۰/۰۴۹	۰/۲۰۴	۰/۱۷۱	۰/۶۵۰	۰/۵۲۲	۱		
IC2	۰/۴۱۷	۰/۰۹۰	۰/۶۱۷	۰/۲۴۷	۰/۳۹۳	۰/۲۶۴	۰/۳۳۱	۱	
Mor11m	۰/۰۱۱	-۰/۲۵۹	-۰/۰۹۸	۰/۱۶۴	۰/۲۳۷	۰/۳۵۹	۰/۳۳۴	-۰/۰۵۹	۱

به این ترتیب معادله خطی نهایی، با استفاده از سری آموزش به صورت زیر دست آمد: (معادله ۴-۱)

$$y = -624.29 + 22.688\,Ss - 799.26\,MATS2e + 2215.1BIC5 + 181.6\,nOH + 992.19\,H2m - 847.94\,R8e + 28.724\,RDF050m - 140.73IC2 - 83.75Mor11m$$

مقادیر عددی توصیف‌کننده‌های وارد شده در مدل MLR نهایی در ادامه آورده شده است (جدول۴-۵).

جدول۴-۵- مقادیر عددی توصیف‌کننده‌های وارد شده در مدل MLR نهایی

شماره ترکیب	Mor11m	IC2	RDF050	R8e	H2m	nOH	BIC5	MATS2e	Ss
۱	۰/۳۰۸	۳/۱۴۰	۱/۵۵۱	۰/۰۰۰	۰/۱۷۰	۰	۰/۷۶۰	۰/۱۱۵	۱۶/۵۰۰
۲	۰/۲۲۱	۳/۱۴۰	۱/۳۷۹	۰/۰۰۰	۰/۱۹۴	۰	۰/۷۶۸	۰/۱۴۴	۱۶/۵۸۰
۳	-۰/۲۳۰	۳/۲۰۰	۰/۱۹۹	۰/۰۰۰	۰/۱۵۶	۰	۰/۷۷۴	۰/۱۶۸	۱۷/۰۸۰
۴	۰/۲۷۱	۳/۲۷۷	۱/۲۲۶	۰/۰۰۰	۰/۱۹۱	۰	۰/۷۷۴	۰/۰۸۳	۱۷/۰۸۰
۵	۰/۱۸۶	۳/۲۷۷	۰/۱۸۸	۰/۰۰۰	۰/۲۰۱	۰	۰/۷۷۴	۰/۱۱۳	۱۷/۰۸۰
۶	۰/۶۸۴	۳/۰۶۵	۰/۳۰۶	۰/۰۰۰	۰/۲۳۹	۰	۰/۷۶۸	۰/۱۱۵	۱۶/۵۸۰
۷	۰/۳۰۳	۳/۳۹۰	۳/۹۵۴	۰/۲۲۰	۰/۲۳۴	۰	۰/۷۶۷	۰/۰۱۷	۲۴/۳۳۰
۸	۰/۲۸۱	۳/۱۱۳	۱/۲۷۰	۰/۱۵۰	۰/۲۰۷	۰	۰/۷۸۴	۰/۰۴۳	۱۷/۸۳۰
۹	۰/۱۹۲	۲/۹۸۸	۲/۳۱۴	۰/۱۳۰	۰/۱۸۹	۰	۰/۷۶۸	۰/۰۷۷	۱۷/۶۷۰

ادامه جدول۴-۵

شماره ترکیب	Mor11m	IC2	RDF050	R8e	H2m	nOH	BIC5	MATS2e	Ss
۱۰	-۰/۰۶۶	۲/۶۷۶	۰/۴۳۱	۰/۰۰۰	۰/۲۸۵	۰	۰/۶۹۳	۰/۱۵۸	۱۹/۳۳۰
۱۱	۰/۳۵۲	۳/۲۰۰	۱/۲۷۸	۰/۰۹۷	۰/۱۹۳	۰	۰/۸۳۸	۰/۰۴۳	۱۸/۱۷۰
۱۲	۰/۰۰۴	۰/۶۲۲	۰/۹۲۶	۰/۱۲۷	۰/۲۱۸	۰	۰/۷۳۶	۰/۰۰۹	۱۸/۰۰
۱۳	۰/۴۲۴	۲/۹۸۸	۰/۷۴۹	۰/۱۲۰	۰/۲۰۰	۰	۰/۷۶۸	۰/۰۷۷	۱۷/۶۷۰
۱۴	۰/۱۳۴	۲/۶۴۶	۲/۴۷۲	۰/۱۳۶	۰/۱۹۴	۰	۰/۷۶۹	۰/۱۱۰	۱۸/۶۷۰
۱۵	۰/۲۸۵	۲/۶۷۰	۳/۰۳۹	۰/۱۴۶	۰/۱۸۰	۰	۰/۷۵۲	۰/۱۱۰	۱۷/۵۰۰
۱۶	۰/۱۳۳	۰/۸۷۳	۰/۰۱۴	۰/۰۰۰	۰/۲۴۵	۰	۰/۸۱۸	۰/۳۷۰	۲۲/۳۳۰
۱۷	۰۸/۸۸۰	۱/۸۲۸	۰/۱۶۵	۰/۰۳۲	۰/۲۵۱	۰	۰/۶۲۶	-۰/۲۳۵	۱۹/۰۰۰
۱۸	-۰/۳۹۱	۲/۸۷۴	۱/۳۲۶	۰/۰۰۰	۰/۲۴۱	۰	۰/۷۶۳	۰/۰۶۵	۲۳/۰۰۰
۱۹	-۰/۰۲۶	۰/۸۷۳	۰/۴۳۹	۰/۰۰۰	۰/۳۸۶	۰	۰/۸۱۸	۰/۳۷۰	۲۲/۳۳۰
۲۰	۰/۰۹۵	۳/۱۱۶	۱/۲۸۴	۰/۱۲۵	۰/۲۲۱	۰	۰/۸۴۰	-۰/۱۳۳	۱۸/۸۳۰
۲۱	۰/۱۲۹	۲/۸۷۴	۰/۹۳۵	۰/۱۱۴	۰/۲۶۳	۰	۰/۷۹۱	۰/۰۴۶	۲۳/۱۷
۲۲	۰/۵۵۷	۳/۴۰۹	۲/۰۸۸	۰/۰۸۳	۰/۴۴۹	۰	۰/۸۱۶	۰/۱۷۲	۲۳/۸۳۰
۲۳	-۰/۱۷۹	۲/۹۴۸	۱/۰۱۷	۰/۰۰۰	۰/۲۹۶	۰	۰/۷۶۳	۰/۰۶۵	۲۳/۰۰
۲۴	۰/۴۳۴	۳/۳۹۱	۱/۸۵۶	۰/۱۳۰	۰/۲۸۰	۱	۰/۷۹۱	۰/۰۹۸	۲۴/۹۰۰
۲۵	۰/۲۵۳	۳/۱۱۱	۲/۳۸۸	۰/۰۰۰	۰/۲۱۰	۰	۰/۸۴۰	-۰/۰۰۹	۱۷/۳۰۰
۲۶	-۰/۲۴۸	۳/۲۲۶	۱/۸۳۵	۰/۱۴۰	۰/۵۰۰	۰	۰/۷۹۰	۰/۲۴۱	۲۹/۸۰۰
۲۷	۰/۴۸۵	۳/۰۹۸	۲/۰۰۷	۰/۱۰۰	۰/۲۶۷	۱	۰/۸۱۶	۰/۰۰۴	۲۱/۸۰۰
۲۸	۰/۴۳۲	۳/۴۶۰	۱/۲۳۷	۰/۱۴۷	۰/۳۱۶	۱	۰/۸۰۵	۰/۰۰۹	۲۲/۶۰۰
۲۹	۰/۷۱۱	۳/۴۹۷	۵/۰۹۷	۰/۱۵۶	۰/۴۷۱	۰	۰/۸۱۷	۰/۰۷۳	۲۵/۷۰۰
۳۱	۰/۳۵۰	۳/۰۹۸	۱/۲۵۶	۰/۱۰۶	۰/۲۹۸	۱	۰/۸۱۶	۰/۰۰۴	۲۱/۸۰۰
۳۲	۰/۲۷۲	۳/۲۰۱	۲/۱۱۸	۰/۱۵۸	۰/۲۸۱	۱	۰/۷۹۱	۰/۰۸۵	۲۲/۷۰۰
۳۳	۰/۴۷۱	۳/۱۵۷	۰/۷۸۴	۰/۲۲۶	۰/۵۳۲	۰	۰/۸۱۴	۰/۰۸۳	۲۶/۵۰۰
۳۴	-۰/۲۵۹	۳/۱۰۶	۰/۰۲۸	۰/۰۰۰	۰/۲۰۶	۱	۰/۷۹۱	۰/۰۲۴	۲۱/۶۰۰
۳۵	۰/۲۹۲	۰/۲۵۳	۲/۸۱۵	۰/۰۷۲	۰/۲۷۹	۰	۰/۸۷۱	-۰/۱۳۷	۲۳/۳۰۰
۳۶	۰/۹۶۰	۳/۵۶۰	۳/۱۹۸	۰/۱۷۸	۰/۴۸۹	۰	۰/۸۲۷	۰/۰۰۲	۲۷/۰۰
۳۷	-۰/۰۰۹	۳/۲۱۷	۳/۹۱۶	۰/۰۶۲	۰/۴۲۹	۰	۰/۸۰۹	۰/۲۷۴	۲۶/۰۰۰
۳۸	۰/۶۱۷	۳/۳۲۳	۴/۳۶۲	۰/۱۵۷	۰/۴۸۱	۰	۰/۸۲۳	۰/۰۷۹	۲۵/۱۰۰
۳۹	۰/۳۲۹	۳/۵۴۳	۱/۴۷۸	۰/۱۳۵	۰/۳۷۷	۰	۰/۸۳۷	۰/۰۵۹	۲۵/۳۰۰

ادامه جدول۴-۵

شماره ترکیب	Mor11m	IC2	RDF050	R8e	H2m	nOH	BIC5	MATS2e	Ss
۴۰	۰/۲۴۲	۳/۳۶۴	۴/۹۸۴	۰/۰۸۴	۰/۴۹۰	۰	۰/۸۰۱	۰/۲۴۳	۳۰/۳۰۰
۴۱	۰/۷۶۹	۳/۰۸۳	۶/۰۳۵	۰/۱۲۵	۰/۴۷۶	۰	۰/۷۸۴	۰/۰۹۴	۲۷/۴۰۰
۴۲	۰/۴۰۸	۳/۳۵۰	۳/۴۷۸	۰/۰۸۱	۰/۳۹۸	۰	۰/۸۲۴	۰/۰۸۱	۲۴/۹۰۰
۴۳	۰/۳۲۹	۳/۴۱۲	۱/۵۹۸	۰/۰۸۶	۰/۳۹۰	۰	۰/۸۶۸	-۰/۰۳۰	۲۷/۱۰۰
۴۴	۰/۳۹۰	۳/۴۶۱	۵/۵۷۸	۰/۱۵۸	۰/۴۳۰	۰	۰/۸۱۶	۰/۰۶۸	۲۴/۱۰۰
۴۵	۰/۷۲۰	۳/۲۸۴	۳/۷۴۸	۰/۱۵۰	۰/۳۷۷	۰	۰/۸۲۳	۰/۰۲۵	۲۶/۵۰۰
۴۶	۰/۷۷۸	۳/۰۴۱	۵/۴۱۹	۰/۱۸۲	۰/۴۳۲	۰	۰/۸۱۶	-۰/۰۱۸	۲۶/۱۰۰
۴۷	۰/۴۳۰	۳/۱۷۲	۵/۳۰۶	۰/۱۵۳	۰/۴۱۰	۰	۰/۸۱۶	۰/۰۶۷	۲۷/۳۰۰
۴۸	۰/۴۷۲	۳/۲۹۶	۴/۲۱۶	۰/۱۱۹	۰/۳۰۰	۱	۰/۷۹۱	-۰/۰۵۹	۲۳/۸۰۰
۴۹	۰/۴۹۲	۳/۲۹۱	۵/۵۱۰	۰/۱۰۴	۰/۴۸۵	۰	۰/۸۱۴	۰/۰۹۵	۲۵/۲۰۰
۵۰	۰/۵۹۶	۳/۲۹۶	۲/۹۳۱	۰/۱۱۰	۰/۳۲۸	۱	۰/۷۹۱	-۰/۰۵۹	۲۳/۸۰۰
۵۱	۰/۳۰۴	۲/۷۸۸	۲/۳۷۷	۰/۱۴۳	۰/۴۹۲	۰	۰/۸۶۷	۰/۱۶۶	۲۹/۳۰۰
۵۲	۰/۰۲۰	۳/۱۴۲	۲/۲۷۸	۰/۰۶۰	۰/۲۸۱	۰	۰/۸۶۴	-۰/۱۲۷	۲۲/۸۰۰
۵۳	۰/۲۱۷	۳/۶۳۷	۲/۱۳۴	۰/۰۵۵	۰/۵۲۱	۰	۰/۸۸۴	۰/۱۰۳	۲۵/۱۷۰
۵۴	۰/۰۵۵	۲/۹۵۳	۰/۰۰۴	۰/۰۰۰	۰/۲۲۰	۱	۰/۸۷۲	۰/۰۳۹	۱۹/۳۰۰
۵۵	۰/۰۱۹	۲/۸۸۲	۴/۹۸۳	۰/۰۸۶	۰/۳۸۰	۰	۰/۸۷۶	۰/۲۲۸	۳۳/۸۳۰
۵۶	۰/۳۰۴	۲/۷۸۸	۲/۳۷۷	۰/۱۴۳	۰/۴۹۲	۰	۰/۸۵۷	۰/۱۶۶	۲۹/۳۳۰
۵۷	۰/۳۲۶	۲/۹۱۸	۲/۷۴۶	۰/۱۶۶	۰/۵۶۸	۰	۰/۸۹۲	۰/۱۵۳	۲۹/۸۳۰
۵۸	۰/۵۲۵	۳/۵۶۸	۳/۳۷۱	۰/۰۹۴	۰/۵۴۵	۰	۰/۸۲۹	-۰/۰۳۴	۲۶/۲۵۰
۵۹	۰/۳۲۸	۳/۲۸۳	۶/۱۹۱	۰/۱۹۹	۰/۴۷۶	۰	۰/۸۱۶	۰/۱۱۵	۲۶/۰۰۰
۶۰	۰/۱۶۳	۳/۲۷۳	۳/۴۴۵	۰/۱۰۵	۰/۳۵۷	۰	۰/۸۱۴	-۰/۱۸۶	۲۸/۵۰۰
۶۱	-۰/۰۶۸	۳/۳۶۸	۲/۱۸۶	۰/۰۹۸	۰/۳۵۱	۰	۰/۸۴۸	-۰/۰۹۹	۲۷/۵۰۰
۶۲	۰/۴۱۵	۳/۲۷۳	۴/۸۲۲	۰/۱۲۸	۰/۵۶۶	۰	۰/۸۱۱	۰/۲۴۶	۳۶/۶۷۰
۶۳	۰/۶۱۷	۳/۴۲۷	۴/۹۹۸	۰/۱۷۶	۰/۵۱۷	۱	۰/۸۲۶	۰/۰۹۲	۳۳/۵۸۰
۶۴	۰/۰۷۴	۲/۳۷۰	۱/۵۶۸	۰/۰۴۴	۰/۳۱۹	۰	۰/۹۲۲	-۰/۰۹۴	۲۴/۶۷۰
۶۵	۰/۳۰۰	۳/۲۴۶	۶/۰۷۳	۰/۱۲۳	۰/۵۱۷	۱	۰/۸۲۳	۰/۰۷۶	۲۸/۲۸۰
۶۶	۰/۵۳۵	۳/۵۱۱	۸/۰۰۴	۰/۱۸۷	۰/۵۳۳	۱	۰/۸۳۵	۰/۰۷۸	۳۰/۲۵۰
۶۷	۰/۱۷۷	۳/۶۸۵	۳/۷۰۱	۰/۱۱۱	۰/۴۳۸	۱	۰/۸۷۸	۰/۰۹۷	۲۹/۰۰۰
۶۸	-۰/۱۷۰	۳/۸۸۰	۶/۱۶۱	۰/۱۳۳	۰/۵۴۰	۰	۰/۸۶۳	۰/۱۱۶	۳۷/۱۷۰
۶۹	۰/۰۱۹	۳/۴۲۱	۶/۷۳۷	۰/۱۴۹	۰/۳۸۴	۰	۰/۸۶۹	۰/۲۵۳	۳۳/۰۰۰
۷۰	۰/۲۷۰	۳/۴۳۳	۶/۷۸۶	۰/۱۵۴	۰/۵۲۹	۱	۰/۸۳۰	۰/۰۸۲	۳۰/۸۳۰

