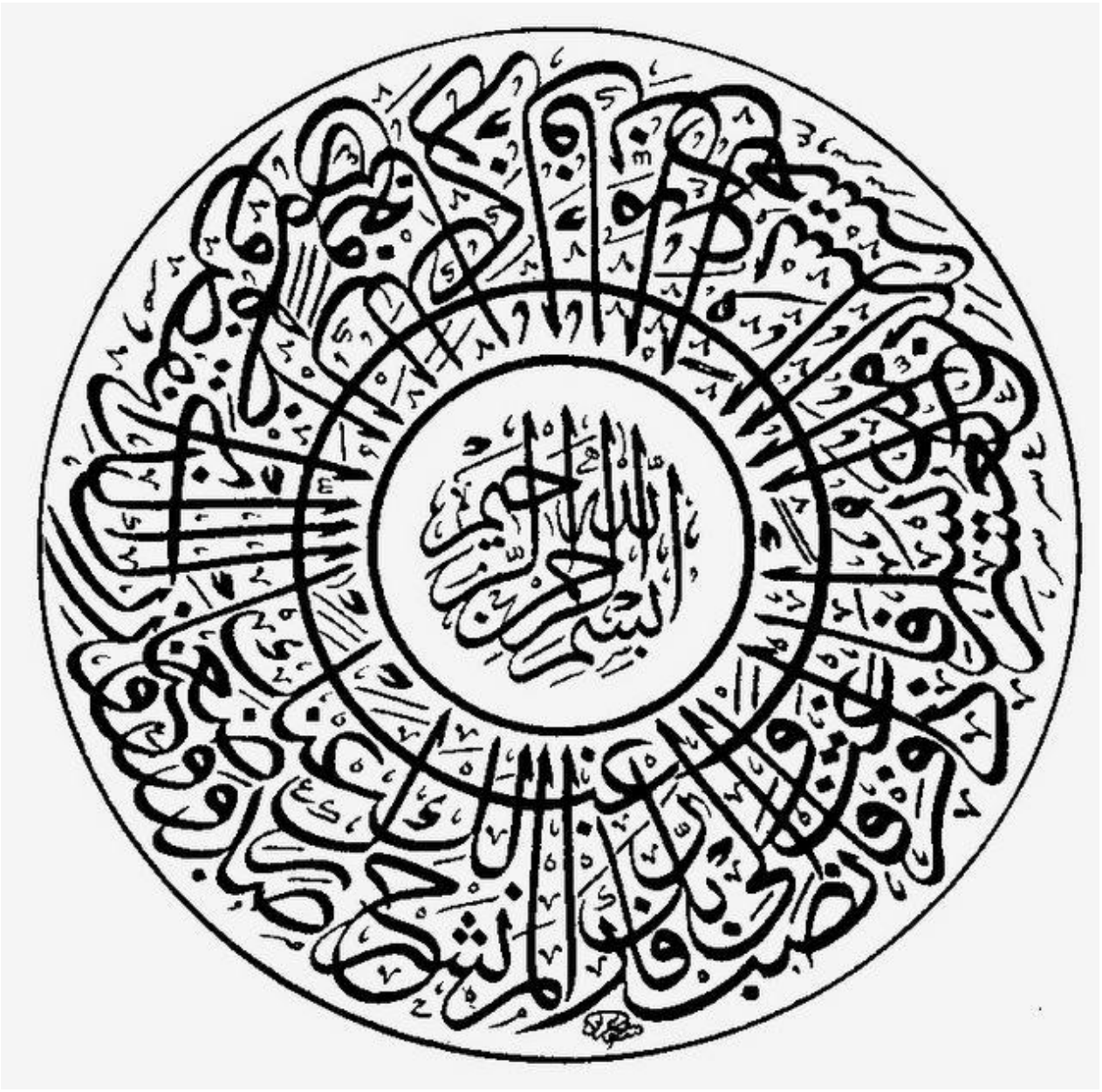


دانشکده: علوم پایه

گروه: شیمی

مطالعه کمی ساختار-فعالیت برخی از ترکیبات سولفونانیلید به عنوان گروهی جدید از داروهای ضد سرطان و فعالیت ضد HIV برخی از ترکیبات جدید

دانشجو: **مهدی محمدرضائی**

استاد راهنما:

دکتر منصور عرب چم جنگلی

استاد مشاور:

دکتر زهرا کلانتر کهدمی

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

دی ماه ۱۳۸۹

به پاس تعبیر عظیم و انسانی شان از کلمه ایثار و از خودگذشتگی

به پاس قلب های بزرگشان که فریادرس است و سرگردانی و ترس در پناهشان به شجاعت می گراید

به پاس عاطفه سرشار و گرمای امیدبخش وجودشان که در این سردترین روزگاران بهترین پشتیبان است

و به پاس محبت های بی دریغشان که هرگز فروکش نمی کند

این مجموعه را به پدر و مادر عزیزم تقدیم می کنم

به مصداق «من لم یشکر المخلوق لم یشکر الخالق» بسی شایسته است

که با درود فراوان خدمت پدر و مادر بسیار عزیز، دلسوز و فداکارم که پیوسته جرعه نوش جام تعیلم و تربیت، فضیلت و انسانیّت آنها بوده‌ام و همواره چراغ وجودشان روشنگر راه من در سختی ها و مشکلات بوده است، از ایشان تقدیر و تشکر نمایم.

از زحمات بی‌دریغ اساتید فرهیخته و بزرگوارم در دانشگاه صنعتی شاهرود بخصوص استاد ارجمندم جناب آقای دکتر عرب که با راهنمایی‌های برادرانه خود در پیشبرد این پروژه راهگشای اینجانب بودند و استاد گرامیم سرکار خانم دکتر کلانتر که با مشاوره‌های خردمندانه و پرمغزشان یاری‌گر من در این راه بوده و همیشه با روی گشاده پاسخگوی سوالات بنده بودند کمال تشکر و سپاسگزاری را داشته باشم.

همچنین از تمامی همکلاسی‌های عزیزم و بچه‌های دانشکده شیمی و نیز دوستان و هم اتاقی‌های گراممیم که لحظاتی سرشار از صفا و صمیمیت را در کنار خود برایم به یادگار گذاشتند و همیشه اینجانب را مورد لطف و محبت خود قرار داده و به من درس صداقت و مهرورزی آموختند بسیار سپاسگزار باشم.

در پایان از زحمات سرکار خانم هنگامه سلیمی که مرا در مراحل مختلف انجام این پروژه یاری نمودند و مهربانانه پاسخگوی پرسش‌های من بودند و تجربیات خود را در اختیار اینجانب قرار دادند کمال تشکر را دارم.

تعهدنامه

اینجانب **مهدی محمدرضائی** دانشجوی دوره کارشناسی ارشد رشته **شیمی تجزیه** دانشکده **شیمی** دانشگاه **صنعتی شاهرود** نویسنده پایان‌نامه: **مطالعه کمی ساختار- فعالیت برخی از ترکیبات سولفونانیلید به عنوان گروهی جدید از داروهای ضد سرطان و فعالیت ضد HIV برخی از ترکیبات جدید** تحت راهنمایی **دکتر منصور عرب چم‌جنگلی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تا کنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود» و یا « **Shahrood University of Technology** » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجودات زنده (یا بافت‌های آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاقی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

*** متن این صفحه نیز باید در ابتدای نسخه‌های تکثیر شده پایان‌نامه وجود داشته باشد.**

چکیده:

تحقیقات اخیر نشان داده‌اند که استروژن نقش بسزایی در بیماری سرطان سینه دارد، بنابراین پژوهش‌هایی در رابطه با داروهایی که بتوانند به نحوی فعالیت این هورمون را در بدن محدود نمایند انجام شده است. این پژوهش‌ها به دو دسته تقسیم می‌شوند، اول پژوهش بر روی ترکیباتی که به طور مستقیم بر فعالیت استروژن در بدن اثر می‌گذارند و دوم تحقیق بر روی داروهایی که از طریق ایجاد محدودیت در تولید استروژن عمل می‌نمایند. یکی از این دسته ترکیبات که می‌توانند تولید استروژن را محدود نمایند، مشتقات سولفونانیلید هستند. در بررسی اول رابطه کمی ساختار-فعالیت ۲۱ آنالوگ سولفونانیلید با استفاده از رگرسیون خطی چندگانه و شبکه عصبی مصنوعی مورد مطالعه قرار گرفت. از میان تعداد زیاد توصیف‌کننده محاسبه شده فقط ۲۰ توصیف‌کننده توسط رگرسیون خطی مرحله‌ای انتخاب گردید که فقط ۶ تا از آنها به عنوان ورودی شبکه استفاده شدند. سری داده‌ها به طور تصادفی به دو سری آموزش و تست که به ترتیب شامل ۱۶ و ۵ ترکیب بودند تقسیم شد. مدل‌های رگرسیون خطی و شبکه عصبی با استفاده از روش ارزیابی تقاطعی ساخته شدند. توانایی پیش‌بینی مدل‌ها با استفاده از سری تست و روش رد مرحله‌ای تک‌تک مورد ارزیابی قرار گرفت و مقادیر MSE به ترتیب برابر با ۰/۰۲۲۷ و ۰/۰۱۳۲ برای MLR و ۰/۰۰۹۸ و ۰/۰۱۶۱ برای ANN به دست آمد. نتایج به دست آمده نشان دهنده قدرت بالای مدل شبکه عصبی مصنوعی در پیش‌بینی توان بازدارندگی ترکیبات مورد بررسی می‌باشند.

در بررسی دوم رابطه ساختار-فعالیت مشتقات ۵-اکسوپیرولیدین-۳-کربوکسامید مورد پژوهش قرار گرفت. این ترکیبات از طریق دخالت در برهمکنش میان CCR5 و HIV مانع از گسترش بیماری می‌شوند. مجموعه داده‌های مورد بررسی شامل ۱۰۴ ترکیب از مشتقات ۵-اکسوپیرولیدین-۳-کربوکسامید می‌باشد. در این تحقیق سری داده‌ها به طور تصادفی به سه سری آموزش، ارزیابی و تست تقسیم شد که هر یک به ترتیب شامل ۶۲، ۲۱ و ۲۱ ترکیب می‌باشند. مدل شبکه عصبی با استفاده از ۱۹ توصیف‌کننده با قدرت بالایی مقادیر pIC_{50} مربوط به هریک از ترکیبات را پیش‌بینی نمود. مقادیر mse برای سری تست و رد مرحله‌ای تک‌تک به ترتیب ۰/۱۲۱۰ و ۰/۱۴۰۵ می‌باشد.

واژه‌های کلیدی: مطالعه ساختار- فعالیت، شبکه عصبی مصنوعی، رگرسیون خطی چندگانه، سولفونانیلید.

نتایج به دست آمده از این پژوهش طی دو پوستر تحت عناوین

1. Quantitative Structure-Activity Relationships of some of Sulfonanilide derivatives as anti-breast cancer agents

2. Non-linear QSAR study of anti-HIV-1 activity of some recently synthesized pipredine-4-carboxamide derivatives

در هفدهمین سمینار شیمی تجزیه ایران در دانشگاه کاشان ارائه گردید.

فهرست مطالب

فهرست اشکال

فهرست جداول

فصل اول

مقدمه

۱-۱- کمومتریکس

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.

از نظر شیمی‌دانان فعالیت‌ها و خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده مولکول انجام گیرد، به آن ارتباط کمی ساختار – ویژگی (QSPR)[۱] می‌گویند مانند یافتن رابطه‌ای بین خواصی نظیر نقطه‌ی جوش و ساختار مولکولی و یا پیش بینی فاکتور بازداری ترکیبات آلی روی ستون های کروماتوگرافی. اما وقتی خواصی از نوع بیولوژیکی (مانند فعالیت دارویی) توصیف می‌شود، اشاره به ارتباط کمی ساختار– فعالیت (QSAR)[۲] خواهد داشت[۳]. در مطالعات QSAR سعی بر این است تا رابطه‌ی هماهنگ میان فعالیت‌های شیمیایی و فیزیکی با ویژگی‌های مولکولی پیدا شود، به گونه‌ای که بتوان این قواعد را برای ارزیابی فعالیت ترکیبات جدید به کار برد. در واقع نتایج این نوع مطالعات علاوه بر شفاف‌سازی نحوه‌ی ارتباط بین خواص مولکول‌ها و ویژگی‌های ساختمانی آنها، به پژوهشگران در پیش‌بینی رفتار مولکول‌های جدید بر اساس رفتار مولکول‌های مشابه کمک می‌کند[۴].

از جمله روش‌هایی که به منظور مطالعه ارتباط خطی ساختار فعالیت مورد استفاده قرار می‌گیرند

[۱] Quantitative Structure-Property Relationship (QSPR)

[۲] Quantitative Structure-Activity Relationship (QSAR)

می‌توان به روش‌های رگرسیون خطی چندگانه[۱]، رگرسیون اجزای اصلی[۲] و روش حداقل مربعات جزئی[۳] اشاره کرد. روش‌های دیگری مانند شبکه‌های عصبی مصنوعی[۴] نیز ارتباط غیرخطی میان ساختار و ویژگی‌های ترکیبات را مورد مطالعه قرار می‌دهند. در تحقیق حاضر مطالعه ارتباط کمّی ساختار- فعالیت ترکیبات موردنظر با استفاده از روش‌های خطی و غیرخطی انجام گرفته و نتایج با هم مقایسه شده‌اند.

۱-۲- ارتباط کمی ساختار – فعالیت

کورن هانش[۵] برای اولین بار در سال ۱۹۶۴ مفهوم رابطه‌ی کمی ساختار – فعالیت یا همان QSAR را مطرح کرد. او نشان داد که توصیف‌کننده‌های کمی[۶] مولکولی در ترکیبات دارویی با ویژگی‌های بیولوژیک این ترکیبات ارتباط نزدیکی دارند. او و دانشمندان پس از وی، از توصیف‌کننده‌های ساده‌ای برای توصیف ساختار ترکیبات استفاده کردند که توانایی توصیف ویژگی‌های الکترونی، هندسی، فضایی و هیدروفوبیک ترکیبات دارویی را داشتند. هر مطالعه QSAR شامل مراحل زیر می‌باشد:

- جمع‌آوری سری داده‌ها
- رسم ساختارهای مولکولی و بهینه‌سازی آنها
- استخراج توصیف‌کننده‌ها
- بدست آوردن بهترین توصیف‌کننده‌ها برای ورود به مدل
- مدل‌سازی

---

[۱] Multiple Linear Regression(MLR)
[۲] Principal Components Regression (PCR)
[۳] Partial Least Square (PLS)
[۴] ArtificialNeural Network (ANN)
[۵] Corren Hansch
[۶] Quantitative Descriptors

- ارزیابی مدل

۱-۲-۱- جمع‌آوری سری داده‌ها

اولین گام در توسعه‌ی QSAR و QSPR گردآوری مجموعه‌ای از ترکیبات است که کمیت‌های تجربی خاصی برای آنها بوسیله محققین تجربی گزارش شده‌است. این کار معمولا از طریق جستجو در مقالات، مجلات و نوشتارهای علمی انجام می‌پذیرد. از میان مولکول‌های گردآوری شده آنهایی که مقادیر بدست آمده برای آنها در شرایط عملی یکسانی بدست آمده است بعنوان سری داده‌ها انتخاب می‌شوند. هر چه تنوع و تعداد این ساختارها بیشتر بوده و بصورتی انتخاب شوند که از نظر ساختاری به مولکول مادر شباهت بیشتری داشته باشند مدل QSAR بدست آمده مطلوب‌تر و از قدرت پیش‌بینی بالاتری برخوردار است.

۱-۲-۲- رسم ساختارهای مولکولی و بهینه‌سازی آنها

کمومتریکس از روش‌های مفیدی که قادر به استخراج اطلاعات موجود در داده هستند، استفاده می‌کند[۵]. به این منظور از شیمی محاسباتی[۱] 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، روشAM1[۲] است که جزء روش‌های محاسباتی نیمه‌تجربی[۱] می‌باشد. در روش‌های نیمه‌تجربی، که در

[۱] Computational chemistry
[۲] Austin Methods

برنامه‌هایی نظیر HyperChem وارد شده‌اند، محاسبات بر اساس مکانیک کوانتومی صورت می‌گیرد. در این روش‌ها یک سری از مقادیر تجربی نیز وارد شده‌اند و فقط الکترون‌های لایه ظرفیت در محاسبه در نظر گرفته می‌شوند. روش‌های نیمه تجربی مختلفی مانند صرف نظر کردن کامل از انتگرال همپوشانی دیفرانسیلی[۲] (CNDO)، صرف‌نظر کردن میانی از همپوشانی دیفرانسیلی[۳] (INDO)، صرف‌نظر کردن از همپوشانی دو اتمی اصلاح شده[۴] (MNDO) ارائه شده است[۶]. روش AM1 همان روش MNDO اصلاح شده است که روی مولکول‌های قطبی نتایج خوبی می‌دهد ولی پیوند هیدروژنی را نمی‌تواند در نظر بگیرد. در این روش هسته و لایه‌های داخلی را به شکل یک هسته مرکزی[۵] در نظر می‌گیریم و محاسبات روی الکترون‌های ظرفیت انجام می‌گردد. سایر پارامترهای مورد نیاز در این روش از اطلاعات تجربی مانند اسپکتروسکوپی استخراج می‌گردند [۳و۷].

۱-۲-۳- استخراج 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. عدد پلات[۶][۸] و شاخص وینر[۷] [۹] 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

---

[۱] Semi-empirical methods
[۲] Complete neglected of differential overlap
[۳] Intermediate neglected of differential over lap
[۴] Modified neglected of diatomic overlap
[۵] Core
[۶] Platt Number
[۷] Wiener Index

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:[۱۰]

- ساده بودن
- مستقل بودن
- تفسیر ساختار مولکول (غنی بودن از نظر اطلاعات)
- عدم همبستگی با سایر توصیف‌کننده‌ها
- 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 و 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[۱۱]. نمایش مولکولی روشی است که با آن یک مولکول از طریق فرآیند قراردادی و قواعد اختصار به صورت نمادین ارائه می‌شود. در نوع دیگری از تقسیم‌بندی، توصیف‌کننده‌ها به دسته‌های صفربعدی، یک‌بعدی، دوبعدی و سه‌بعدی طبقه‌بندی می‌شوند که در اینجا اشاره‌ای مختصر به این توصیف‌کننده‌ها خواهد شد.

❖ 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،

[۱] Molecular representation

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، تعداد یک عنصر شیمیایی خاص(C, N, O, H, F, …)، تعداد گروه‌های عاملی خاص، تعداد پیوندهای ساده، دوگانه، سه گانه، آروماتیک و...، تعداد کل حلقه‌ها بر اساس تعداد اتم‌های آنها (حلقه های شش تایی، پنج تایی و ...) در مولکول، وزن مولکولی و متوسط وزن اتمی، مثال‌هایی از توصیف‌کننده‌های ساختاری می‌باشند.

❖ توصیف‌کننده‌های یک بعدی

نمایش فهرست ویژگی‌های زیرساختاری[1] 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 مورد 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.

❖ 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

[1] Substructure List Representation

توصیف‌کننده‌ها شامل زیرگروه‌های توپولوژیکی، شمارش‌های مسیرهای مولکولی[۱]، توصیف‌کننده‌های BCUT[۲]، ضرایب بار[۳] و توصیف‌کننده‌های خود همبستگی دو بعدی[۴] می‌شود.