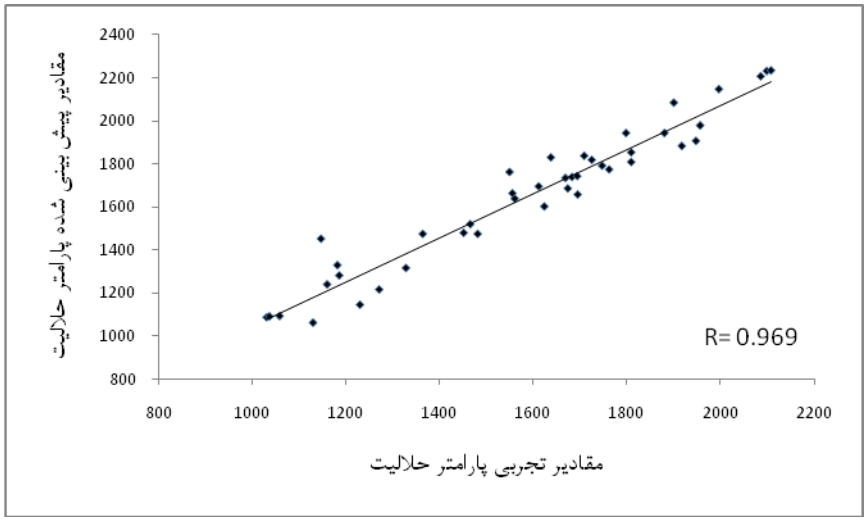
با استفاده از این مدل مقادیر شاخص بازداری برای سری ارزیابی و تست نیز پیش‌بینی شد. در جداول ۴-۶، ۴-۷ و ۴-۸ مقادیر تجربی، پیش‌بینی شده، خطای مطلق و درصد خطای نسبی به ترتیب برای سری آموزش، سری ارزیابی و سری تست آورده شده است. نمودارهای مقادیر پیش‌بینی شده در مقابل مقادیر تجربی پارامتر حلالیت برای سری آموزش، ارزیابی و تست نیز به ترتیب در شکل‌های ۴-۱، ۴-۲ و ۴-۳ رسم شد. ضریب همبستگی برای نمودارهای یاد شده به ترتیب ۰/۹۶۸، ۰/۹۶۷ و ۰/۹۵۶ می‌باشد. این مقادیر نشان‌دهنده کارامدی روش MLR در پیش‌بینی شاخص بازداری ترکیبات مورد نظر می‌باشد.

جدول۴-۶- مقایسه مقادیر تجربی و پیش‌بینی شده شاخص بازداری با روش MLR برای سری آموزش

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی	خطای مطلق	درصد خطای نسبی
۱	۱۰۳۲	۱۰۸۶/۵	۵۴/۵	۵/۳
۲	۱۰۳۸	۱۰۹۰/۸	۵۲/۸	۵/۱
۳	۱۰۶۰	۱۰۹۲/۰	۳۲/۰	۳/۰
۶	۱۱۳۱	۱۰۶۱/۱	-۶۹/۹	-۶/۲
۷	۱۱۴۸	۱۴۵۱/۲	۳۰۲/۲	۲۶/۳
۸	۱۱۶۱	۱۲۳۹/۱	۷۸/۱	۶/۷
۱۱	۱۱۸۳	۱۳۲۸/۷	۱۴۵/۷	۱۲/۳
۱۲	۱۱۸۷	۱۲۸۰/۴	۹۳/۴	۷/۳
۱۳	۱۲۳۱	۱۱۴۴/۶	-۸۶/۴	-۷/۰
۱۶	۱۲۷۲	۱۲۱۵/۴	-۵۶/۶	-۴/۴
۱۷	۱۳۲۹	۱۳۱۵/۲	-۱۳/۸	-۱/۴
۱۸	۱۳۶۵	۱۴۳۷/۹	۱۰۸/۹	۸/۰
۲۱	۱۴۵۲	۱۴۷۸/۴	۲۶/۴	۱/۸
۲۲	۱۴۶۶	۱۵۱۸/۸	۵۲/۸	۳/۶
۲۳	۱۴۸۲	۱۴۷۳/۶	-۸/۴	-۰/۶
۲۶	۱۵۵۰	۱۷۶۱/۳	۲۱۱/۳	۱۳/۶
۲۷	۱۵۵۶	۱۶۶۲/۲	۱۰۶/۲	۶/۸
۲۸	۱۵۶۱	۱۶۳۷/۳	۷۶/۳	۴/۹
۳۱	۱۶۱۲	۱۶۹۴	۸۲/۰	۵/۱
۳۲	۱۶۲۴	۱۶۰۱/۲	-۲۲/۸	-۱/۴
۳۳	۱۶۳۸	۱۸۲۸/۸	۱۹۰/۸	۱۱/۶
۳۶	۱۶۶۹	۱۷۳۳/۷	۶۴/۷	۳/۹

ادامه جدول۴-۶

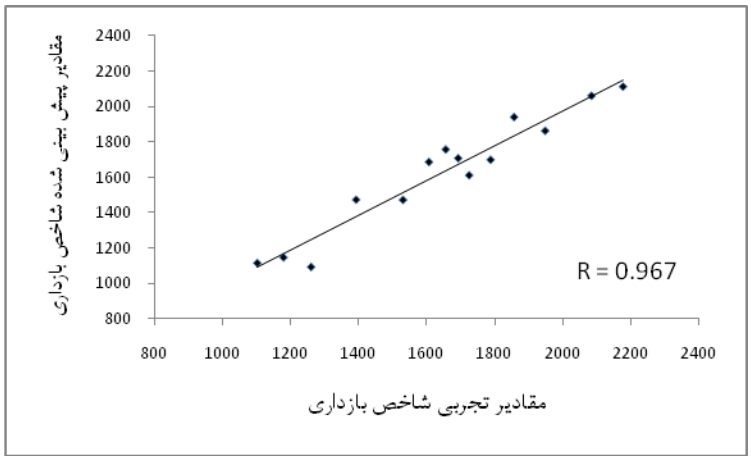
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۳۷	۱۶۷۴	۱۶۸۴/۷	۱۰/۷	۰/۶
۳۸	۱۶۸۳	۱۷۳۸/۱	۵۵/۱	۳/۳
۴۱	۱۶۹۴	۱۷۴۲/۲	۴۸/۲	۲/۸
۴۲	۱۶۹۵	۱۶۵۶/۵	-۳۸/۵	-۲/۳
۴۳	۱۷۰۹	۱۸۳۶/۳	۱۲۷/۳	۷/۴
۴۶	۱۷۲۵	۱۸۱۷/۳	۹۲/۳	۵/۴
۴۷	۱۷۴۷	۱۷۹۰/۴	۴۳/۴	۲/۵
۴۸	۱۷۶۲	۱۷۷۳/۰	۱۱/۰	۰/۶
۵۱	۱۷۹۸	۱۹۴۲/۰	۱۴۴/۰	۸/۰
۵۲	۱۸۰۹	۱۸۰۷/۷	-۱/۳	-۰/۱
۵۳	۱۸۰۹	۱۸۵۲/۵	۴۳/۵	۲/۴
۵۶	۱۸۸۰	۱۹۴۲/۰	۶۲/۰	۳/۳
۵۷	۱۹۰۰	۲۰۸۳/۲	۱۸۳/۲	۹/۶
۵۸	۱۹۱۷	۱۸۸۲/۲	-۳۴/۸	-۱/۸
۶۱	۱۹۴۷	۱۹۰۵/۵	-۴۱/۵	-۲/۱
۶۲	۱۹۵۶	۱۹۷۷/۴	۲۱/۴	۱/۱
۶۳	۱۹۹۶	۲۱۴۶/۱	۱۵۰/۱	۷/۵
۶۶	۲۰۸۵	۲۲۰۵/۸	۱۲۰/۸	۵/۸
۶۷	۲۰۹۸	۲۲۳۰/۲	۱۳۲/۲	۶/۳
۶۸	۲۱۰۷	۲۲۳۳/۰	۱۲۶/۰	۶/۰



شکل۴-۱- رسم مقادیر پیش‌بینی شده شاخص بازداری توسط روش MLR بر حسب مقادیر تجربی برای سری آموزش

جدول۴-۷- مقایسه مقادیر تجربی و پیش‌بینی شده شاخص بازداری با روش MLR برای سری ارزیابی

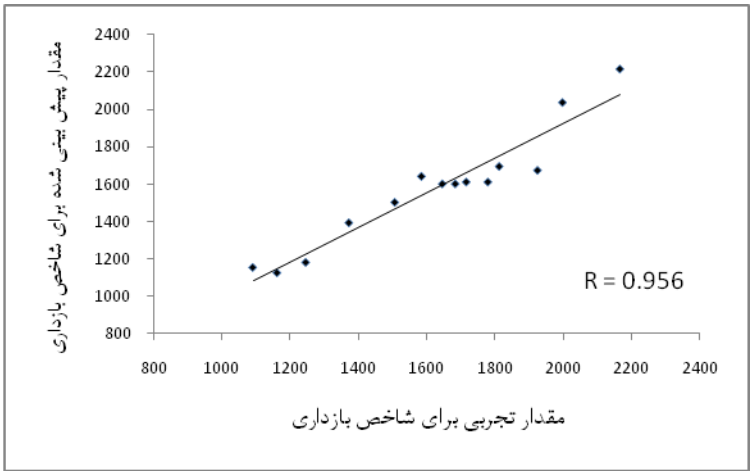
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۵	۱۱۰۲	۱۱۱۵/۵	۱۳/۵	۱/۲
۱۰	۱۱۷۹	۱۱۴۷/۱	-۳۱/۹	-۲/۷
۱۵	۱۲۶۰	۱۰۹۳/۰	-۱۶۷/۰	-۱۳/۲
۲۰	۱۳۹۲	۱۴۷۳/۶	-۸۱/۶	-۵/۹
۲۵	۱۵۳۰	۱۴۷۳/۰	-۵۷/۰	-۳/۷
۳۰	۱۶۰۶	۱۶۸۷/۱	۸۱/۱	۵/۰
۳۵	۱۶۵۵	۱۷۵۸/۲	۱۰۳/۲	۶/۲
۴۰	۱۶۹۲	۱۷۰۸/۳	۱۶/۳	۱/۰
۴۵	۱۷۲۴	۱۶۱۲/۰	-۱۱۲/۰	-۶/۵
۵۰	۱۷۸۷	۱۶۹۹/۸	-۸۷/۲	-۴/۹
۵۵	۱۸۵۶	۱۹۴۱/۵	۸۵/۵	۴/۶
۶۰	۱۹۴۷	۱۸۶۳/۹	-۸۳/۱	-۴/۳
۶۵	۲۰۸۳	۲۰۶۱/۷	-۲۱/۳	-۱/۰
۷۰	۲۱۷۶	۲۱۱۳/۲	-۶۲/۸	-۲/۹



شکل۴-۲- رسم مقادیر پیش‌بینی شده شاخص بازداری توسط روش MLR بر حسب مقادیر تجربی برای سری ارزیابی

جدول۴-۸- مقایسه مقادیر تجربی و پیش‌بینی شده شاخص بازداری با روش MLR برای سری تست

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۴	۱۰۹۲	۱۱۵۲/۲	۶۰/۲	۵/۵
۹	۱۱۶۳	۱۱۲۳/۴	-۳۹/۶	-۳/۴
۱۴	۱۲۴۷	۱۱۷۹/۴	-۶۷/۶	-۵/۴
۱۹	۱۳۷۳	۱۳۹۲/۰	۱۹/۰	۱/۴
۲۴	۱۵۰۷	۱۵۰۱/۶	-۵/۴	-۰/۴
۲۹	۱۵۸۵	۱۶۴۱/۱	۵۶/۱	۳/۵
۳۴	۱۶۴۶	۱۶۰۰/۴	-۴۵/۶	-۲/۷
۳۹	۱۶۸۴	۱۵۳۳/۲	-۱۵۰/۸	-۹/۰
۴۴	۱۷۱۶	۱۶۱۰/۴	-۱۰۵/۶	-۶/۲
۴۹	۱۷۷۹	۱۷۲۲/۷	-۵۶/۳	-۳/۲
۵۴	۱۸۱۲	۱۶۹۴/۵	-۱۱۷/۵	-۶/۵
۵۹	۱۹۲۴	۱۶۷۳/۱	-۲۵۰/۹	-۱۳/۰
۶۴	۱۹۹۷	۲۰۳۷/۴	۴۰/۴	۲/۰
۶۹	۲۱۶۵	۲۲۱۶/۷	۵۱/۷	۲/۴



شکل۴-۳- رسم مقادیر پیش‌بینی شده شاخص بازداری توسط روش MLR بر حسب مقادیر تجربی برای سری تست

۴-۳-۱- ارزیابی مدل به روش حذف مرحله‌ای تک تک داده‌ها

برای ارزیابی مدل به دست آمده علاوه بر استفاده از سری تست، روش حذف مرحله‌ای تک تک داده‌ها نیز به‌کار گرفته شد. مقادیر محاسبه‌شده با مقادیر تجربی مقایسه گردید و نمودار مربوطه نیز رسم شد (جدول ۴-۹ و شکل ۴-۴). نزدیک بودن این نتایج به نتایج به دست آمده از مدل منتخب نشان دهنده صحت مدل می‌باشد.