❖ **توصیف‌کننده‌های سه بعدی**

نمایش سه بعدی، یک مولکول را معمولاً به عنوان یک شیء هندسی انعطاف ناپذیر در فضا می‌بیند و نه تنها یک نمایش از ماهیت و اتصال اتم‌ها، بلکه از صورتبندی فضائی مولکول نیز فراهم می‌کند که به آنها نمایش‌های هندسی[۵] گفته می‌شود. توصیف‌کننده‌های مولکولی مشتق شده از این نمایش، توصیف‌کننده‌های سه بعدی نامیده می‌شوند. زیرگروه‌های این دسته عبارتند از: شاخص‌های آروماتیکی[۶]، خصوصیات مولکولی راندیک[۷] و توصیف‌کننده‌ها 3D-MoRSE[۸].

۱-۲-۴- بدست آوردن بهترین توصیف‌کننده‌ها برای ورود به مدل

یکی از مهمترین مراحل QSAR تعیین و انتخاب توصیف‌کننده‌های مناسب می‌باشد. زیرا توصیف‌کننده‌های زائد کار برازش و مدل‌سازی را طولانی می‌کنند و تاثیری در بهبود نتایج نخواهند داشت. به این منظور لازم است در انتخاب توصیف‌کننده‌ها دقت لازم به عمل آید و توصیف‌کننده‌هایی انتخاب شوند که با پارامتر موردنظر ارتباط نزدیکی داشته و بتوانند تغییرات آن را به خوبی توجیه نمایند. یکی از روش‌های انتخاب توصیف‌کننده‌های مناسب روش برازش مرحله‌ای می‌باشد که با محاسبه‌ی ضرایب همبستگی مشخص می‌شوند. ضریب همبستگی میزان نزدیکی اطلاعات بدست

---

[۱] Molecular Walk Count

[۲] Burden-CAS-University of Texas eigenvalues

[۳] Charge indices

[۴] 2D autocorrelation

[۵] Geometrical Representation

[۶] Aromaticity indices

[۷] Randic molecular properties

[۸] 3-D MOlecular Representation of Structures based on Electron diffraction

آمده به خط رگرسیون را نشان می‌دهد. مقدار ضریب همبستگی بین صفر و یک تغییر می‌کند. صفر یعنی هیچ ارتباطی بین فعالیت و توصیف‌کننده وجود ندارد و یک یعنی ارتباط کامل است. پس توصیف‌کننده‌هایی با ضرایب همبستگی بزرگ در برازش منحنی برای ساختن مدل جهت پیش‌گویی خواص استفاده می‌شوند[۳]. در ادامه برخی از روش‌های خطی انتخاب توصیف‌کننده‌ها به طور اجمالی معرفی می‌شوند.

❖ 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.

❖ روش رو به جلو[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.

❖ روش رو به عقب[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.

---

[۱] Enter
[۲] Forward
[۳] Backward

این فرآیند زمانی متوقف خواهد شد که حذف متغیری که از شرط فرآیند پیروی کند باعث کاهش ارزش مدل گردد.

❖ روش مرحله‌ای[1]

این روش مشابه روش انتخاب جلو برنده است، ابتدا متغیری که بالاترین میزان همبستگی را با متغیر وابسته دارد وارد مدل می‌شود، ولی با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله مورد بررسی قرار می‌گیرند. برای متغیرهای وارد شده در مدل، آمارهF محاسبه می‌شود و متغیر با کوچک‌ترین آماره F قبل از ورود متغیر جدید حذف خواهد شد و مدل مناسب به دست می آید[۱۰].

۱-۲-۵- مدل‌سازی

مدل یک رابطه ریاضی است که بیانگر رابطه بین متغیر وابسته و متغیر(های) مستقل می‌باشد. روش‌های متفاوتی در مدل‌سازی به کار گرفته می‌شوند. که می‌توان آنها را در دو دسته کلی قرار داد:

الف) روش‌های آماری چند متغیره

این روش‌ها خطی می‌باشند و وقتی کارآیی مناسبی دارند که داده‌ها دارای رفتار خطی باشند. در این روش از روش‌های کالیبراسیون چند متغیره استفاده می‌شود. چندین تکنیک متفاوت برای برازش خطی چند متغیره وجود دارد که عبارتند از MLR، PLS، PCR و ... که در اینجا اشاره مختصری به روش MLR می‌شود.

[1] Stepwise

۱-۲-۶- مدل‌سازی به روش MLR

انتخاب نوع آنالیز چند متغیره تحت تأثیر نوع توصیف‌کننده‌های مورد استفاده می‌باشد. 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 در تابعی که برای مدل‌سازی مناسب است سهمی دارند. 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 بدست می‌آیند. یک مدل رگرسیون خطی بر اساس فرضیات زیر استوار است:

- مقادیر متغیر مستقل از قبل توسط محققین تعیین شده و در طول رگرسیون تغییر نخواهد کرد.
- خطای اندازه‌گیری متغیر مستقل ناچیز و قابل چشم‌پوشی است.
- به ازای هر مقدار متغیر مستقل زیرمجموعه‌ای از مقادیر وابسته با پراکندگی نرمال وجود دارد.
- واریانس در زیرمجموعه‌های متغیر وابسته یکسان است.
- میانگین‌های زیرمجموعه‌های متغیر وابسته روی یک خط واقعند. این فرض که فرض خطی بودن نیز نام دارد به شکل زیر بیان می‌شود:

$$\mu_{y|x} = \alpha + \beta x$$

که $\mu_{y|x}$ میانگین زیرمجموعه متغیر وابسته (y) برای مقدار خاصی از متغیر مستقل (x) است. همچنین α و β ضرایب رگرسیون هستند.

- مقادیر متغیر وابسته از نظر آماری مستقل‌اند.

فرض‌هایی که ذکر شد را می‌توان با این رابطه بیان کرد که همان مدل رگرسیون خطی است:

$$y = \alpha + \beta x + e$$

که y یک مقدار از متغیر وابسته، x یک مقدار از متغیر مستقل و e نشانگر خطاست.

روشی که غالبا برای بدست آوردن معادله‌ی خط بین متغیرهای مستقل و وابسته بکار می‌رود، به روش حداقل مربعات[1] موسوم است. اصولا برای رسم یک خط دانستن مقادیر α و β در معادله مربوط به آن کافی است. این مقادیر را می‌توان از دستگاه معادلات زیر که به معادلات نرمال معروف است بدست آورد:

$$\begin{cases} \sum y_i = n\alpha + \beta \sum x_i \\ \sum x_i y_i = \alpha \sum x_i + \beta \sum x_i^2 \end{cases} \quad (۳-۱)$$

n بیان‌گر تعداد مقادیر مشخص تجربی یا همان رکوردهاست. با جایگزین کردن مقادیر تجربی متغیرهای مستقل و وابسته در معادلات نرمال مقادیر α و β به دست می‌آیند. با استفاده از روابط زیر می‌توان مقادیر α **و** β **را** بصورت مستقیم هم به دست آورد:

$$\beta = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \quad (۴-۱)$$

$$\alpha = \frac{\sum y_i - \beta \sum x_i}{n} \quad (۵-۱)$$

خطی که با این روش به دست می‌آید دارای بهترین معادله برای مرتبط کردن متغیرهای مستقل و وابسته است؛ هرچند همه‌ی مشاهدات تجربی بر روی این خط واقع نمی‌شوند. ویژگی چنین خطی این است که مجموع مربعات انحراف عمودی نقاط مشاهده شده $(\bar{y}_i)$ از این خط، همواره کمتر از مجموع مربعات انحراف عمودی نقاط مشاهده شده $(\bar{y}_i)$ از هر خط مفروض دیگر است.

[1] Least Squares Method

پس از این مقدمه، به بررسی رگرسیون خطی با چند متغیر می‌پردازیم که مدل مورد استفاده ما در این تحقیق است. در این نوع رگرسیون فرض‌هایی مشابه آنچه در حالت قبل دیدیم، در نظر گرفته می‌شود. معادله‌ی این نوع رگرسیون به این صورت است:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i \qquad (۶-۱)$$

که y_i یکی از مقادیر زیرمجموعه‌ی متغیر وابسته است. مقادیر β ضراب متغیرهای مستقل‌اند. e_i هم یک مقدار تصادفی است که میانگین آن صفر و واریانس آن برابر واریانس زیرمجموعه‌های متغیر وابسته است. مقدار این متغیر بیانگر میزان خطای رابطه‌ی رگرسیون است.

مقادیر β را می‌توان با بسط روابط فوق از رابطه زیر محاسبه نمود:

$$\beta = (x' \times x)^{-1} \times x' \times y \qquad (۷-۱)$$

در این رابطه β ماتریسی ستونی است که اولین آرایه آن برابر با مقدار β_0 و سایر اعضای آن به ترتیب با مقادیر $\beta_1, \beta_2, \beta_3, \ldots, \beta_k$ در رابطه (۱-۶) برابر خواهند بود. x نیز ماتریسی است که تمامی اعضای ستون اول آن برابر یک می باشند و سایر ستون‌های آن را متغیرهای مستقل تشکیل می‌دهند و y نیز ماتریس ستونی شامل متغیرهای وابسته است.

برای ارزیابی معادله رگرسیون می‌توان از شاخص r^2 استفاده کرد که به صورت زیر محاسبه می‌گردد:

$$r^2 = \frac{var(y')}{var(y)} \qquad (۸-۱)$$

که در این رابطه $var(y')$ واریانس مقادیر بدست آمده از معادله رگرسیون و $var(y)$ واریانس مقادیر تجربی است. از این رو هرچه مقدار این شاخص بزرگتر باشد مدل صحیح‌تر خواهد بود[۱۳].