جدول۴-۹- ارزیابی مدل خطی به روش حذف مرحله‌ای تک تک داده‌ها

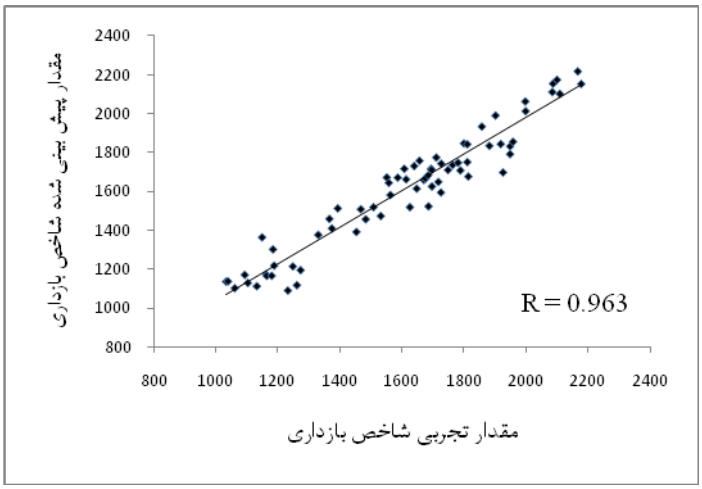
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۱	۱۰۳۲	۱۱۳۳/۷	۱۰۱/۷	۹/۸
۲	۱۰۳۸	۱۱۳۵/۲	۹۷/۲	۹/۴
۳	۱۰۶۰	۱۱۰۰/۱	۴۰/۱	۳/۸
۴	۱۰۹۲	۱۱۶۸/۶	۷۶/۶	۷/۰
۵	۱۱۰۲	۱۱۲۶/۸	۲۴/۸	۲/۲
۶	۱۱۳۱	۱۱۱۰/۳	-۲۰/۷	-۱/۸
۷	۱۱۴۸	۱۳۶۱/۹	۲۱۳/۹	۱۸/۶
۸	۱۱۶۱	۱۱۷۱/۱	۱۰/۱	۰/۹
۹	۱۱۶۳	۱۱۶۳/۸	۰/۸	۰/۱
۱۰	۱۱۷۹	۱۱۶۴/۲	-۱۴/۸	-۱/۳
۱۱	۱۱۸۳	۱۳۰۰/۰	۱۱۷/۰	۹/۹
۱۲	۱۱۸۷	۱۲۱۶/۱	۲۹/۱	۲/۴
۱۳	۱۲۳۱	۱۰۸۷/۷	-۱۴۳/۳	-۱۱/۶
۱۴	۱۲۴۷	۱۲۱۲/۲	-۳۴/۸	-۲/۸
۱۵	۱۲۶۰	۱۱۱۵/۴	-۱۴۴/۶	-۱۱/۵
۱۶	۱۲۷۲	۱۱۹۳/۰	-۷۹/۰	-۶/۲
۱۷	۱۳۲۹	۱۳۷۴/۹	۴۵/۹	۳/۴
۱۸	۱۳۶۵	۱۴۵۷/۲	۹۲/۲	۶/۸
۱۹	۱۳۷۳	۱۴۰۷/۷	۳۴/۷	۲/۵
۲۰	۱۳۹۲	۱۵۱۱/۰	۱۱۹/۰	۸/۵
۲۱	۱۴۵۲	۱۳۸۹/۵	۶۲/۵	۴/۳
۲۲	۱۴۶۶	۱۵۰۶/۹	۴۰/۹	۲/۸
۲۳	۱۴۸۲	۱۴۵۵/۰	-۲۷/۰	-۱/۸
۲۴	۱۵۰۷	۱۵۱۷/۵	۱/۵	۰/۷
۲۵	۱۵۳۰	۱۴۷۱/۹	-۵۸/۱	-۳/۸
۲۶	۱۵۵۰	۱۶۷۰/۴	۱۲۰/۴	۷/۸
۲۷	۱۵۵۶	۱۶۴۲/۶	۸۶/۶	۵/۶
۲۸	۱۵۶۱	۱۵۸۰/۱	۱۹/۱	۱/۲
۲۹	۱۵۸۵	۱۶۶۹/۴	۸۴/۴	۵/۳

ادامه جدول۴-۹

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۳۰	۱۶۰۶	۱۷۱۵/۱	۱۰۹/۱	۶/۸
۳۱	۱۶۱۲	۱۶۶۱	۴۹/۰	۳/۰
۳۲	۱۶۲۴	۱۵۱۷/۳	-۱۰۶/۷	-۶/۶
۳۳	۱۶۳۸	۱۷۲۹/۵	۹۱/۵	۵/۶
۳۴	۱۶۴۶	۱۶۱۲/۷	-۳۳/۳	-۲/۰
۳۵	۱۶۵۵	۱۷۵۶/۹	۱۰۱/۹	۶/۲
۳۶	۱۶۶۹	۱۶۵۸/۰	-۱۱/۰	-۰/۷
۳۷	۱۶۷۴	۱۶۶۵/۱	-۸/۹	-۰/۵
۳۸	۱۶۸۳	۱۶۸۱/۹	-۱/۱	-۰/۱
۳۹	۱۶۸۴	۱۵۲۱/۵	-۱۶۲/۵	-۹/۶
۴۰	۱۶۹۲	۱۷۱۵/۳	۲۳/۳	۱/۴
۴۱	۱۶۹۴	۱۷۱۰/۱	۱۶/۱	۱/۰
۴۲	۱۶۹۵	۱۶۲۴/۰	-۷۱/۰	-۴/۴
۴۳	۱۷۰۹	۱۷۷۳/۴	۶۴/۴	۳/۸
۴۴	۱۷۱۶	۱۶۴۷/۶	-۶۸/۴	-۴/۰
۴۵	۱۷۲۴	۱۵۹۳/۱	-۱۳۰/۹	-۷/۶
۴۶	۱۷۲۵	۱۷۴۱/۶	۱۶/۶	۱/۰
۴۷	۱۷۴۷	۱۷۰۸/۹	-۳۸/۱	-۱/۸
۴۸	۱۷۶۲	۱۷۳۵/۲	-۲۶/۸	-۲/۳
۴۹	۱۷۷۹	۱۷۴۶/۶	-۳۲/۴	-۱/۸
۵۰	۱۷۸۷	۱۷۰۶/۶	-۸۰/۴	-۴/۵
۵۱	۱۷۹۸	۱۸۴۶/۱	۴۸/۱	۲/۷
۵۲	۱۸۰۹	۱۷۴۹/۷	-۵۹/۳	-۳/۳
۵۳	۱۸۰۹	۱۸۴۱/۴	۳۲/۴	۱/۸
۵۴	۱۸۱۲	۱۶۷۵/۳	-۱۳۶/۷	۵
۵۵	۱۸۵۶	۱۹۳۲/۸	۷۶/۸	۴/۱
۵۶	۱۸۸۰	۱۸۳۳/۸	-۴۶/۲	-۲/۵
۵۷	۱۹۰۰	۱۹۸۹/۷	۸۹/۷	۴/۷
۵۸	۱۹۱۷	۱۸۴۲/۶	-۷۴/۴	-۳/۹

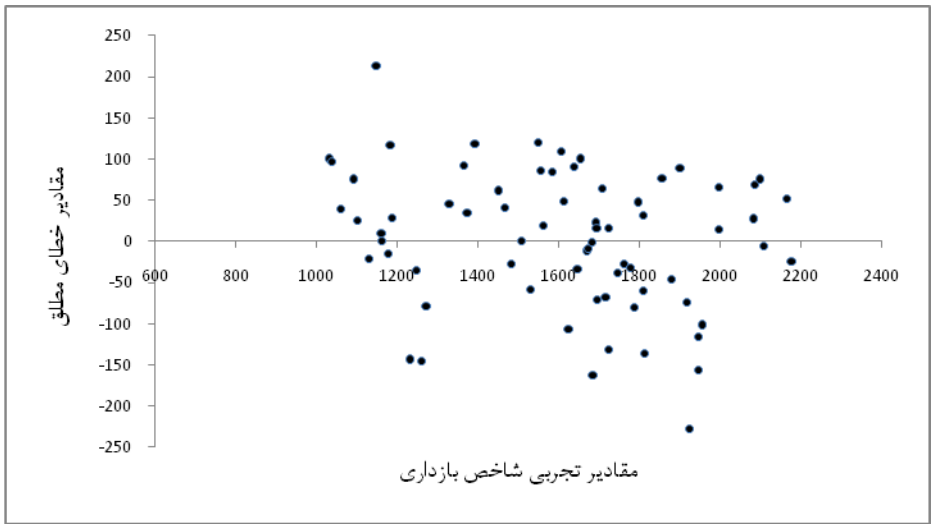
ادامه جدول۴-۹

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۵۹	۱۹۲۴	۱۶۹۶/۵	-۲۲۷/۵	-۱۱/۸
۶۰	۱۹۴۷	۱۸۳۱	-۱۱۶/۰	-۶/۰
۶۱	۱۹۴۷	۱۷۹۰/۷	-۱۵۶/۳	-۸/۰
۶۲	۱۹۵۶	۱۸۵۴/۲	-۱۰۱/۸	-۵/۲
۶۳	۱۹۹۶	۲۰۶۱/۷	۶۵/۷	۳/۳
۶۴	۱۹۹۷	۲۰۱۲/۳	۱۵/۳	۰/۸
۶۵	۲۰۸۳	۲۱۱۰/۷	۲۷/۷	۱/۳
۶۶	۲۰۸۵	۲۱۵۳/۹	۶۸/۹	۳/۳
۶۷	۲۰۹۸	۲۱۷۳/۸	۷۵/۸	۳/۶
۶۸	۲۱۰۷	۲۱۰۱/۴	-۵/۶	-۰/۳
۶۹	۲۱۶۵	۲۲۱۷/۰	۵۲/۰	۲/۴
۷۰	۲۱۷۶	۲۱۵۲/۲	-۲۳/۸	-۱/۱



شکل۴-۴- رسم مقادیر پیش بینی شده شاخص بازداری در ارزیابی مدل بر حسب مقادیر تجربی برای سری داده‌ها

همچنین نمودار مقادیر باقیمانده محاسبه شده در ارزیابی مدل MLR بر حسب مقادیر تجربی شاخص بازداری نیز در شکل ۴-۵ رسم شد. تقارن نسبی داده‌ها حول خط صفر نشان‌دهنده عدم وجود خطای معین می‌باشد.



شکل۴-۵- رسم مقادیر خطای مطلق در ارزیابی مدل MLR به روش LOO در مقابل مقادیر تجربی شاخص بازداری

۴-۴- مدل‌سازی با استفاده از شبکه عصبی مصنوعی(ANN)

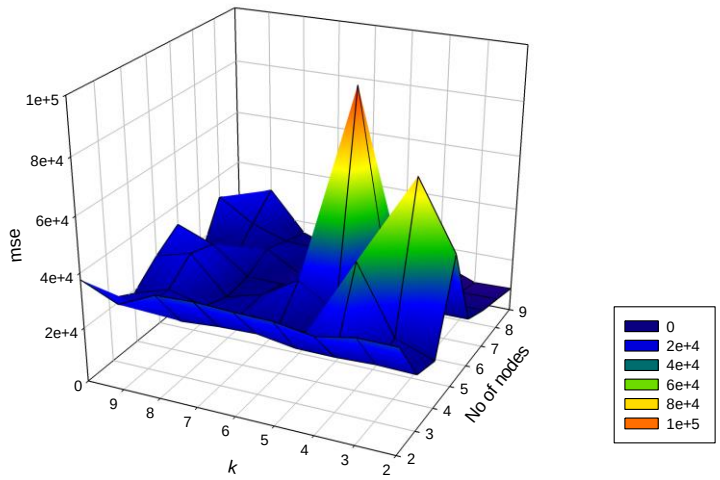
از آنجایی که ممکن است مدل خطی بهترین مدل برای توصیف رفتار سری داده‌ها نباشد، شبکه عصبی مصنوعی برای مدل‌سازی غیرخطی داده‌ها مورد استفاده قرار گرفت. در این تحقیق برنامه نویسی و اجرای یک شبکه سه لایه در محیط نرم افزار MATLAB نسخه 2008b صورت گرفت. در این روش تابع کارایی شبکه، میانگین مجذور خطای استاندارد (mse) می‌باشد. برای به دست آوردن بهترین معادله و کمترین خطا، پارامترهای شبکه (تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه، تعداد گره‌ها در لایه مخفی، مومنتم و تعداد دورهای آموزش) بهینه‌سازی شدند.