ب) روش‌های غیرخطی

جهت کم کردن انحراف بین مقادیر برازش شده و مقادیر صحیح بهتر است از یک برازش غیرخطی در مدل‌سازی استفاده شود. از میان روش‌های غیرخطی روش شبکه عصبی بطور گسترده در مطالعات QSAR و QSPR به کار گرفته شده است[۳].

۱-۳- شبکه‌های عصبی

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 اعمال 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 کاهش یابد، در زمان لازم برای اجرا صرفه‌جویی می‌شود. دوم اینکه برخی مسأله‌ها با شیوه‌های موازی بهتر حل می‌شوند. برای مثال بینایی فعالیتی است که در آن همه تصویر باید به یکباره پردازش شود، چون تفسیر قسمتی از تصویر به تفسیر قسمت‌های دیگر بستگی دارد[۱۴].

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

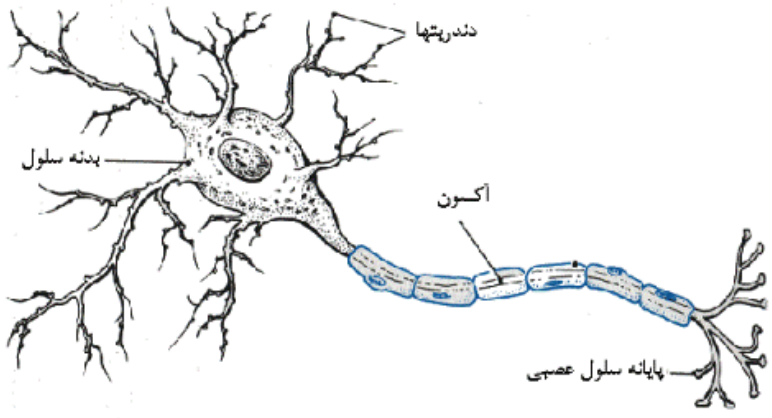
[۱] Neuron

۱-۳-۱- 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 کل اکسیژن بدن را مصرف می‌کند برای بسیاری از رفتارهای خودآگاه و ناخودآگاه همچون تفکر، تنفس، خواندن، دویدن و... استفاده می‌شود. مغز از ۱۰۰ تریلیون نرون که با بیش از $۱۰^{۱۶}$ ارتباط به هم متصل شده‌اند تشکیل شده است. نرون‌ها ساده‌ترین 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

دندریت‌ها به عنوان مناطق دریافت سیگنال‌های الکتریکی، شبکه‌هایی تشکیل یافته از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بیشمار می‌باشند. به همین علت آنها را شبکه‌های دریافت‌کننده‌ی درخت مانند نیز می‌نامند. دندریت‌ها سیگنال‌های الکتریکی را به هسته سلول منتقل می‌کنند. بدنه سلول، انرژی لازم را برای فعالیت نرون فراهم نموده و بر روی سیگنال‌های دریافتی عمل می‌کند.



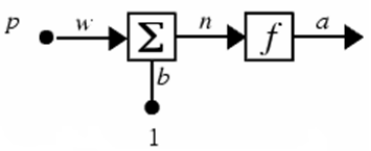
شکل ۱-۱- شمای سلول عصبی زیستی

[۱] Dendrite

آکسون‌ها[1] 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. محل تلاقی یک آکسون از یک سلول به دندریت‌های سلول دیگر را سیناپس می‌نامند. 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 [۱۵]. شکل ۱-۱ نشان‌دهنده ساختار کلی یک نرون می‌باشد.

۱-۳-۲- 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[2] نحوه ارتباط بین اجزای شبکه را تعیین نمود. معمولاً برای نشان دادن مدل ریاضی ساده‌ای از نرون شبکه عصبی مصنوعی از شکل زیر استفاده می‌شود:



شکل ۱-۲- مدل ریاضی نرون

[1] Axon
[2] Connection weight

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

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[۱۶].

۱-۳-۲-۱- توابع انتقال

سه تابع انتقال رایج در شبکه‌های عصبی، در این پایان‌نامه برای بهینه‌سازی شبکه به‌کار رفته است. این توابع عبارتند از:

❖ **تابع انتقال خطی (PURLIN)**

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 ۱-۳ 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 ۱-۳ ب).

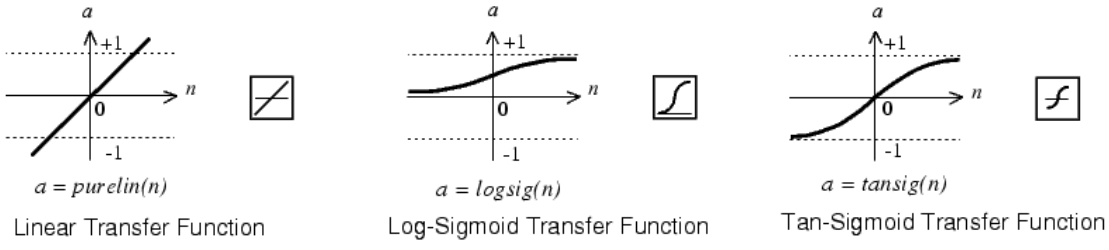
❖ **تابع انتقال تانژانت هایپربولیک (TANSIG)**

این تابع انتقال مقادیر ورودی را در محدوده ∞- و ∞+ دریافت کرده و خروجی بین ۱ و ۱- تولید می‌نماید (شکل ۱-۳ ج).

---

[۱] Transfer function

[۲] bias



الف ب ج

شکل ۱-۳- برخی از توابع انتقال مورد استفاده در شبکه‌های عصبی

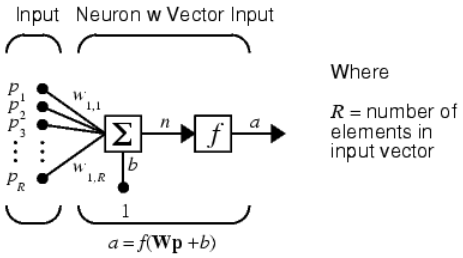
۱-۳-۳- معماری شبکه‌های عصبی[۱]

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

$$n = W_{1,1}P_1 + W_{1,2}P_2 + \cdots + W_{1,R}P_R + b \qquad (۹-۱)$$

در نهایت b مقدار بایاس است که به حاصلضرب ماتریسی w در p اضافه شده است. به ترکیب وزن‌ها، بایاس، تابع انتقال و عملیات ضرب و جمع انجام شده یک لایه از شبکه می‌گویند، شایان توجه است که بردار ورودی در بیشتر نوشتارها، یک لایه محسوب نمی‌شود. اگر از تابع انتقال خاصی استفاده شود، نماد آن تابع در کادر مربوط به تابع قرار می‌گیرد. شکل ۱-۴ نمایش خلاصه‌ای از یک نرون تک‌لایه را نشان می‌دهد.

[۱] Network architecture

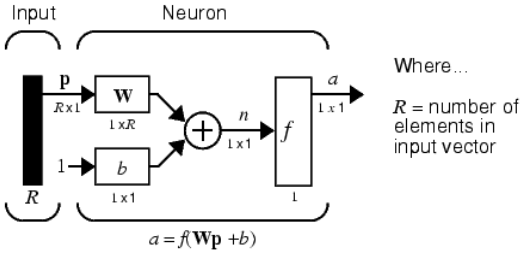


شکل ۱-۴- نمایش یک نرون تک‌لایه

دو یا چند نرون می‌توانند در یک لایه با هم ترکیب شوند و یک شبکه می‌تواند از یک یا چندلایه این‌چنینی تشکیل شود. در شبکه‌های مذکور اعضای بردار ورودی p به همه نرون‌ها اعمال می‌شوند و پس از ضرب در بردار وزن‌ها و جمع با بایاس به تابع انتقال اعمال شده و خروجی حاصل می‌گردد. خروجی شبکه بالا یک بردار 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.

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

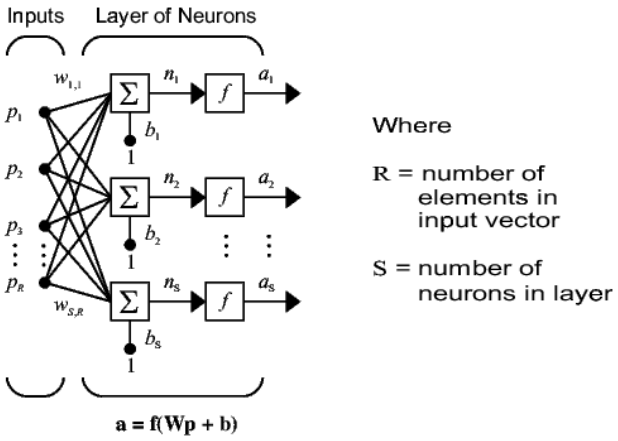
یک شبکه تک‌لایه با S نرون و R ورودی به صورت خلاصه در شکل ۱-۵ نشان داده شده است.



شکل ۱-۵- نمایش یک شبکه تک‌لایه با R ورودی و S نرون

۱-۳-۳-۱- ورودی‌ها و لایه‌ها

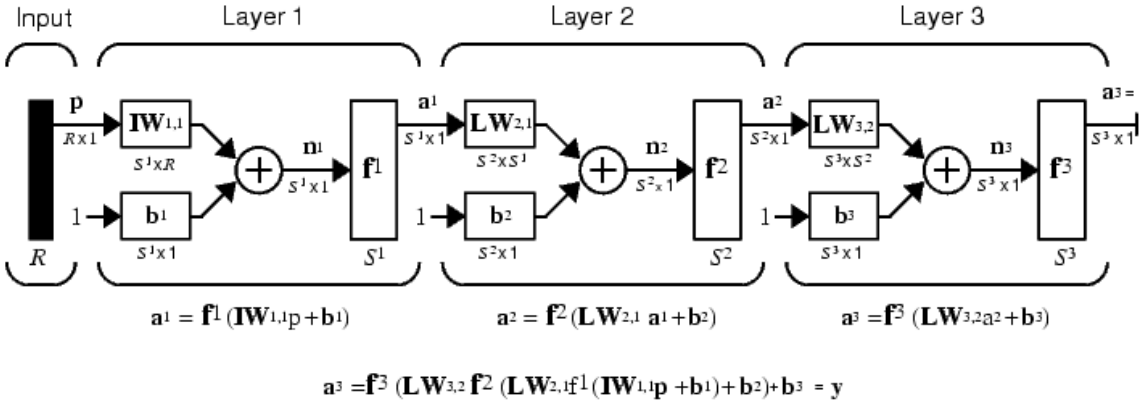
در شبکه‌های چندلایه لازم است ماتریس‌های وزن مربوط به ورودی‌های نرون‌ها و ماتریس‌های وزن بین لایه‌ها از هم قابل تشخیص باشند. بنابراین مبدأ و مقصد ماتریس‌های وزن باید معلوم باشند. این دو نوع ماتریس را به اختصار با IW (وزن‌های ورودی[۱]) و LW (وزن‌های لایه‌ها[۲]) نشان می‌دهیم. بنابراین تصویر یک شبکه یک‌لایه با S نرون و R ورودی به صورت زیر در خواهد آمد (شکل ۱-۶).



شکل ۱-۶- نمایش یک شبکه تک‌لایه با R ورودی و S نرون

با افزایش تعداد لایه‌ها به بیش از یک لایه، تصویر شبکه بصورت شکل ۱-۷ درخواهد آمد که 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.

[۱] Input Weights
[۲] Layer Weights



شکل ۱-۷- نمایش یک شبکه چند لایه

۱-۳-۳-۲- ساختار داده‌های[۱] مورد استفاده [۱۴]

در این بخش تأثیر فرمت ورودی‌ها در شبیه‌سازی شبکه مورد بررسی قرار می‌گیرد. دو گونه بردار ورودی قابل بحث اند: بردارهای ورودی همزمان[۲]، که ارائه داده‌ها به شبکه در یک زمان اتفاق می‌افتد، نه در توالی از زمان‌های مشخص و بردارهای ورودی ترتیبی[۳] که برخلاف بردارهای همزمان، در اینجا ترتیب بردارهای ورودی مهم است.

شبکه ایستا[۴]، شبکه‌ای است که در آن پسخورد[۵] و تأخیر[۶] وجود نداشته باشد. در اینگونه شبکه‌ها ورودی‌ها معمولاً همزمان خواهند بود.

شبکه‌های پویا[۷] 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، داده‌های ترتیبی در یک آرایه سلولی و داده‌های همزمان در بردار قرار می‌گیرند. در شبکه مورد استفاده برای این تحقیق، داده‌ها به صورت همزمان به شبکه ارائه شده‌اند.

---

[۱] Data structures
[۲] Concurrently inputs
[۳] Sequential inputs
[۴] Static
[۵] Feedback
[۶] Delay
[۷] Dynamic

۱-۳-۴- آموزش شبکه عصبی

معمولاً شبکه‌های عصبی تنظیم می‌شوند یا اینکه آموزش می‌یابند تا اعمال یک ورودی خاص، منجر به دریافت خروجی موردنظر یا همان هدف شود. همانطور که در شکل ۱-۸ 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 می‌گیریم:

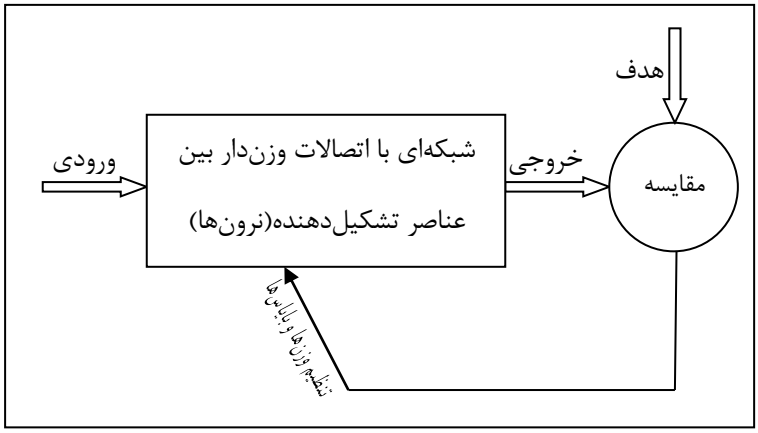
$$\{p_1, t_1\}, \{p_2, t_2\}, \ldots \{p_Q, t_Q\}$$

که در آن 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 وزن-ها 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.

عموماً برای آموزش شبکه‌های عصبی از قواعد یادگیری نظارت شده استفاده می‌شود اما می‌توان شبکه‌ها را با روش آموزش غیر نظارتی نیز آموزش داد[۱۶].

[۱] Supervised Training

[۲] Unsupervised training



شکل ۱-۸- شمای کلی عملکرد شبکه عصبی مورد بحث

۱-۳-۴-۱- 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 به‌روز 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.

در این تحقیق شیوه آموزش دسته‌ای در شبکه‌های پویا به کار گرفته شده و توابع مورد استفاده برای آموزش trainbr وtrainlm می‌باشند [۱۶].

[۱] Incremental training
[۲] Batch training

۱-۳-۵- شبکه‌های پس انتشار[۱]

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[۱۷].

این شبکه‌ها که از آنها، به اختصار تحت عنوان BP نیز یاد می‌شود، شبکه‌هایی چندلایه با قاعده یادگیری ویدرو - هوف[۲] 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[۳] حرکت می‌کنند. در واقع لغت پس انتشار نیز به رفتار شبکه BP در محاسبه شیب در شبکه‌های غیر خطی چندلایه اشاره دارد. الگوریتم‌های مختلفی مانند روش‌های نیوتن[۴] 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] Back Propagation (BP)

[۲] Widrow Hoff

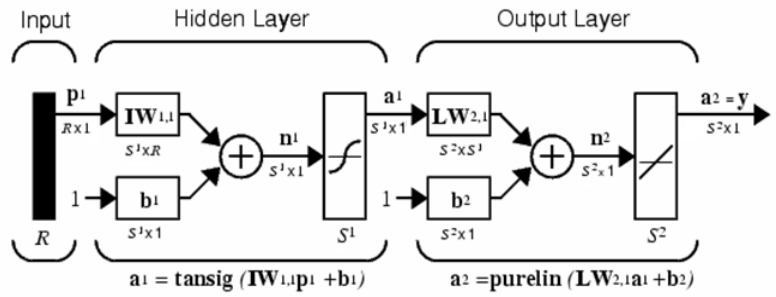
[۳] Performance Function

[۴] Newton methods

- فراهم کردن داده‌های آموزشی
- ایجاد شبکه بهینه
- آموزش شبکه
- شبیه‌سازی شبکه با داده‌های جدید

۱-۳-۵-۱- شبکه‌های پیش‌خور[۱]

شبکه‌های پیش‌خور اغلب داری 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 ۱-۹، یک شبکه دو لایه tansig/purelin نشان داده شده است. این شبکه می‌تواند برای تقریب زدن هر تابع با تعداد محدود ناپیوستگی استفاده شود. ایجاد و آموزش یک شبکه پیش‌خور با استفاده از جعبه ابزار MATLAB دارای مراحلی است که در اینجا اشاره مختصری به این مراحل می‌شود:



شکل ۱-۹- شبکه دو لایه tansig/purelin

[۱] Feed forward

- ایجاد شیئی شبکه: 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 ماتریس R×۲ به عنوان مقادیر مینیمم و ماکزیمم هر ورودی از ماتریس ورودی (R) 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.
- 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. فرآیند آموزش به یک سری مثال‌ها از رفتار مورد انتظار شبکه نیاز دارد که شامل ورودی شبکه (P) و هدف (T) می‌باشد. در طول فرآیند 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.

$$mse = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t - a_i)^2 \qquad (۱۰-۱)$$

در رابطه (۱-۱۰)،t مقدار خروجی مورد انتظار برای نمونه ورودی *i*ام و a_i مقدار محاسبه شده توسط شبکه برای آن است.

[۱] Mean Square Error

- شبیه سازی: تابع sim پس از اتمام آموزش به کار گرفته می‌شود و شبکه را شبیه‌سازی می‌کند. این تابع شبکه آموزش دیده و بردار ورودی را به عنوان ورودی دریافت کرده و خروجی شبکه را بر می‌گرداند[۱۸].

۱-۳-۵-۲- 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.

۱-۳-۵-۲-۱- 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:

$$\mathbf{x}_{k+1} = \mathbf{x}_{k} - \alpha_k g_k \qquad (۱۱-۱)$$

که $\mathbf{x}_{k}$ یک بردار از وزن‌ها و بایاس‌های فعلی در k امین تکرار است، g_k شیب فعلی[۱] و α_k سرعت یادگیری[۲] است[۱۸].

دو روش مختلف برای اجرای این الگوریتم پیاده‌سازی شده است، روش گام به گام و روش دسته‌ای که قبلاً به آنها اشاره شد.

[۱] Current Gradient
[۲] Learning rate

۱-۳-۵-۲-۲- الگوریتم لونبرگ- مارکوارت[۱]

الگوریتم لونبرگ-مارکورت روشی ساده و در عین حال توانمند در تقریب زنی یک تابع است و اساسا شامل حل معادله زیر می‌باشد:

$$(\boldsymbol{J}^t\boldsymbol{J} + \mu I)\delta = \boldsymbol{J}^t\boldsymbol{E} \qquad (۱۲-۱)$$

در رابطه فوق **J** ماتریس ژاکوبین[۲] برای سیستم، μ ضریب میرایی لونبرگ[۳]، δ بردار بروزرسانی وزن‌ها و **E** بردار خطا شامل خطاهای خروجی برای هر بردار ورودی مورد استفاده برای آموزش شبکه می‌باشد. δ نشان می‌دهد که چقدر باید وزن‌های شبکه را تغییر دهیم تا به راه حل (احتمالا) بهتری دست یابیم. J^tJ را نیز می‌توان به عنوان تقریبی از ماتریس هسیان[۴] تلقی نماییم.