۴-۴-۱- بهینه‌سازی تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه و تعداد نرون‌های لایه مخفی

برای بهینه‌سازی و شناسایی بهترین پارامترهای شبکه، شبکه‌هایی با ورودی‌های از ۲ تا ۹ توصیف‌کننده ایجاد شدند. هر شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تنظیم بائزین با تعداد گره‌های از ۲ تا ۱۰ در لایه پنهان آموزش داده شد. در حالیکه برای به دست آوردن بهترین تابع انتقال در لایه پنهان مدل‌های شبکه عصبی نیز از توابع لگاریتم زیگموئید، تانژانت هایپربولیک و خطی به عنوان توابع انتقال استفاده شد. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطای شبکه به عنوان معیار انتخاب گردید. در تمام شبکه‌ها تعداد دورهای آموزشی یکسان و برابر ۱۰ در نظر گرفته شد. به این ترتیب مجموعاً ۴۳۲ شبکه سه لایه آموزش داده شده و سپس برای پیش‌بینی سری ارزیابی مورد استفاده قرار گرفتند. نتایج این پیش‌بینی در جداول ۴-۱۰ تا ۴-۱۵ آورده شده و نمودار سه‌بعدی این جداول نیز در شکل‌های ۴-۶ تا ۴-۱۱ رسم شده است. با توجه به نتایج به دست آمده، الگوریتم آموزشی تنظیم بائزین نسبت به الگوریتم آموزشی لونبرگ مارکوات دارای میانگین مربع خطاهای کمتری می‌باشد. این الگوریتم علاوه بر سرعت آموزش بسیار زیاد دارای قدرت تعمیم‌پذیری بالا می‌باشد، به این معنی که قادر است میان ترکیباتی که در آموزش به کار گرفته شده و آنهایی که در آموزش به کار گرفته نشده، تطابق خوبی برقرار نماید. همانطور که در جداول ۴-۱۰ تا ۴-۱۵ و شکل‌های ۴-۶ تا ۴-۱۱ ملاحظه می‌شود، در این تحقیق شبکه با تابع آموزش trainbr، تابع انتقال logsig، ۹ متغیر ورودی و تعداد ۳ نرون در لایه مخفی، کمترین میانگین خطاها را نشان داده و به عنوان شبکه بهینه انتخاب شد.

جدول ۴- ۱۰- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی بائزین (trainbr) و تابع تبدیل لگاریتم زیگموئید (logsig) با تعداد متغیر های ورودی متفاوت و تعداد گره های مختلف در لایه پنهان

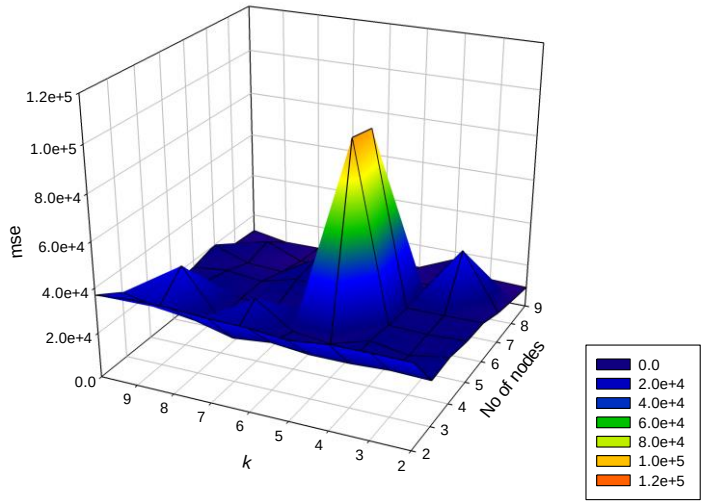
		تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K							
		۲	۳	۴	۵	۶	۷	۸	۹
تعداد گره‌ها	۲	۳۷۷۷۲	۲۰۴۷۲	۱۷۴۹۷	۴۸۵۲۷	۱۷۹۶۰	۱۴۰۸۷	۱۱۴۰۳	۸۶۰۲
	۳	۳۷۸۸۶	۲۰۸۶۰	۲۲۷۰۲	۷۲۷۷۵	۱۷۹۱۱	۱۴۰۸۷	۱۱۱۷۸	**۸۲۳۷**
	۴	۳۵۲۵۱	۲۰۳۰۰	۴۷۱۰۷	۱۸۶۳۲	۱۷۷۳۷	۱۵۱۸۴	۱۱۱۵۰	۱۳۶۷۰
	۵	۳۷۴۳۳	۲۰۴۳۶	۱۸۰۸۸	۱۸۸۸۷	۹۴۱۲۳	۱۵۸۴۰	۱۵۲۴۲	۱۰۵۸۴
	۶	۳۷۷۰۴	۲۲۶۳۶	۱۷۹۴۸	۲۸۰۸۹	۱۷۱۰۳	۲۷۷۷۴	۱۲۰۴۱	۹۲۷۱
	۷	۳۸۳۳۲	۲۲۲۱۰	۱۹۵۲۳	۱۹۹۱۱	۱۶۸۹۲	۱۳۷۱۷	۱۷۷۵۴	۱۰۷۷۶
	۸	۳۷۷۷	۲۱۱۷۴	۱۹۶۶۱	۱۶۰۶۰	۱۷۱۹۳	۲۱۴۷۸	۱۸۳۶۷	۱۵۴۵۵
	۹	۳۱۶۰۶	۲۳۴۵۴	۲۳۰۱۳	۳۰۱۴۸	۲۷۱۹۷	۱۴۹۸۲	۲۲۹۲۷	۳۰۷۸۴
	۱۰	۳۷۶۶۹	۲۳۷۸۴	۱۸۷۹۱	۲۶۴۸۰	۳۲۵۰۳	۱۳۵۱۸	۳۱۳۷۷	۲۵۶۴۶



شکل ۴-۶- نمایش سه بعدی شبکه هایی با الگوریتم آموزشی بائزین و تابع تبدیل لگاریتم زیگموئید

جدول ۴-۱۱- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی بائزین (trainbr) و تابع تبدیل تانژانت هایپربولیک (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

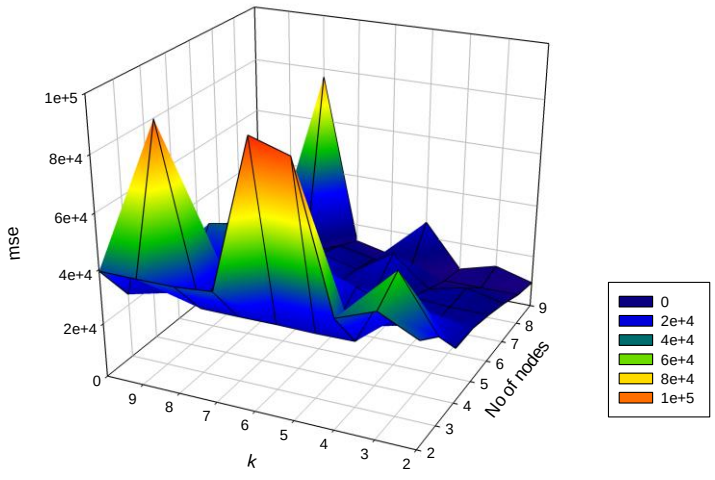
		تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K							
		۲	۳	۴	۵	۶	۷	۸	۹
تعداد گره‌ها	۲	۳۷۴۲۵	۲۰۲۳۹	۲۰۸۲۳	۱۸۶۱۸	۱۷۹۰۳	۱۳۸۸۳	۱۱۳۷۷	۹۱۰۵
	۳	۳۷۵۲۱	۲۰۷۵۹	۱۷۷۴۰	۱۸۰۳۹	۱۷۸۱۶	۳۹۴۸۴	۱۱۲۳۴	۹۶۹۶
	۴	۳۷۳۷۸	۱۹۸۸۲	۲۰۵۱۶	۱۸۵۹۳	۱۷۷۰۶	۱۸۰۵۰	۱۱۲۵۹	۱۶۳۳۹
	۵	۳۵۳۶۸	۱۹۸۶۳	۱۵۴۵۸	۹۶۶۴۳	۹۴۲۶۸	۱۴۲۶۷	۱۰۳۱۷	۹۰۴۸
	۶	۳۶۹۵۴	۲۱۵۵۰	۱۷۴۰۷	۱۷۴۴۳	۱۹۵۳۶	۱۶۵۴۴	۱۳۵۰۴	۸۳۳۴
	۷	۳۷۲۴۳	۱۹۸۰۴	۲۸۸۰۹	۱۶۶۳۱	۲۱۰۲۶	۲۲۴۷۷	۱۱۲۴۰	۸۹۰۸
	۸	۳۸۲۷۷	۲۴۰۰۲	۱۶۳۹۸	۱۶۸۸۹	۱۶۶۷۸	۱۳۱۳۳	۱۰۳۷۰	۹۰۷۲
	۹	۳۸۳۷۱	۲۶۲۰۸	۳۶۸۹۳	۲۱۲۸۴	۱۸۵۳۷	۱۴۱۷۹	۱۶۹۷۶	۸۹۷۳
	۱۰	۳۶۹۳۰	۲۹۷۰۹	۱۹۳۵۷	۱۵۶۷۰	۱۶۹۵۷	۱۹۹۲۷	۱۲۸۱۴	۱۱۹۴۲



شکل ۴-۷- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بائزین و تابع تبدیل تانژانت هایپربولیک

جدول ۴-۱۲- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی بائزین (trainbr) و تابع تبدیل خطی (purelin) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

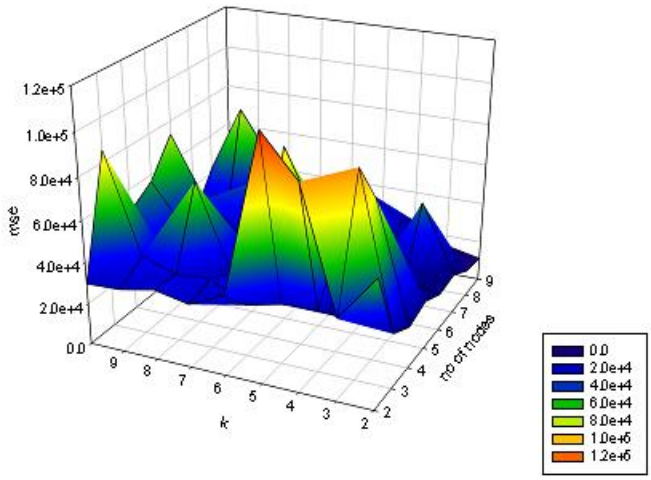
		تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K							
		۲	۳	۴	۵	۶	۷	۸	۹
تعداد گره‌ها	۲	۳۸۵۴۱	۳۲۲۴۸	۲۰۴۴۹	۲۰۲۴۶	۱۷۹۴۲	۱۵۰۹۷	۱۱۴۷۱	۹۴۱۴
	۳	۴۵۴۱۳	۵۲۳۰۹	۳۲۴۲۸	۱۸۶۸۲	۱۷۹۵۴	۱۸۴۲۴	۱۱۴۲۲	۱۳۳۹۵
	۴	۴۰۰۰۸	۲۴۳۹۴	۲۲۳۳۲	۲۱۱۸۴	۱۹۰۰۵	۱۵۱۰۹	۱۱۴۹۴	۹۴۹۰
	۵	۹۰۵۷۶	۲۴۲۰۳	۲۰۴۵۴	۲۷۰۸۱	۱۸۰۲۳	۲۶۲۶۳	۱۱۸۷۱	۲۶۱۵۴
	۶	۹۴۹۵۰	۲۴۴۱۵	۴۲۴۱۴	۱۹۵۰۳	۱۸۷۹۷	۱۷۳۲۶	۱۴۸۷۰	۱۰۰۷۹
	۷	۴۰۵۱۶	۲۴۳۳۴	۲۵۱۵۸	۲۰۰۶۶	۳۰۶۲۰	۲۱۰۰۵	۱۷۷۱۴	۱۳۲۷۶
	۸	۳۹۳۴۱	۲۴۴۱۱	۳۱۹۷۲	۴۲۷۸۴	۲۱۰۶۶	۱۶۲۱۸	۱۲۰۷۱	۷۶۹۹۵
	۹	۳۹۳۲۴	۸۸۸۳۸	۲۰۵۱۸	۴۰۲۸۳	۱۹۷۷۵	۳۹۷۵۴	۱۱۷۲۴	۹۹۳۵
	۱۰	۳۹۳۸۳	۲۳۸۹۹	۲۵۱۴۷	۲۰۲۰۵	۱۸۷۲۵	۱۶۸۴۷	۱۱۴۳۶	۹۴۴۵



شکل ۴-۸- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بائزین و تابع تبدیل خطی

جدول ۴-۱۳- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت (trainlm) و تابع تبدیل لگاریتم زیگموئید(logsig) با تعداد متغیر های ورودی متفاوت و تعداد گره های مختلف در لایه پنهان

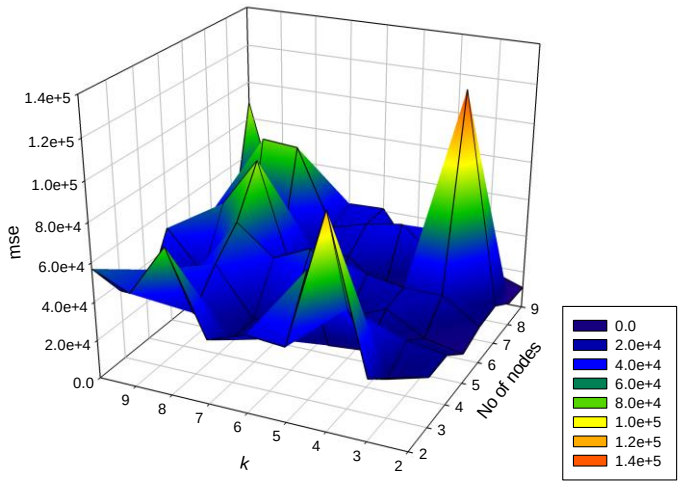
		تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K							
		۲	۳	۴	۵	۶	۷	۸	۹
تعداد گره‌ها	۲	۶۰۷۸۵	۲۶۳۷۴	۲۰۰۸۴	۲۳۳۷۱	۱۸۰۶۰	۱۹۴۱۲	۱۲۸۳۳	۱۱۱۹۳
	۳	۳۸۹۰۶	۹۷۲۶۰	۷۶۵۹۴	۱۸۳۶۳	۲۱۵۷۰	۵۲۴۰۸	۲۳۹۸۱	۱۴۷۶۷
	۴	۹۵۳۵۳	۲۹۸۸۱	۱۷۸۱۳	۴۳۲۱۲	۱۵۸۶۵	۱۴۹۸۷	۳۲۵۶۸	۲۴۸۹۸
	۵	۱۱۴۱۶۹	۲۹۰۰۳	۲۶۲۸۱	۵۱۳۴۴	۳۴۹۳۹	۲۹۰۴۵	۴۴۳۴۶	۱۱۸۰۲
	۶	۳۸۳۷۵	۲۴۹۰۴	۳۳۷۸۵	۸۴۸۱۸	۵۵۵۵۴	۴۷۰۵۷	۳۳۵۸۳	۱۵۱۳۲
	۷	۳۰۶۶۲	۳۴۰۲۶	۲۸۴۱۱	۳۰۴۵۱	۴۴۳۴۴	۵۴۸۸۱	۲۹۱۴۳	۲۴۴۹۵
	۸	۳۳۹۹۹	۳۲۸۴۷	۶۹۵۶۰	۲۷۳۶۹	۳۰۵۵۶	۴۸۳۸۵	۵۲۴۰۵	۱۳۸۳۱
	۹	۳۰۵۱۶	۳۸۳۸۸	۳۰۵۰۸	۳۵۲۱۹	۴۳۴۰۲	۱۷۳۴۱	۷۶۰۳۶	۳۶۷۹۹
	۱۰	۲۹۸۰۳	۸۵۴۶۶	۴۹۱۸۴	۵۶۷۷۴	۷۳۴۷۳	۲۳۵۳۸	۴۲۴۰۲	۳۰۹۹۹



شکل ۴- ۹- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ -مارکوارت و تابع تبدیل لگاریتم زیگموئید

جدول ۴-۱۴- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت (trainlm) و تابع تبدیل تانژانت هایپربولیک (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

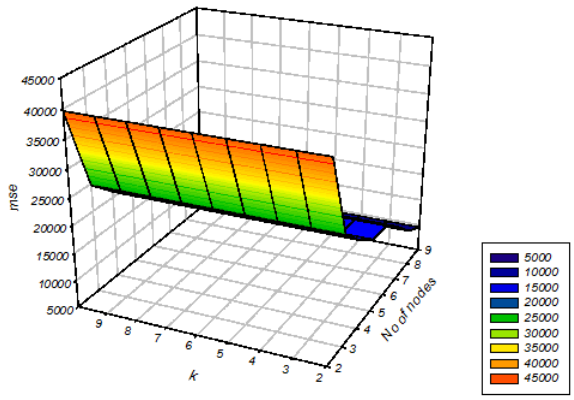
		تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K							
		۲	۳	۴	۵	۶	۷	۸	۹
تعداد گره‌ها	۲	۳۷۰۰۰	۲۵۷۱۲	۲۵۸۰۱	۱۵۵۹۲	۱۹۵۳۵	۲۰۵۲۸	۱۲۰۵۵	۱۱۰۷۹
	۳	۳۱۵۸۰	۲۰۱۶۵	۲۹۱۴۷	۱۹۵۱۰	۳۲۰۷۷	۱۲۷۹۷۹	۱۴۱۶۷	۱۷۹۳۶
	۴	۱۰۶۸۳۳	۵۱۱۹۴	۱۹۳۸۴	۳۰۸۳۵	۳۲۴۶۲	۱۱۶۹۶	۲۹۰۰۹	۳۵۱۳۲
	۵	۳۸۳۳۷	۶۶۴۲۹	۱۷۷۴۸	۲۲۲۲۴	۳۵۶۸۰	۱۶۷۳۱	۴۴۶۸۴	۱۵۹۵۳
	۶	۴۰۰۷۹	۲۷۵۹۶	۲۹۲۰۱	۵۴۷۲۷	۱۹۴۳۰	۳۲۱۳۷	۴۱۳۱۴	۴۱۴۲۰
	۷	۳۳۲۸۳	۲۲۸۱۷	۲۲۸۶۶	۴۵۹۹۰	۲۴۴۸۷	۳۲۳۲۴	۳۵۷۶۷	۴۰۸۲۶
	۸	۷۵۳۷۴	۳۲۴۹۳	۴۸۳۱۸	۶۱۱۲۶	۸۵۱۳۶	۱۹۳۲۹	۷۶۹۱۸	۳۲۴۴۴
	۹	۴۸۴۰۵	۵۰۹۵۳	۳۰۴۶۸	۳۰۰۲۳	۵۹۲۱۷	۱۰۴۲۱	۷۸۵۲۷	۲۱۴۳۵
	۱۰	۵۶۹۱۸	۳۵۹۹۰	۲۲۶۳۸	۵۱۰۰۸	۳۵۶۴۴	۴۲۴۷۵	۲۴۲۸۷	۸۸۶۲۵



شکل۴-۱۰- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ- مارکوارت و تابع تبدیل تانژانت هایپربولیک

جدول ۴- ۱۵- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت (trainlm) و تابع تبدیل خطی (purelin) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

		تعداد متغیرهای ورودی به شبکه یا توصیف کننده‌ها K							
		۲	۳	۴	۵	۶	۷	۸	۹
تعداد گره‌ها	۲	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۳	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۴	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۵	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۶	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۷	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۸	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۹	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴
	۱۰	۳۹۶۶۹	۲۴۷۰۷	۲۱۰۰۶	۱۸۲۹۴	۱۷۹۸۸	۱۵۱۳۳	۱۱۵۹۴	۱۱۵۹۴



شکل ۴-۱۱- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ -مارکوارت و تابع تبدیل خطی

۴-۴-۲- بهینه‌سازی پارامتر mu

در این مرحله تابع آموزش trainbr، تابع انتقال logsig و تعداد ۳ نرون در لایه مخفی به عنوان تعداد بهینه نرون‌های شبکه قرار داده شد و برای بهینه سازی mu مقدار این پارامتر بین ۰/۰۰۰۱ تا ۱ تغییر داده شد و سپس برای هر مورد مقدار mse سری‌های ارزیابی و تست محاسبه شده و بر حسب mu رسم گردید (جدول۴-۱۶). نتایج نشان می‌دهد که مقدار ۰/۰۰۵ برای پارامتر mu که همان مقدار پیش‌فرض الگوریتم بائزین نیز می‌باشد، می‌تواند به عنوان نقطه بهینه در نظر گرفته شود (جدول۴-۱۰).

جدول۴-۱۶- مقادیر خطای شبکه با تغییر mu

mu	mse سری ارزیابی
۰/۰۰۰۱	۶۳۹۸/۹۶
۰/۰۰۰۲	۸۱۶۹/۸۱
۰/۰۰۰۳	۸۶۴۴/۲۳
۰/۰۰۰۴	۴۸۵۱۰/۵۴
۰/۰۰۰۵	**۶۱۴۸/۸۴**
۰/۰۰۰۶	۶۲۲۸/۴۶
۰/۰۰۰۷	۲۱۸۱۴/۸۴
۰/۰۰۰۸	۲۱۴۹۸/۳۹
۰/۰۰۰۹	۶۴۲۰/۲۷
۰/۰۰۱	۶۳۹۸/۹۶
۰/۰۰۲	۸۱۶۹/۸۱
۰/۰۰۳	۸۶۴۴/۲۳
۰/۰۰۴	۴۸۵۱۰/۵۵
۰/۰۰۵	۶۱۴۸/۸۴
۰/۰۰۶	۶۲۲۸/۴۶
۰/۰۰۷	۲۱۸۱۴/۸۴
۰/۰۰۸	۲۱۴۹۸/۳۹
۰/۰۰۹	۶۴۲۰/۲۷

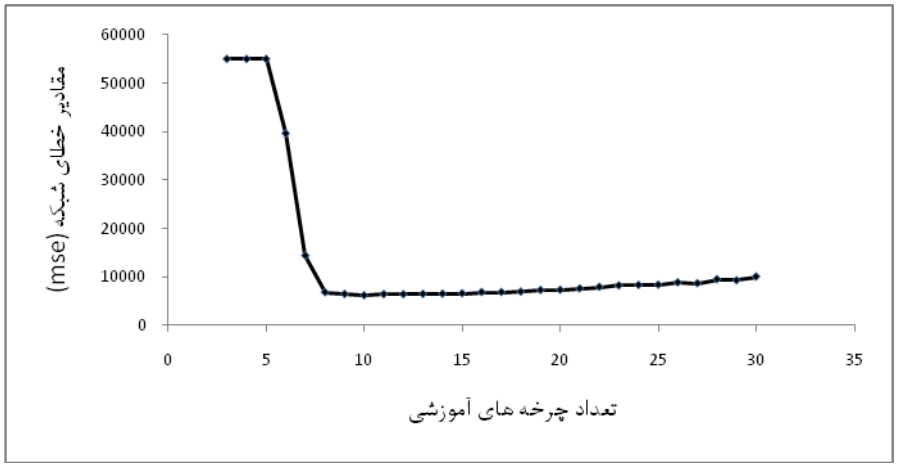
mu	mse سری ارزیابی
۰/۰۱	۶۳۹۸/۹۶
۰/۰۲	۸۱۶۹/۸۱
۰/۰۳	۸۶۴۴/۲۳
۰/۰۴	۴۸۵۱۰/۵۵
۰/۰۵	۶۱۴۸/۸۳
۰/۰۶	۶۲۲۸/۴۶
۰/۰۷	۲۱۸۱۴/۸۴
۰/۰۸	۲۱۴۹۸/۳۹
۰/۰۹	۶۴۲۰/۲۷
۰/۱	۶۳۹۸/۹۶
۰/۲	۸۱۶۹/۸۱
۰/۳	۸۶۴۴/۲۳
۰/۴	۴۸۵۱۰/۵۵
۰/۵	۶۱۴۸/۸۳
۰/۶	۶۲۲۸/۴۶
۰/۷	۲۱۸۱۴/۸۴
۰/۸	۲۱۴۹۸/۳۹
۰/۹	۶۴۲۰/۲۷
۱	۶۳۹۸/۹۶

۴-۴-۳- بهینه سازی تعداد چرخه‌های آموزشی

تعداد چرخه های آموزشی پارامتری است که برای جلوگیری از آموزش بیش از حد شبکه بهینه‌سازی می‌شود. برای این منظور، ابتدا مقدار بهینه تعداد نرون‌ها (۳) و مقدار بهینه مومنتم (۰/۰۰۵) در شبکه مورد نظر با تابع آموزش قرار داده شد و با تغییر تعداد چرخه‌های آموزشی از ۳ تا ۳۰، مقادیر خطای شبکه ها به دست آمد. این نتایج برای در جدول ۴-۱۷ سری ارزیابی و تست آورده شده است.

جدول ۴-۱۷- مقادیر خطای شبکه با تغییر تعداد چرخه‌های آموزشی

تعداد چرخه‌های آموزشی	mse سری ارزیابی
۳	۵۵۰۹۸/۱۴
۴	۵۵۰۹۸/۱۴
۵	۵۵۰۹۸/۱۴
۶	۳۹۶۸۵/۵۷
۷	۱۴۳۸۴/۶۰
۸	۶۷۹۳/۹۳
۹	۶۴۲۲/۴۴
۱۰	**۶۱۴۸/۸۳**
۱۱	۶۳۸۷/۲۷
۱۲	۶۳۶۵/۶۹
۱۳	۶۴۴۷/۲۳
۱۴	۶۴۷۳/۳۸
۱۵	۶۵۶۵/۴۴
۱۶	۶۸۲۶/۹۱
۱۷	۶۸۱۹/۲۶
۱۸	۶۹۱۲/۰۵
۱۹	۷۲۳۲/۸۹
۲۰	۷۲۹۵/۵۹
۲۱	۷۵۶۷/۰۷
۲۲	۷۹۰۱/۵۲
۲۳	۸۱۸۷/۰۳
۲۴	۸۲۸۴/۳۳
۲۵	۸۳۳۹/۰۵
۲۶	۸۸۴۴/۸۴
۲۷	۸۶۶۳/۶۸
۲۸	۹۴۹۳/۳۸
۲۹	۹۲۵۵/۱۳
۳۰	۱۰۰۳۱/۷۲



شکل ۴-۱۲- منحنی مقادیر خطای شبکه در مقابل تعداد چرخه‌های آموزشی

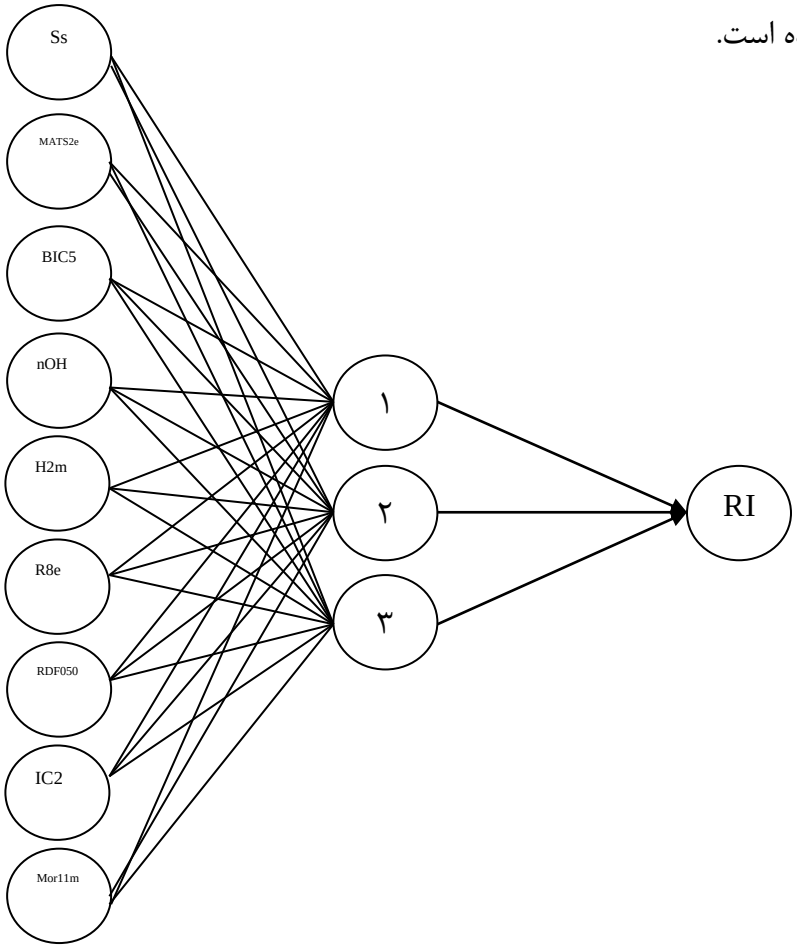
با توجه به نتایج به دست آمده و منحنی رسم شده برای میانگین مربعات خطای سری ارزیابی (شکل ۴-۱۲) تعداد بهینه چرخه‌های آموزشی ۱۰ انتخاب گردید.