ژاکوبین ماتریسی است از تمام مشتقات جزئی مرتبه اول از یک تابع بردار- مقدار، و در مورد شبکه عصبی، ماتریسی N در W است که در آن N تعداد ورودی‌ها در سری آموزش و W تعداد کل پارامترهای شبکه عصبی (وزن‌ها + بایاس‌ها) است. این ماتریس را می‌توان بوسیله گرفتن مشتق جزئی از هر خروجی با توجه به وزن ایجاد نمود و به شکل زیر می‌باشد:

$$J = \begin{bmatrix} \frac{\partial F(x_1,w)}{\partial w_1} & \cdots & \frac{\partial F(x_1,w)}{\partial w_W} \\ \vdots & \ddots & \vdots \\ \frac{\partial F(x_N,w)}{\partial w_1} & \cdots & \frac{\partial F(x_N,w)}{\partial w_W} \end{bmatrix}$$

شکل ۱-۱۰- شماتیک کلی از یک ماتریس ژاکوبین

[۱] Levenberg -Marquardt
[۲] Jacobian Matrix
[۳] Levenberg's damping factor
[۴] Hessian Matrix

در اینجا $\mathbf{F(x_i, w)}$ تابع شبکه است و با استفاده از بردار وزن $\mathbf{w}$ برای iامین بردار ورودی از سری آموزش محاسبه می‌شود، w_j نیز jامین عنصر از بردار وزن $\mathbf{w}$ شبکه می‌باشد.

در اجرای لونبرگ-مارکورت سنتی، ژاکوبین با استفاده از روش تفاضل متناهی تقریب زده می‌شود. اما در شبکه‌های عصبی، می‌توان این ماتریس را بطور بسیار کارآمدتری بوسیله قانون مشتق‌گیری زنجیری در حساب دیفرانسیل، و مشتقات اول توابع فعالیت محاسبه نمود.

μ، ضریب میرایی، در هر دور آموزش تنظیم می‌گردد و فرآیند بهینه‌سازی را هدایت می‌کند. اگر E بسرعت کاهش یابد، برای نزدیک کردن این تابع به الگوریتم گوس-نیوتون[۱]، مقادیر کوچک‌تری از μ استفاده می‌شود، در حالیکه اگر در یک تکرار، کاهش قابل توجهی در مقدار خطا مشاهده نشود، تا رسیدن به مرحله‌ای که به مسیر نزول گرادیان نزدیک‌تر شود، μ افزایش می‌یابد.

زمانی که مقدار عددی μ صفر باشد، این تابع تبدیل به یک روش نیوتن برای تقریب ماتریس هسیان می‌شود. زمانی که مقدار عددی μ بزرگ باشد، این تابع تبدیل به روش گرادیان نزولی با گام کوچک می‌شود. روش نیوتن نسبت به روش شیب توأم، دقیق‌تر است. بنابراین μ بعد از هر گام موفق (کاهش در تابع کارآیی) کاهش یافته و تنها وقتی که گام آزمایشی تابع کارآیی را افزایش دهد، افزایش داده می شود. در این روش، تابع کارآیی همیشه در هر تکراری از الگوریتم کاهش می‌یابد. پارامترهای آموزشی برای الگوریتم لونبرگ-مارکوارت (trainlm) عبارتند از: دوره‌ها[۲]، نمایش[۳]، هدف[۴]، زمان[۵]، گرادیان-مینیمم[۶]، رد شدن- ماکزیمم[۷]،mu، کاهش- mu[۸]، افزایش-mu[۹] و ماکزیمم- mu[۱۰].

---

[۱] Gauss-Newton algorithm
[۲] Epochs
[۳] show
[۴] goal
[۵] time
[۶] min-grad
[۷] max-fail
[۸] mu-dec
[۹] mu-inc
[۱۰] mu-max

پارامتر mu مقدار آغازین μ است. این مقدار در هر گام که تابع کارآیی کاهش یابد درmu-dec و هر زمان که افزایش یابد در mu-inc ضرب می‌شود. اگر mu از mu-max بزرگ‌تر شود، الگوریتم متوقف می‌شود. این روش یکی از سریعترین روش‌های پیاده سازی شده در MATLAB می‌باشد و برای یک شبکه متوسط (با چند صد پارامتر موثر) دارای کارایی بسیار بالایی است[۱۸].

۱-۳-۵-۲-۲-۱- تقریب زدن هسیان

برای مسئله حداقل مربعات، عموما نیازی به محاسبه هسیان نیست. همانطور که پیش از این گفته شد، این ماتریس را می‌توان با استفاده از ماتریس ژاکوبین و به کمک رابطه زیر تقریب زد:

$$\boldsymbol{H} \approx \boldsymbol{J}^t\boldsymbol{J} \qquad (۱۳-۱)$$

اگر خطاهای باقیمانده در راه حل مسئله کوچک باشند این رابطه تقریب بسیار خوبی از هسیان را ارائه می‌دهد ولی اگر خطاهای باقیمانده به اندازه کافی کوچک نباشد این تقریب ممکن است موجب کندی همگرایی گردد. ماتریس هسیان را همچنین می‌توان در بکار بردن فرآیند تنظیم مورد استفاده قرار داد که بعدا به آن می‌پردازیم.

۱-۳-۵-۲-۲-۲- حل رابطه لونبرگ-مارکورت

سهم اصلی لونبرگ در این روش، تعریف ضریب میرایی μ است. پیش از اینکه سیستم برای گرادیان حل شود، این مقدار با هر یک از اعضای قطری ماتریس هسیان جمع می‌گردد. عموما μ با مقادیر کوچکی مانند ۰/۰۰۱ شروع می‌شود. سپس رابطه لونبرگ-مارکورت با استفاده از یک تجزیه LU[۱]حل می‌شود. البته این سیستم فقط در شرایطی که تقریب ماتریس هسیان منفرد نشده باشد

[۱] LU decomposition

(معکوس داشته باشد) می‌تواند حل شود. در غیر اینصورت این رابطه را می‌توان با استفاده از تجزیه مقادیر منفرد (SVD)[1] حل نمود.

پس از اینکه رابطه حل شد، وزن‌های w با استفاده از δ بروزرسانی شده و خطاهای شبکه برای هر یک از ورودی‌ها در سری آموزش دوباره محاسبه می‌شوند. اگر مجموع مربعات خطاهای جدید کاهش یافته باشد μ نیز کاهش می‌یابد و دور تمام می‌شود. در غیر اینصورت وزن‌ها کاهش داده می‌شوند و روش با یک مقدار بیشتری از μ تکرار می‌شود.

تنظیم μ با استفاده از فاکتور تنظیم v انجام می‌گردد. اگر μ نیاز به افزایش داشته باشد در v ضرب می‌شود و اگر نیاز به کاهش داشته باشد بر v تقسیم می‌گردد. این فرآیند تا زمانی که خطا کاهش یابد تکرار می‌شود. وقتی خطا کاهش یافت این دور خاتمه می‌یابد. فاکتور v در واقع همان پارامترهای mu-dec و mu-inc است، که پیش از این در مورد آنها توضیح داده شد.

۱-۳-۵-۲-۲-۳- الگوریتم عمومی لونبرگ مارکورت

همانطور که قبلا گفته شد، اساسا لونبرگ-مارکورت شامل حل رابطه (۱-۱۲) با تفاضل مقادیر μ است تا زمانی‌که مجموع مربعات خطا کاهش یابد. بطور کلی هر دور آموزش شامل مراحل اصلی زیر خواهد بود:

۱. محاسبه ژاکوبین (با استفاده از تفاضل متناهی یا قانون زنجیری)

۲. محاسبه گرادیان خطا

$$\boldsymbol{g} = \boldsymbol{J}^{\boldsymbol{t}}\boldsymbol{E} \qquad (۱-۱۴)$$

۳. تقریب زدن هسیان با استفاده از حاصلضرب خارجی ژاکوبین

[1] Singular Value Decomposition (SVD)

$$\boldsymbol{H} = \boldsymbol{J}^{t}\boldsymbol{J} \qquad (۱۵-۱)$$

۴. حل رابطه $(\boldsymbol{H} + \mu I)\delta = \boldsymbol{g}$ برای پیدا کردن δ.

۵. بروزرسانی وزن‌های شبکه w با استفاده از δ.

۶. محاسبه مجدد مجموع مربعات خطاهای شبکه با استفاده از وزن‌های بروزرسانی شده.

۷. اگر مجموع مربعات خطاها کاهش نیافته بود

- از وزن‌های جدید صرف نظر نموده و مقدار μ را با استفاده از v افزایش داده و به مرحله ۴ بازمی‌گردد.

۸. در غیر اینصورت مقدار μ را با استفاده از v کاهش داده متوقف می‌گردد.

۱-۳-۵-۳- 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 می‌شود. یکی از مشکلاتی که در طی آموزش شبکه عصبی رخ می‌دهد برازش اضافی[۲] است، یعنی خطای مربوط به سری آموزشی به یک مقدار خیلی کوچک رسانده می‌شود، اما زمانی که داده‌های جدید به شبکه ارائه می‌شوند، 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] [۱۸].

[۱] Generalization
[۲] Overfitting
[۳] Regularization
[۴] Early Stopping

۱-۳-۵-۳-۱- 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 به‌روز 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. همان گونه که خطای سری آموزشی کاهش می‌یابد، خطای ارزیابی نیز به طور طبیعی در طول فاز اولیه آموزش 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 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 فرآیند 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 به صورت تصادفی به سری‌های مورد نظر تقسیم شده‌اند.