مشخصات شبکه بهینه در جدول ۴-۱۸ نشان داده شده است.

جدول۴-۱۸- مشخصات شبکه بهینه

تابع آموزش	trainbr
تابع انتقال لایه پنهان	logsig
تابع انتقال لایه خروجی	purelin
تابع کارآیی	mse
تابع مقدار اولیه وزن‌دهی و بایاس	rands
بهبود تعمیم	Early Stopping
تعداد نرون های لایه پنهان	۳
پارامتر mu	۰/۰۰۵
تعداد چرخه‌های آموزشی	۱۰

ساختار شبکه عصبی به دست آمده پس از بهینه سازی فاکتورهـای بحـث شـده در بـالا بـه طـور شـماتیک در شکل۴-۱۳ نشان داده شده است.



شکل ۴-۱۳- تصویر شماتیک ساختار شبکه عصبی مصنوعی به دست آمده پس از بهینه سازی

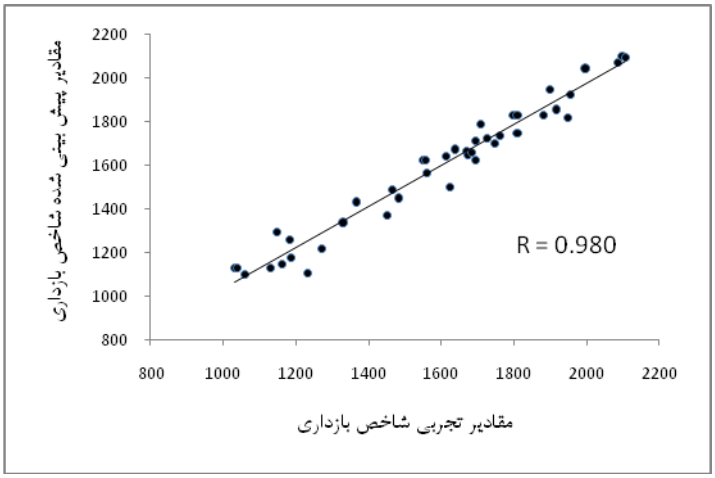
با استفاده از این شبکه مقادیر شاخص بازداری برای سری ارزیـابی و تسـت پـیش‌بینـی شـد. در جـدول ۴-۱۹، ۴-۲۰ و ۴-۲۱ به ترتیب مقادیر محاسبه شده برای سری آموزش، سری ارزیابی و سری تست آورده شده اسـت. نمودارهای مقادیر پیش‌بینی شده در مقابل مقادیر تجربی شاخص بازداری برای سری آموزش، ارزیـابی و تسـت نیز به ترتیب در شکل‌های ۴-۱۴، ۴-۱۵ و ۴-۱۶ رسم شد. ضریب همبسـتگی بـرای نمودارهـای یـاد شـده بـه ترتیب ۰/۹۸۰، ۰/۹۷۰ و ۰/۹۶۴ می‌باشد.

جدول ۴-۱۹- مقایسه مقادیر تجربی و پیش‌بینی شده برای سری آموزش

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطلق	درصد خطای نسبی
۱	۱۰۳۲	۱۱۲۹/۲	۹۷/۲	۹/۴
۲	۱۰۳۸	۱۱۲۷/۷	۸۹/۷	۸/۶
۳	۱۰۶۰	۱۱۰۱/۶	۴۱/۶	۳/۹
۶	۱۱۳۱	۱۱۳۰/۳	-۰/۷	-۰/۱
۷	۱۱۴۸	۱۲۹۴/۸	۱۴۶/۸	۱۲/۸
۸	۱۱۶۱	۱۱۴۷/۶	-۱۳/۴	-۱/۲
۱۱	۱۱۸۳	۱۲۵۷/۰	۸۴/۰	۷/۱
۱۲	۱۱۸۷	۱۱۷۷/۲	-۹/۸	-۰/۸
۱۳	۱۲۳۱	۱۱۰۳/۹	-۱۲۷/۱	-۱۰/۳
۱۶	۱۲۷۲	۱۲۱۷/۵	-۵۴/۵	-۴/۳
۱۷	۱۳۲۹	۱۳۳۷/۹	۸/۹	۰/۷
۱۸	۱۳۶۵	۱۴۳۲/۵	۶۷/۵	۴/۹
۲۱	۱۴۵۲	۱۳۷۱/۹	-۸۰/۱	-۵/۵
۲۲	۱۴۶۶	۱۴۹۰/۳	۲۴/۳	-۱/۶
۲۳	۱۴۸۲	۱۴۵۰/۷	-۳۱/۳	-½
۲۶	۱۵۵۰	۱۶۲۶/۳	۷۶/۳	۴/۹
۲۷	۱۵۵۶	۱۶۲۴/۹	۶۸/۹	۴/۴
۲۸	۱۵۶۱	۱۵۶۳/۷	۲/۷	۰/۲
۳۱	۱۶۱۲	۱۶۴۰	۲۸	۱/۷
۳۲	۱۶۲۴	۱۵۰۰/۵	-۱۲۳/۵	-۷/۶
۳۳	۱۶۳۸	۱۶۷۴/۹	۳۶/۹	۲/۲
۳۶	۱۶۶۹	۱۶۶۷/۰	-۲	-۰/۱
۳۷	۱۶۷۴	۱۶۴۷/۳	۲۶/۷	۱/۶
۳۸	۱۶۸۳	۱۶۵۹/۲	-۲۳/۸	-۱/۴
۴۱	۱۶۹۴	۱۷۱۱/۶	۱۷/۶	۱
۴۲	۱۶۹۵	۱۶۲۴/۷	-۷۰/۳	-۴/۱
۴۳	۱۷۰۹	۱۷۸۹/۱	۸۰/۱	۴/۷
۴۶	۱۷۲۵	۱۷۲۶/۳	۱/۳	۰/۱

ادامه جدول ۴-۱۹

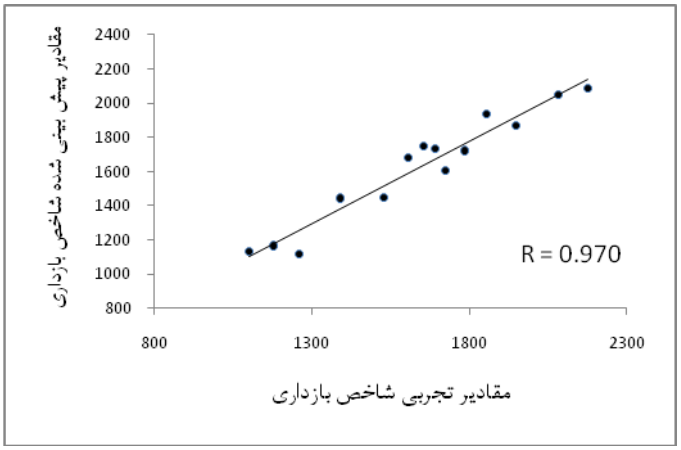
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	تفاضل	درصد خطای نسبی
۴۷	۱۷۴۷	۱۷۰۳/۰	-۴۴	-۲/۵
۴۸	۱۷۶۲	۱۷۳۶/۵	-۲۵/۵	-۱/۵
۵۱	۱۷۹۸	۱۸۳۲/۱	۳۴/۱	۱/۹
۵۲	۱۸۰۹	۱۷۴۷/۵	-۶۱/۵	-۳/۴
۵۳	۱۸۰۹	۱۸۲۹/۱	۲۰/۱	۱/۱
۵۶	۱۸۸۰	۱۸۳۲/۱	-۴۷/۹	-۲/۵
۵۷	۱۹۰۰	۱۹۵۰/۹	۵۰/۹	۲/۷
۵۸	۱۹۱۷	۱۸۵۸/۰	-۵۹	-۳/۱
۶۱	۱۹۴۷	۱۸۲۰/۰	-۱۲۷/۰	-۶/۵
۶۲	۱۹۵۶	۱۹۲۳/۳	-۳۲/۷	-۱/۷
۶۳	۱۹۹۶	۲۰۴۶/۰	۵۰	۲/۵
۶۶	۲۰۸۵	۲۰۷۴/۳	-۱۰/۷	-۰/۵
۶۷	۲۰۹۸	۲۱۰۳/۴	۵/۴	۰/۳
۶۸	۲۱۰۷	۲۰۹۵/۸	-۱۱/۲	-۰/۵



شکل۴-۱۴- رسم مقادیر پیش‌بینی شده توسط روش ANN بر حسب مقادیر تجربی برای سری آموزش

جدول ۴-۲۰- مقایسه مقادیر تجربی و پیش‌بینی شده برای سری ارزیابی

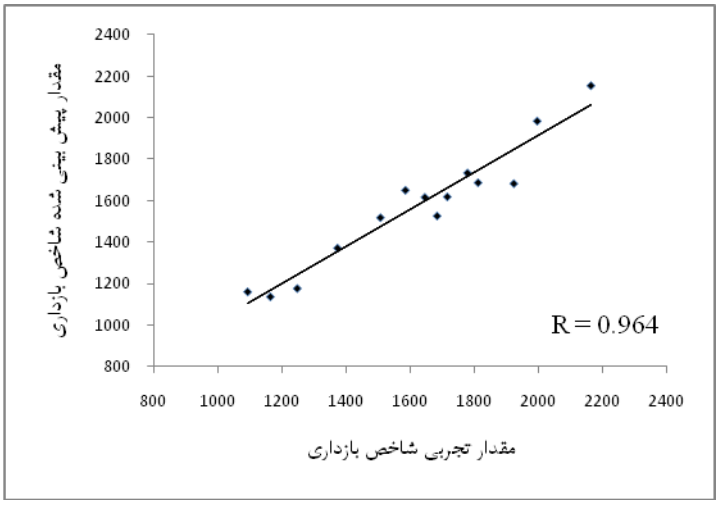
شماره ترکیب	مقدار تجربی	مقدار محاسبه شده	مقدار تفاضل	درصد خطای نسبی
۵	۱۱۰۲	۱۱۳۳/۲	۳۱/۲	۲/۸
۱۰	۱۱۷۹	۱۱۶۵/۲	-۱۳/۸	-۱/۲
۱۵	۱۲۶۰	۱۱۱۴/۴	-۱۴۵/۶	-۱۱/۶
۲۰	۱۳۹۲	۱۴۴۴/۷	۵۲/۷	۳/۸
۲۵	۱۵۳۰	۱۴۴۹/۳	-۸۰/۷	-۵/۳
۳۰	۱۶۰۶	۱۶۸۰/۱	۷۴/۱	۴/۶
۳۵	۱۶۵۵	۱۷۴۸/۴	۹۳/۴	۵/۶
۴۰	۱۶۹۲	۱۷۳۱/۹	۳۹/۹	۲/۴
۴۵	۱۷۲۴	۱۶۰۸/۵	-۱۱۵/۵	-۶/۷
۵۰	۱۷۸۷	۱۷۲۱/۶	-۶۵/۴	-۳/۷
۵۵	۱۸۵۶	۱۹۳۵/۵	۷۹/۵	۴/۳
۶۰	۱۹۴۷	۱۸۶۹/۸	-۷۷/۲	-۴/۰
۶۵	۲۰۸۳	۲۰۴۷/۸	-۳۵/۲	-۱/۷
۷۰	۲۱۷۶	۲۰۸۹/۵	-۸۶/۵	-۴/۰



شکل۴-۱۵- رسم مقادیر پیش‌بینی شده توسط روش ANN بر حسب مقادیر تجربی برای سری ارزیابی

جدول ۴-۲۱- مقایسه مقادیر مورد انتظار و پیش‌بینی شده سری تست

شماره ترکیب	مقدار تجربی	مقدار محاسبه شده	خطای مطلق	درصد خطای نسبی
۴	۱۰۹۲	۱۱۶۲/۹	۷۰/۹	۶/۵
۹	۱۱۶۳	۱۱۳۹/۱	-۲۳/۹	-۲/۱
۱۴	۱۲۴۷	۱۱۷۹	-۶۸	-۵/۴
۱۹	۱۳۷۳	۱۳۷۳/۳	۰/۳	۰/۰
۲۴	۱۵۰۷	۱۵۱۹/۷	۱۲/۷	۰/۸
۲۹	۱۵۸۵	۱۶۵۱/۱	۶۶/۱	۴/۲
۳۴	۱۶۴۶	۱۶۱۷/۴	-۲۸/۶	-۱/۷
۳۹	۱۶۸۴	۱۵۲۷/۸	-۱۵۶/۲	-۹/۳
۴۴	۱۷۱۶	۱۶۱۹/۸	-۹۶/۲	-۵/۶
۴۹	۱۷۷۹	۱۷۳۴/۵	-۴۴/۵	-۲/۵
۵۴	۱۸۱۲	۱۶۷۸/۵	-۱۲۴/۵	-۶/۹
۵۹	۱۹۲۴	۱۶۸۳/۱	-۲۴۰/۹	-۱۲/۵
۶۴	۱۹۹۷	۱۹۸۳/۷	-۱۳/۳	-۰/۷
۶۹	۲۱۶۵	۲۱۵۳/۸	-۱۱/۲	-۰/۵



شکل۴-۱۶- رسم مقادیر پیش‌بینی شده توسط روش ANN بر حسب مقادیر تجربی برای سری تست

۴-۴-۴- ارزیابی مدل به دست آمده از شبکه

۴-۴-۴-۱- ارزیابی شبکه به روش حذف مرحله‌ای تک تک داده‌ها

برای ارزیابی مل، علاوه بر به‌کارگیری سری تست، روش حذف مرحله‌ای تک تک داده‌ها نیز مورد استفاده قرار گرفت و نتایج آن با نتایج محاسبه شده توسط مدل غیرخطی برای سری داده‌ها مقایسه گردید (جدول ۴-۲۲ و شکل ۴-۱۷). تقارن نسبی داده‌ها حول خط صفر نشان‌دهنده عدم وجود خطای معین می‌باشد.

جدول۴-۲۲- ارزیابی مدل غیر خطی به روش حذف مرحله‌ای تک تک داده‌ها

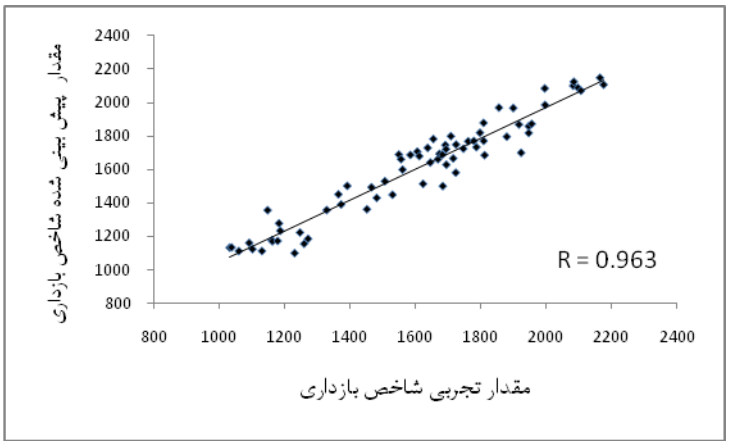
شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	تفاضل	درصد خطای نسبی
۱	۱۰۳۲	۱۱۲۹/۹	۹۷/۹	۹/۵
۲	۱۰۳۸	۱۱۳۰/۸	۹۲/۸	۸/۹
۳	۱۰۶۰	۱۱۰۹/۱	۴۹/۱	۴/۶
۴	۱۰۹۲	۱۱۵۷/۸	۶۵/۸	۶/۰
۵	۱۱۰۲	۱۱۲۰/۸	۱۸/۸	۱/۷
۶	۱۱۳۱	۱۱۰۹/۶	-۲۱/۴	-۱/۹
۷	۱۱۴۸	۱۳۵۳/۷	۲۰۵/۷	۱۷/۹
۸	۱۱۶۱	۱۱۷۴/۳	۱۳/۳	۱/۱
۹	۱۱۶۳	۱۱۶۷/۳	۴/۳	۰/۴
۱۰	۱۱۷۹	۱۱۶۹/۶	-۹/۴	-۰/۸
۱۱	۱۱۸۳	۱۲۷۴/۸	۹۱/۸	۷/۸
۱۲	۱۱۸۷	۱۲۳۰/۵	۴۳/۵	۳/۷
۱۳	۱۲۳۱	۱۰۹۶/۷	-۱۳۴/۳	-۱۰/۹
۱۴	۱۲۴۷	۱۲۲۰/۱	-۲۶/۹	-۲/۲
۱۵	۱۲۶۰	۱۱۵۲/۹	-۱۰۷/۱	-۸/۵
۱۶	۱۲۷۲	۱۱۸۲/۱	-۸۹/۹	-۷/۱
۱۷	۱۳۲۹	۱۳۵۴/۲	-۲۵/۲	-۱/۹
۱۸	۱۳۶۵	۱۴۴۸/۵	۸۳/۵	۶/۱
۱۹	۱۳۷۳	۱۳۸۸/۲	۱۵/۲	۱/۱
۲۰	۱۳۹۲	۱۴۹۹/۱	۱۰۷/۱	۷/۷
۲۱	۱۴۵۲	۱۳۵۹/۲	-۹۲/۸	-۶/۴
۲۲	۱۴۶۶	۱۴۹۱/۰	۲۵/۰	۱/۷

ادامه جدول۴-۲۲

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	تفاضل	درصد خطای نسبی
۲۳	۱۴۸۲	۱۴۲۷/۵	-۵۴/۵	-۳/۷
۲۴	۱۵۰۷	۱۵۲۷/۲	۲۰/۲	۱/۳
۲۵	۱۵۳۰	۱۴۴۶/۲	-۸۳/۸	-۵/۵
۲۶	۱۵۵۰	۱۶۸۶/۷	۱۳۶/۷	۸/۸
۲۷	۱۵۵۶	۱۶۵۹	۱۰۳/۰	۶/۶
۲۸	۱۵۶۱	۱۵۹۶/۵	۳۵/۵	۲/۳
۲۹	۱۵۸۵	۱۶۸۴/۳	۹۹/۳	۶/۳
۳۰	۱۶۰۶	۱۷۰۴/۱	۹۸/۱	۶/۱
۳۱	۱۶۱۲	۱۶۷۷/۲	۶۵/۲	۴/۰
۳۲	۱۶۲۴	۱۵۱۱/۴	-۱۱۲/۶	-۶/۹
۳۳	۱۶۳۸	۱۷۲۶/۶	۸۸/۶	۵/۴
۳۴	۱۶۴۶	۱۶۳۹/۱	-۶/۹	-۰/۴
۳۵	۱۶۵۵	۱۷۸۰/۰	۱۲۴/۰	۷/۶
۳۶	۱۶۶۹	۱۶۵۸/۳	-۱۰/۷	-۰/۶
۳۷	۱۶۷۴	۱۶۹۱/۳	۱۷/۳	۱/۰
۳۸	۱۶۸۳	۱۶۸۶/۷	۳/۷	۰/۲
۳۹	۱۶۸۴	۱۴۹۷/۹	-۱۸۶/۱	-۱۱/۰
۴۰	۱۶۹۲	۱۷۴۳/۴	۵۱/۴	۳/۰
۴۱	۱۶۹۴	۱۷۱۹/۲	۲۵/۲	۱/۵
۴۲	۱۶۹۵	۱۶۲۶/۱	-۶۸/۹	-۴/۱
۴۳	۱۷۰۹	۱۷۹۶/۹	۸۷/۹	۵/۱
۴۴	۱۷۱۶	۱۶۶۳/۹	۵۲/۱	-۳/۰
۴۵	۱۷۲۴	۱۵۷۸/۶	-۱۴۵/۴	-۸/۴
۴۶	۱۷۲۵	۱۷۴۶/۵	۲۱/۵	۱/۲
۴۷	۱۷۴۷	۱۷۲۱/۹	-۲۵/۱	-۱/۴
۴۸	۱۷۶۲	۱۷۶۶/۰	۴/۰	۰/۲
۴۹	۱۷۷۹	۱۷۶۸/۲	-۱۰/۸	-۰/۶
۵۰	۱۷۸۷	۱۷۳۱/۱	-۵۵/۹	-۳/۱
۵۱	۱۷۹۸	۱۸۱۷/۱	۱۹/۱	۱/۱
۵۲	۱۸۰۹	۱۷۶۸/۴	-۴۰/۶	-۲/۲
۵۳	۱۸۰۹	۱۸۷۶/۲	۶۷/۲	۳/۷
۵۴	۱۸۱۲	۱۶۸۲/۹	-۱۲۹/۱	-۷/۱
۵۵	۱۸۵۶	۱۹۶۷/۶	۱۱۱/۶	۶/۰
۵۶	۱۸۸۰	۱۷۹۳/۷	-۸۶/۳	-۴/۶

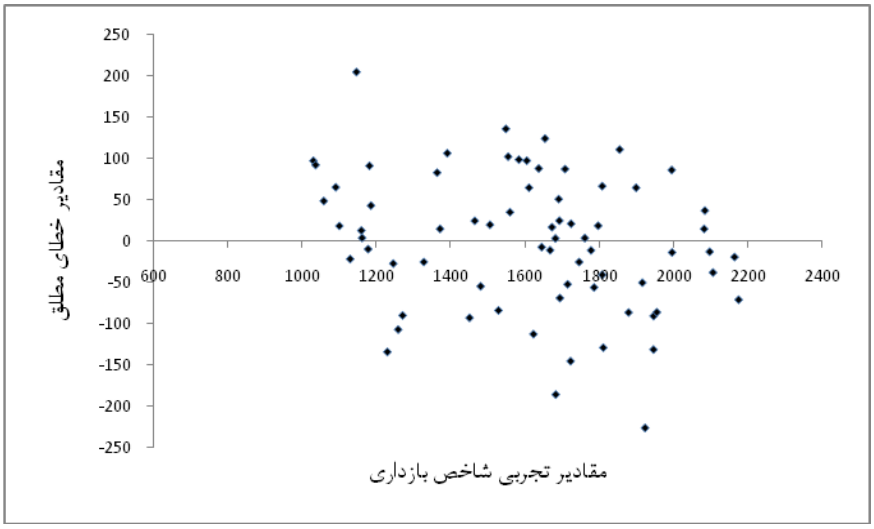
ادامه جدول۴-۲۲

شماره ترکیب	مقدار واقعی	مقدار پیش‌بینی شده	خطای مطق	درصد خطای نسبی
۵۷	۱۹۰۰	۱۹۶۵/۲	۶۵/۲	۳/۴
۵۸	۱۹۱۷	۱۸۶۶/۹	-۵۰/۱	-۲/۶
۵۹	۱۹۲۴	۱۶۹۷/۵	-۲۲۶/۵	-۱۱/۸
۶۰	۱۹۴۷	۱۸۵۶/۳	-۹۰/۷	-۴/۶
۶۱	۱۹۴۷	۱۸۱۵/۷	-۱۳۱/۳	-۶/۷
۶۲	۱۹۵۶	۱۸۶۹/۹	-۸۶/۱	-۴/۴
۶۳	۱۹۹۶	۲۰۸۲/۷	۸۶/۷	۴/۳
۶۴	۱۹۹۷	۱۹۸۳/۶	-۱۳/۴	-۰/۷
۶۵	۲۰۸۳	۲۰۹۸/۲	۱۵/۲	۰/۷
۶۶	۲۰۸۵	۲۱۲۲/۵	۳۷/۵	۱/۸
۶۷	۲۰۹۸	۲۰۸۵/۵	-۱۲/۵	-۰/۶
۶۸	۲۱۰۷	۲۰۶۹/۲	-۳۷/۸	-۱/۸
۶۹	۲۱۶۵	۲۱۴۶/۱	-۱۹/۰	-۰/۹
۷۰	۲۱۷۶	۲۱۰۵/۲	-۷۰/۸	-۳/۲



شکل ۴-۱۷- رسم مقادیر پیش‌بینی شده در ارزیابی مدل ANN بر حسب مقادیر تجربی برای سری داده‌ها

همچنین نمودار مقادیر باقیمانده محاسبه شده در ارزیابی مدل MLR بر حسب مقادیر تجربی شاخص بازداری نیز در شکل ۴-۱۸ رسم شد. تقارن نسبی داده‌ها حول خط صفر نشان‌دهنده عدم وجود خطای معین می‌باشد.



شکل۴-۱۸- رسم مقادیر خطای مطلق محاسبه شده در ارزیابی مدل ANN در مقابل مقادیر تجربی شاخص بازداری

۴-۴-۴-۲- ارزیابی شبکه به روش Y- تصادفی

در این روش، مقادیر پاسخ به صورت تصادفی در محدوده پاسخ‌ها تغییر داده شد. سپس همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از شاخص R^2 مورد بررسی قرار گرفت. این کار ده بار تکرار شد. در جدول ۴-۲۳ مقادیر R^2 برای سری ارزیابی و تست آورده شده است. مقادیر پایین R^2 نشان‌دهنده عدم وجود همبستگی شانسی در مدل به دست آمده از شبکه هستند.

جدول۴-۲۳- مقادیر R^2 برای سری ارزیابی و تست در روش Y-Randomization

شماره تکرار آزمون	R^2 برای سری ارزیابی	R^2 برای سری تست
۱	۰/۰۰	۰/۰۰
۲	۰/۰	۰/۰۰
۳	۰/۰۲	۰/۰۱
۴	۰/۰۰	۰/۱۱
۵	۰/۰۴	۰/۰۰
۶	۰/۱۹	۰/۰۰
۷	۰/۰۰	۰/۱۵
۸	۰/۰۶	۰/۰۲
۹	۰/۲۴	۰/۰۵
۱۰	۰/۰۰	۰/۰۰

۴-۵- بررسی نتایج و نتیجه‌گیری

در ادامه به بررسی و مقایسه نتایج به دست آمده در این فصل می‌پردازیم.

۴-۵-۱- مقایسه روش های رگرسیون خطی چندگانه و شبکه عصبی مصنوعی

جدول ۴-۲۴، به طور مقایسه‌ای پارامترهای آماری حاصل از دو روش MLR و ANN را نشان می‌دهد.

جدول ۴-۲۴- مقایسه برخی شاخص‌های آماری دو روش ANN و MLR

	R_t	R_{val}	R_{tr}	F_t	F_{val}	F_{tr}	SE_t	SE_{val}	SE_{tr}	R_{cv}
روش MLR	۰/۹۴۹	۰/۹۷	۰/۹۵۶	۹۹/۴۳۲	۱۷۷/۷۸۸	۴۷۷/۴۹۴	۹۴/۷۰۳	۸۵/۱۸۲	۸۵/۲۴۲	۰/۹۶۴
روش ANN	۰/۹۶۴	۰/۹۷	۰/۹۸	۱۵۸/۱۹۵	۱۹۲/۷۸۳	۹۹۴/۱۹۰	۸۷/۷۱۵	۸۱/۷۱۸	۶۱/۹۶۴	۰/۹۶۳

با توجه به نتایج این جدول، F و ضرایب همبستگی بزرگتر و خطای استاندارد کوچکتر سری آموزش، ارزیابی و تست برای شبکه عصبی مصنوعی، برتری این روش بر روش رگرسیون خطی چندگانه در پیش‌بینی شاخص بازداری را نشان می‌دهد.