۱-۳-۵-۳-۲- 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].

۱-۳-۵-۳-۲-۱- تنظیم خودکار

تعیین پارامترهای کارایی به صورت خودکار بسیار مطلوب می‌باشد. یک راه دستیابی به این فرآیند چهارچوب کاری بائزین[۱] می‌باشد. در این چهارچوب وزن‌ها و بایاس‌های شبکه مقادیر تصادفی با یک

[۱] Performance Function Modification
[۲] Automated Regularization

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[۱۸].

trainbr یک تابع آموزش برای شبکه است که وزن‌ها و بایاس را بر اساس بهبود الگوریتم لونبرگ-مارکوارت بروزرسانی می‌کند. این تابع، ترکیبی خطی از مربع خطاها و نیز وزن‌ها را به حداقل رسانده و سپس ترکیب صحیحی از آنها را برای ایجادیک شبکه با قابلیت تعمیم‌پذیری خوب مشخص 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 پارامتر شبکه در آموزش موثر بوده‌اند. 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 بائزین به الگوریتم لونبرگ-مارکوارت تغییراتی در آن به وجود می‌آورد. این الگوریتم به صورت زیر محاسبه می‌شود:

۱. محاسبه ژاکوبین.

۲. محاسبه گرادیان خطا.

۳. محاسبه هسیان با استفاده از حاصلضرب خارجی ژاکوبین.

۴. محاسبه تابع ارزش به صورت زیر:

$$C = \beta E_d + \alpha E_w \qquad (۱۶-۱)$$

۵. در رابطه فوق E_d جمع مربعات خطا وE_w جمع مربعات وزن‌ها می‌باشد.

۶. حل رابطه زیر برای پیدا کردن δ

$$(\boldsymbol{H} + \lambda I)\, \delta = \boldsymbol{g} \qquad (۱۷-۱)$$

---

[۱] Bayezian Regularization

۷. بروزرسانی وزن‌های شبکه w با استفاده از δ

۸. محاسبه مجدد تابع ارزش با استفاده از وزن‌های بروز رسانده شده

۹. اگر مقدار ارزش کاهش نیافته بود:

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

۱۰. در غیر این صورت μکاهش داده می‌شود.

۱۱. پارامترهای دیگر بائزین با استفاده از فرمول‌های مک‌کی[۱] و پولند[۲] بروزرسانی می‌شوند:

$$\alpha = W - (\alpha * tr(H^{-1})) \quad (۱۸-۱)$$

$$\beta = (N - \gamma)/2.0E_d \quad (۱۹-۱)$$

$$\alpha = \frac{W}{2.0}E_w + tr(H^{-1}) \quad \text{[بروزرسانی با فرمول اصلاح شده پولند]} \quad (۲۰-۱)$$

$$[\alpha=\gamma/(2.0E_w) \quad \text{[بروزرسانی با فرمول اصلی مک‌کی]} \quad (۲۱-۱)$$

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

۱-۳-۶- ارزیابی مدل

هدف از ارزیابی مدل، ارائه شاخص‌ها و پارامترهایی است که بتوان به کمک آنها دقت و صحت نتایج بدست آمده بوسیله مدل را مورد ارزیابی و بررسی قرار داد. از این شاخص‌ها ابتدا برای بهینه‌سازی پارامترهای مدل و سپس برای گزارش میزان اعتمادپذیری آن استفاده می‌شود. برای

[۱] MacKay

[۲] Poland

ارزیابی شبکه‌های عصبی معمولا داده‌ها را به سه سری آموزش، ارزیابی و تست تقسیم می‌کنند. ابتدا شبکه به وسیله سری آموزش، آموزش داده می‌شود و وزن‌ها و بایاس‌ها انتخاب می‌شوند سپس بوسیله سری ارزیابی کیفیت شبکه مورد بررسی قرار می‌گیرد. در صورت دریافت پاسخ نامناسب از شبکه مرحله آموزش وسپس ارزیابی آن با تغییر پارامترهای شبکه تا رسیدن به نتیجه مطلوب تکرار می‌گردد. پس از این مرحله جهت آزمودن قابلیت انطباق‌پذیری شبکه، آن را با نمونه‌های سری تست می‌سنجند.

در صورت کم بودن تعداد نمونه‌ها در سری داده‌ها از روش دیگری با عنوان روش اعتبارسنجی متقابل[1] استفاده می‌گردد. در اینجا سری داده‌ها به دو سری آموزش و تست تقسیم می‌شوند و از سری ارزیابی صرف نظر می‌گردد. نحوه عمل بدین ترتیب است که شبکه به روش رد مرحله‌ای تک‌تک مقادیری را برای سری آموزش پیش‌بینی می‌کند، سپس پارامترهای آماری برای مقادیر پیش‌بینی شده محاسبه می‌شوند و بر اساس اطلاعات به دست آمده بهترین شبکه انتخاب شده و جهت آزمودن قابلیت تعمیم‌پذیری آن، به وسیله نمونه‌های سری تست مورد سنجش قرار می‌گیرد. این روش محدود به زمانی که تعداد اعضای سری داده‌ها کم باشند نیست و در مواقعی هم که تعداد اعضای سری داده‌ها به اندازه کافی زیاد باشند نیز می‌تواند مورد استفاده قرار گیرد.

[1] Cross Validation

فصل دوم

بررسی سرطان وHIV

۲-۱- سرطان

سلول، واحد اساسی و ساختمانی حیات است که همانند کیسه‌ای حاوی پروتئین‌ها، اسیدهای چرب، کربوهیدرات‌ها و ماده حیات به نام دی ان ای[۱] است. قابلیت رشد، تکثیر و همانندسازی از ویژگی‌های سلول‌های زنده است. ساختار ژنتیکی هر سلول سرعت رشد، تقسیم و زمان مرگ آن را تعیین می‌کند. در حالت طبیعی، جایگزینی سلول‌های فرسوده با سلول‌های جوان از یک برنامه منظم تبعیت می‌کنند و فرایند رشد و تجدید سلولی به طور ثابت در بدن اتفاق می‌افتد.

سرطان نوعی بیماری است که در آن سلول‌ها توانایی تقسیم و رشد عادی خود را از دست می‌دهند و این موضوع منجر به تسخیر، تخریب و فاسد شدن بافت‌های سالم می‌شود. از اجتماع این سلول‌های سرطانی و تخریب سلول‌های بافت‌های سالم توده‌ای به نام تومور ایجاد می‌شود. اگر تومور به لایه‌ای محدود ختم شود و به سایر بافت‌ها و ارگان‌ها سرایت نکند تومور خوش خیم (غیرسرطانی) است و اگر تومورگسترده شده یا به طور بالقوه قابلیت پخش شدن و احاطه کردن سایر بافت‌ها و ارگان‌ها را داشته باشد بدخیم یا سرطانی نامیده می‌شود. برخی از شکل‌های سرطان متاستاز[۲] می‌کنند به این معنی که خصوصیت تهاجمی پیدا کرده و به سایر بافت‌های بدن، عمدتا از طریق خون و لنف، سرایت می‌کنند و تومورهای جدیدی را ایجاد می‌نمایند.

در مجموع چهار گروه ژن مسؤول تقسیم سلولی به شمار می‌روند:

۱. آنکوژن‌ها[۳] (ژن‌های عامل تومور): این ژن‌ها درشرایط عادی در فرستادن پیام به سلول برای تکثیر نقش دارند. اختلال و تغییر در این سلول‌ها منجر به تکثیر نامنظم سلول شده و سلول سرطانی بشمار می‌رود.

---

[۱] DNA

[۲] Metastasize

[۳] Oncogene

۲. ژن‌های سرکوبگر تومور[۱]: این ژن‌ها پروتئین‌های خاصی را تولید می‌کنند که در شرایط عادی وظیفه معکوس آنکوژن‌ها را داشته و به سلول پیام توقف تکثیر می‌دهد. یکی از مهم‌ترین ژن‌های این گروه ژنی به نام p53 است.

۳. ژن‌های خودکشی[۲]: خودکشی سلول‌ها یا مرگ سلول یکی از مهم‌ترین عوامل پیچیده سلولی است که به سلول توانایی خودکشی در شرایط غیرمعمول را می‌دهد تا مانع شیوع تکثیر و آسیب‌دیدگی به سایر سلول‌ها شود. هنگامیکه ژن‌های خودکشی آسیب پیدا کنند دیگر قادر به فعالیت خود برای نابود کردن سلول معیوب نبوده و سلول سرطانی محسوب می‌شود.

۴. ژن‌های ترمیمی دی ان ای[۳]: این ژن‌ها مسئول ترمیم دی ان ای آسیب دیده و معیوب هستند که با ترشح پروتئین‌های متفاوت زمینه ترمیم دی ان ای آسیب دیده را فراهم می‌کنند. اما زمانی که خود این ژنهای ترمیمی دی ان ای آسیب می‌بینند، سلول، دیگر توانایی ترمیم خود را از دست داده و اختلالات ژنتیکی و ترمیم نشدن دی ان ای منجربه سرطان می‌شود.

۲-۱-۱- انواع سرطان‌ها

اغلب سرطان‌ها به سه دسته عمده تقسیم می‌شوند :

❖ **کارسینوم**[۴]: شامل سرطان‌هایی می‌شود که از سلول‌هایی که سازنده پوست هستند (مثل سرطان پوست) و یا لایه داخلی اعضا را می‌پوشانند (مثل سرطان ریه) و یا سازنده غدد هستند (مثل سرطان سینه) منشا می‌گیرد.