۴-۵-۲- بررسی ارتباط توصیف‌کننده‌های ساختاری وارد شده در مدل با شاخص بازداری ترکیبات روغن‌های ضروری

در این قسمت با توجه به توصیف‌کننده‌های وارد شده در مدل‌های MLR و ANN یک بررسی اجمالی روی اثرات مختلف موجود در شاخص بازداری ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده شامل ۹ توصیف کننده است که هرکدام بیانگر خصوصیات متفاوتی از ملکول‌های مورد بررسی است.

۴-۵-۲-۱- توصیف‌کننده Ss

این توصیف کننده از گروه توصیف کننده‌های ساختاری می‌باشد. همانطور که در بخش ۳-۴-۲-۱ گفته شد، حالت الکتروتوپولوژیکی را می‌توان به صورت معیاری از احتمال برهم‌کنش با سایر ملکول‌ها نیز تفسیر کرد. به همین دلیل انتظار داریم افزایش مقادیر این توصیف کننده باعث افزایش برهم‌کنش ملکول با فاز ساکن و افزایش شاخص بازداری گردد. اثر متوسط مثبت برای این توصیف کننده (۰/۳۵۱۷۴) این فرضیه را تأیید می‌کند[۳۹].

۴-۵-۲-۲- توصیف‌کننده MATS2e

این توصیف‌کننده جزء توصیف کننده های خودهمبستگی است که قبلاً در بخش ۳-۴-۲-۵ به آنها اشاره شد. این توصیف کننده با الکترونگاتیویته ساندرسون وزن‌دار شده است. تأثیر این توصیف کننده روی شاخص بازداری منفی است.

۴-۵-۲-۳- توصیف‌کننده‌های BIC5 و IC2

این توصیف‌کننده‌ها جزء شاخص‌های اطلاعاتی هستند و دربردارنده اطلاعاتی راجع به ماهیت شیمیایی، پیوند اتم‌ها در فضا، توپولوژی مولکولی و تقارن می‌باشند.

اثر بزرگ و مثبت توصیف‌کننده BIC5 روی شاخص بازداری نشان دهنده تأثیر پیوندها و نحوه قرار گرفتن اتم‌ها در فضای مولکولی، روی مکانیسم بازداری می‌باشد. تأثیر توصیف‌کننده IC2 روی شاخص بازداری منفی است.

۴-۵-۲-۴- توصیف‌کننده nOH

این توصیف‌کننده تأثیر تعداد گروه‌های عاملی OH را روی شاخص بازداری نشان می‌دهد. وجود گروه OH تأثیر مثبتی روی افزایش شاخص بازداری دارد، اما به دلیل کم بودن تعداد مولکول‌های دارای گروه OH این تأثیر کمتر مشهود است.

۴-۵-۲-۵- توصیف‌کننده‌های H2m و R8e

این توصیف‌کننده‌ها از گروه توصیف‌کننده‌های GETAWAY[۱] می‌باشد. دو نوع از توصیف‌کننده‌های GETAWAY طراحی شده‌اند، یک گروه از این توصیف‌کننده‌ها به نام H-GETAWAY از یک نمایش جدید از ساختار مولکولی به نام ماتریس تأثیر مولکولی[۲] که با نماد H نمایش داده می‌شود به دست می‌آیند. این توصیف‌کننده‌ها به الکترونگاتیویته، اندازه اتمی و محل قرار گرفتن اتم‌ها در مولکول بستگی دارند. هرچه الکترونگاتیویته، اندازه اتمی و فاصله بین اتم و مرکز مولکول بیشتر باشد، مقدار این توصیف‌کننده بیشتر است. گروه دیگر توصیف‌کننده‌های R-GETAWAY هستند که از ماتریس اثر/ فاصله (R) گرفته شده‌اند و در آن عناصر ماتریس تأثیر مولکولی با عناصر ماتریس فضایی[۳] ترکیب شده‌اند[۴۵].

۴-۵-۲-۶- توصیف‌کننده RDF050m

این توصیف‌کننده جزء توصیف‌کننده‌های گروه RDF یا تابع توزیع شعاعی[۴] مربوط به یک دسته از اتم‌ها و معادل توزیع احتمال یافتن یک اتم در فضای کروی به شعاع R است. معادله ۴-۲ نحوه محاسبه این توصیف‌کننده‌ها را نشان می‌دهد.

$$g(R) = f.\sum_{i}^{N-1}\sum_{j>i}^{N} A_i A_j e^{-\beta.(R-r_{ij})^2}$$ معادله ۴-۲

که f یک فاکتور مقیاس و N تعداد اتم‌های مولکول است. همچنین، r_{ij} فاصله بین دو اتم i و j و A یک ویژگی اتمی است. β یک فاکتور تسهیل کننده است که توزیع احتمال فاصله بین اتمی را مشخص می‌کند و می‌توان آن را به فاکتور دما برای تعریف انرژی جنبشی اتمی تعبیر کرد. $g(R)$ در نقاط گسسته‌ای با فواصل معین حساب

۱.Geometry,Topology, and Atom Weights Assembly

۲.Molecular Influence Matrix

۳.Geometry Matrix

۴.Radial Distribution Function

می‌شود. این توصیف کننده‌ها ساختار سه بعدی مولکول را توصیف می‌کنند. تأثیر این توصیف کننده روی شاخص بازداری منفی است[۴۶].

۴-۵-۲-۷- توصیف‌کننده Mor11m

این توصیف کننده جزء توصیف کننده‌های 3D-MoRSE است که قبلاً در بخش۳-۴-۳-۲ معرفی شدند. این توصیف کننده با جرم اتمی (m) وزن دار شده است، و اثر آن روی شاخص بازداری کوچک ومنفی (۰/۰۱۶۵۶۷-) است.

۴-۶- آینده نگری

محققان می‌توانند تأثیر دسته خاصی از توصیف کننده‌ها را به طور ویژه بر روی مکانیسم بازداری مورد بررسی قرار دهند. به این ترتیب می‌توان تأثیر یک پارامتر خاص را بر روی تعداد بیشتری از داده‌ها تعمیم داد.

می‌توان از نرم‌افزارهایی با قابلیت‌های بیشتر از HypeChem و همچنین روش‌های دیگری برای انتخاب متغیرها، مانند روش الگوریتم طرح‌ریزی پی‌درپی[۱] بهره گرفت.

۱.Successive Projection Algorithm (SPA)

منابع:

[۱].Brereton R.G. (1990), "*Chemometrics*", Ellis Horwood, Chichester, pp 308

[۲].Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 59

[۳].Brandrup J. Immergut E.H., Grulke E. A. (1999), *Polymer Handbook",* John Wiley, New York pp*VII/675*

[۴].Hansen C.M (2004) "Fifty years with solubility parameters - past and future", Progress *in Organic Coatings*, 51, pp, 77-84

[۵].Burt s. (2004), "Essential oils: their antibacterial properties and potential application in foods- a review" *International Journal of Food Microbiology*, 94, pp 223-253

[۶].Politeo O. Juki M. Milos M. (2006), "Chemical Composition and Antioxidant Activity of Essential Oils of Twelve Spice Plants", *Croatica Chemica Acta***,** 79, 4, pp 545-552

[۷].رووד، دین، (۱۳۸۳)، "*روش‌های کروماتوگرافی: راهنمای عملی استفاده، نگهداری و رفع عیب از دستگاه‌های کروماتوگرافی گازی با ستون موئینه* " امامی س. رحیمی ح.، چاپ اول، انتشارات نورپردازان تهران، ص ۵۲.

[۸]Tantishaiyakul V., Worakul N., Wongpoowarak W. (2006), "Prediction of solubility parameters using partial least square regression" , *International Journal of Pharmaceutics*, 325, pp 8–14

[۹]. F.Gharagheizi, 2008,''QSPR Studies for Solubility Parameter by Means of Genetic Algorithm-Based Multivariate Linear Regression and Generalized Regression Neural Network '', QSAR Comb. Sci, 27 ,pp 165-170

[۱۰].Stefanis E., Panayiotou C., (2008), *Int J Thermophys*, 29, pp 568–58

[۱۱].Liao L., Mei H., Li J., Li Z., (2008), "Estimation and prediction on retention times of components from essential oil of Paulownia tomentosa flowers by molecular electronegativity-distance vector (MEDV*)", J Mol Struct,* 850, pp 1–8

[۱۲].Riahi S. Ganjali M.R., Pourbasheer S., Norouzi P.,(2008), "QSRR Study of GC Retention Indices of Essential Oil Compounds by Multiple LinearRegression with a Genetic Algorithm", Chromatographia, 67, pp 917–922

[۱۳].Riahi S. Ganjali M.R., Pourbasheer S., Norouzi P., (2009), "Investigation of different linear and nonlinear chemometric methods for modeling of retention index of essential oil components: Concerns to support vector machine" *Journal of Hazardous Materials* ,166 , pp 853–859

[۱۴].Otto M., (2007), "Chemometrics", *WILEY-VCH Verlag GmbH & Co. KGaA*, Weinheim, pp 1-11

[۱۵].Howery D.G., Hirsch R.F., (1983), "Chemometrics in the chemistry curriculum", *J. Chem. Ed***.** 60, pp 656-659

[۱۶].Wold S. "chemometrix: what do we mean with it, and what do we want from it?" And Brown, S. D., "Has the 'Chemometricsrevolution' ended? Some views on the past, present and future of Chemometrics", *papers on chemometrics: Philosiphy, History and Directions of InCINC'94*. http://WWW.emsl.pnl.gov:2080/docs/hncinic/homepage.html

[۱۷].Moosavi- Movahedi A.A. , Gharanfoli M. , Nazeri K. , Shamsipur M. , Chamani J. , Hemmateenejad B. , Alavi M. , Shokrollahi A. , Habib- Rezaei M. , Sorrenson C. , Sheibani N. , (2005), *Colloids Surf.* , 43, pp 150-157

[۱۸].Hemmatinejad B. (2005) "Chemometrics in Iran" *Chemometrics and Intelligent Laboratory Systems*, 81, pp 202-208

[۱۹].Katrrritsky A.R. , Karelson M. , Lobanov V,S , (1997), "QSPR as a mean of predicting and understanding chemical and physical properties in terms of structure" *Pure & Appl. Chem.* Vol.69 No.2 pp 245-248

[۲۰].Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 98.

[۲۱].Platt J.R. (1947). "Influence of Neighbor Bonds on Additive Bond Properties in Paraffins". *J.Chem.Phys*., 15, pp 419-420.

[۲۲].Wiener H. (1947c). "Structural Determination of Paraffin Boiling Points". *J.Am.Chem.Soc*., 69, pp 17-20.

[۲۳]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 307

[۲۴]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 304

[۲۵]. Pitts W., MacCullouch S., (1947), Bull. Math. Biophysic. 9, pp 127

[۲۶].Karelson Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 368-371

[۲۷].Karelson M., (2000), *"Molecular Descriptors in QSAR/QSPR"*, Wiley-VCH ,New York

[۲۸]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 461

[۲۹].Hyperchem professional Release 7 for Windows, Molecular Modeling System, Hypercube, Inc., 2002. <http://www.hyper.com>.

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

[۳۱]. SPSS 14.0 for Windows Evaluation version, Release 14.0.0, SPSS Inc., 2005.

[۳۲]. MATLAB, (The Language of Technical Computing). Version 7.5.0.342(R2007b), The Math Works Inc., 2007.

[۳۳]. جنی ر. بینز و. "هوش مصنوعی از الف تای"()، بیگدلی قمی س. محسنی م. بدیع ک.، دفتر تحقیقات یاسین، ص

[۳۴].منهاج م."مبانی شبکه‌های عصبی"، ()،چاپ ، مرکز نشر پروفسور حسابی، ص

[۳۵].کیا م. (۱۳۸۷)،"شبکه‌های عصبی در MATLAB "، چاپ اول، انتشارات کیان رایانه، ص ۲۹-۵۱

[۳۶].جنی ر. بینز و. "هوش مصنوعی از الف تای"()، بیگدلی قمی س. محسنی م. بدیع ک.، دفتر تحقیقات یاسین، ص

[۳۷].کیا م. (۱۳۸۷)،"شبکه‌های عصبی در MATLAB "، چاپ اول، انتشارات کیان رایانه، ص ۷۵-۹۳

[۳۸]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 159-163

[۳۹]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 492-497

[۴۰]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 513

[۴۱]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 84-90

[۴۲]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 17-20

[۴۳].Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 513

[۴۴].Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 84-90

[۴۵].González M.P., Terán C. Teijeira M., González-Moa M.J. (2005), "GETAWAY descriptors to predicting A2A adenosine receptors agonists", *European Journal of Medicinal Chemistry*, 40, pp 1080–1086

[۴۶]. Todeschini R. and Consonni V. (2000), "*Handbook of Molecular Descriptors*", John Wiley, New York, pp 366-367

Abstract:

In this project, a quantitative structure – property relationship (QSPR) study was conducted on the solubility parameters of 70 polymer solvents in the first part and on the retention indices (RI) of 70 essential oil compounds in second part. The solubility parameter (δ) is an intrinsic physicochemical property of a substance. It provides an easy numerical method of fast prediction the basic properties of materials. One of the essential functions of the solubility parameter is for evaluating the possibility of mixing between substances.

Plant essential oils and their extracts have been extensively employed in folkmedicine, for flavoring food, and in thefragrance and pharmaceutical industries The stepwise regression method was used as descriptor selection. Tow linear (Multiple Linear Regression; MLR) and nonlinear (Artificial Neural Network; ANN) methods were used to constructions of odels. The models were validated using external set, as well as leave one out and Y-Randomization techniques. Both linear and nonlinear methods have good prediction ability, although ANN model has more accurate results. In the first section of this study, the correlation coefficient (R) of the test set obtained by MLR and ANN models were 0.968 and 0.975 respectively. In the second part, the correlation coeficient of the test set obtained by MLR and ANN models were 0.949 and 0.964 respectively.

Keywords: quantitative Structure – Property Relationship (QSPR), Multiple Linear Regression (MLR), Artificial Neural Network (ANN), solubility parameter, Retention Indice (RI)

Shahrood University of Technology

Faculty of Chemistry

Prediction of solubility parameters of some organic solvents using linear and nonlinear QSPR methods

Hengameh Salimi

Superviser:

Dr. N. Goodarzi

Advisor:

Dr. Gh. Bagherian dehaghi

Date: July – 2010