[۱] Tumour suppressor genes
[۲] Suicide genes
[۳] DNA repairing genes
[۴] Carcinoma

❖ **سارکوم[1]:** این دسته سرطان‌هایی هستند که از بافت همبند مثل غضروف، استخوان و ماهیچه منشاء می‌گیرند. از این‌رو سرطان استخوان یا سرطان ماهیچه را در هر نقطه از بدن سارکوم می‌گویند.

❖ **لوسمی[2] و لنفوم‌ها[3]:** شامل سرطان‌هایی است که از سلول‌های تشکیل دهنده خون و سلول‌های ایمنی منشاء می‌گیرند.

عواملی که به ایجاد سلول‌های سرطانی کمک می‌کنند عبارتند از:

۱. استعداد میزبان: عوامل ژنتیکی از جمله نقایصی در کروموزوم‌ها و یا انتقال ژن معیوب به جنین.
۲. عوامل ایمنولوژیک(ایمنی): مانند نارسایی مکانیسم ایمنی طبیعی بدن.
۳. داروهای سرکوبگر ایمنی: موجب سرکوب مکانیسم ایمنی طبیعی شده و زمینه ابتلا به سرطان را فراهم می‌کنند.
۴. عوامل محیطی: تماس با مواد سرطان‌زا مانند آزبست،پرتونگارها و رادیو- پرتوهای یونیزه شامل امواج الکترومغناطیسی.
۵. ویروس‌های القا کننده سرطان: ویروس‌هایی هستند که قابلیت تغییر دادن شکل سلول آلوده شده با ویروس را دارند و در نتیجه منجر به تکثیرخارج از کنترل سلول‌های مورد نظر می‌شوند. این تکثیر فزاینده موجب تومور یا سرطان می‌شود.
۶. عوامل ترشح هورمونی: عوامل ترشح هورمونی اغلب موجب تسریع روند بدخیمی بیماری می‌شوند.

---

[1] Sarcoma
[2] Leukemia
[3] Lymphocytes

۲-۱-۲- سرطان سینه

سرطان سینه، شایع ترین سرطان در زنان و اولین علت مرگ ناشی از سرطان در زنان ۴۰ تا ۴۴ ساله است. این بدخیمی، ۳۳٪ سرطان‌های خانم‌ها راتشکیل داده و مسؤول ۱۹٪ از مرگ‌های وابسته به سرطان می‌باشد. آمار و شواهد حاکی ازافزایش مداوم شیوع سرطان سینه از اواسط دهۀ ۱۹۴۰ هستند. به عنوان مثال، در ایالات متحدۀ امریکا این میزان از نسبت ۱ به ۱۳ در سال ۱۹۷۰، به ۱ به ۱۱ در سال ۱۹۸۰ و ۱به ۸ در سال ۱۹۹۶ رسیده است.

۲-۱-۲-۱- اپیدمیولوژی

سرطان سینه یک مسألۀ مهم اپیدمیولوژیک با گسترش جهانی است. زنان ساکن کشورهای صنعتی در مقایسه با سایرین، ریسک ابتلای بالاتری دارند که البته کشور ژاپن در این میان یک استثناء است. تشخیص سرطان سینه یکی از ناخوشایندترین وقایعی است که ممکن است در طول زندگی یک زن رخ دهد.

۲-۱-۲-۲- علل

در این بیماری، تکثیر بدخیم سلول‌های اپیتلیالی[۱] پوشانندۀ مجاری یا لبول‌های[۲] سینه رخ می‌دهد. مانند تمام بدخیمی‌های اپیتلیالی، میزان بروز سرطان سینه نیز با افزایش سن تدریجاً بالا رفته، ولی از سن قطع قاعدگی شیب این منحنی کاهش می‌یابد. سن شروع قاعدگی، اولین بارداری و نیز یائسگی، سه تاریخ مهم تأثیرگذار بر وقوع سرطان سینه در زنان هستند. بدین ترتیب که سن پایین‌تر شروع قاعدگی، سن بالاتر اولین بارداری و همچنین سن بالاتر بروز یائسگی، ریسک ابتلا را افزایش می‌دهند که این موارد نشانگر وابستگی این سرطان به هورمون‌های جنسی است. زنانی که

---

[۱] Epithelial

[۲] Lobule

سن شروع قاعدگی آنها ۱۶ سالگی بوده است، حدود ۶۰-٪۵۰ افرادی که در سن ۱۲ سالگی قاعده شده‌اند بـه سرطان سینه مبتلا می‌شوند.

به همین ترتیب، این ریسک در زنانی که سن قطع قاعدگی آنان ۱۰ سال زودتر از متوسط سن طبیعی آن (۵۲ سالگی) است، به ٪۳۵ سایرین می‌رسد یا در زنانی که به واسطهٔ عمل جراحی برداشتن تخمدان‌ها دچار قطع قاعدگی زودرس شده‌اند، ریسک سرطان سینه یک سوم مواردی است که منوپاز طبیعی در آنان، در سن ۵۰ سالگی یا بیشتر رخ می‌دهد. همچنین افرادی که اولین بارداری خود را در سن زیر ۲۰ تجربه کرده‌اند، نسبت به زنان نولیپار، ۴۰-۳۰٪ کمتر ریسک ابتلا به سرطان سینه دارند. علت این امر، قرارگیری مداوم در معرض استروژن داخلی در غیاب غلظت کافی پروژسترون است. بنابراین، طول دورهٔ قاعدگی و به ویژه آن بخش از این دوره که پیش از اولین بارداری طی می‌شود، از عوامل مؤثر برایجاد سرطان سینه هستند. این فاکتور می تواند مسؤول ۸۰-۷۰٪ علت تفاوت در میزان وقوع سرطان سینه بین ملیت‌های مختلف باشد. برای مثال، مبتلایان به سرطان سینه درایران نسبت به این بیماران در کشورهای غربی الگوهای متفاوتی دارند. میانگین سنی این بیماران در کشور ما کمتر از سایرین گزارش شده است. براساس مطالعات متعدد، وجود سابقه خانوادگی سرطان سینه و وضعیت ازدواج فرد، دو عامل مهم تأثیرگذار در بیماران ایرانی بوده‌اند. سهم بیشتر مبتلایان جوان در ایران می‌تواند تا حد زیادی به دلیل ساختار جوان جمعیت کشور ما توجیه شود. همچنین دو عامل قاعدگی در سن بالاتر و اولین بارداری در سن پایین‌تر، عوامل حفاظتی مهم در برابر ابتلا در سنین بالاتر هستند. بنابراین در حضور این دو فاکتور، شیوع این سرطان به سمت سنین پایین‌تر گرایش می‌یابد. در این بین بخصوص زمان اولین حاملگی فول‌ترم، مهمترین عامل حفاظت کننده است. براساس نتایج تحقیقات، زایمان نکردن ریسک سرطان سینه را در سن پایین‌تر کاهش داده و بالعکس موجب افزایش ریسک آن در سنین بالاتر می‌شود. تفاوت‌های جغرافیایی موجود درمیزان بروز و مرگ و میر ناشی از سرطان سینه نشان داده‌اند که ریسک

فاکتورهای این بیماری در نواحی مختلف متفاوت بوده و برای ابتلا بـه سرطان سینه، عمـدتاً فاکتورهای محیطی نسبت به فاکتورهای ژنتیکی قوی‌تر عمل می‌کنند.

❖ عوامل وراثتی:

نقش فاکتورهای ارثی و ژنتیکی به عنوان عوامل مستعد کننده در ابتلا به سرطان سینه تأیید شده است. یک سوم از کل مبتلایان، دارای سابقهٔ مثبت وقوع سرطان سینه در یک نفر یا تعداد بیشتری از وابستگان درجهٔ اول یا دوم خود هستند. در افرادی که خواهر یا مادر مبتلا به سرطان سینه داشته‌اند، ریسک ابتلا ۴ برابر است. اما این بدین معنی نیست که در صورت سابقه‌ی ابتلای مادر یا خویشاوند درجه یک، شما هم حتماً مبتلا خواهید شد. سرطان سینه ناشی از ژن‌های دخیل در این بیماری و فاکتورهای وراثتی، معمولاً در سنین پایین (زیر ۴۰ سالگی)، در چند نفر از خویشاوندان بیمار و در یک یا هر دوسینه (البته نه به طور همزمان) مشاهده می‌شود. ژن‌های مختلف شناسایی شده می‌توانند خصوصیات بافت‌شناسی و شدت تهاجم تومور را تعیین کنند.

❖ رژیم غذایی:

مصرف مواد غذایی پر چرب و سرخ کرده، احتـمال بـروز سرطان سینه را تا دو برابرافزایش می‌دهد. کیفیت و کمیت چربی رژیم غذایی، هر دو بر ایجاد این بیماری مؤثرند. همچنین بین مصرف متوسط الکل با بروز این بیماری رابطه دیده شده که مکانیسم آن هنوز ناشناخته است.

❖ مصرف هورمون:

تعیین نقش بالقوهٔ هورمون‌ها در ابتلا به سرطان سینه از اهمیت بسیاری برخورداراست، چرا که میلیون‌ها نفر از زنان، مرتباً در حال مصرف داروهای ضد بارداری خوراکی و یا دریافت درمان‌های جایگزینی هورمونی در دوران یائسگی هستند. این عقیده وجود دارد که مصرفOCP (قرص ضدبارداری خوراکی) حتی به صورت طولانی مدت، تأثیری درافزایش میزان بروز سرطان سینه ندارد. ولی مطالعاتی نیز نشان داده‌اند که مصرف طولانی مدت قرص‌های ترکیبی در افرادی که سن خیلی

پایین داشته ویا بارداری را پشت سرنگذاشته‌اند، می‌تواند در بـروز سرطان مـؤثـر باشد. اما پیشنهاد می‌کنند کـه نقش ایـن عوامل در ایجاد سرطان سینه حتی درصورت مثبت بودن، بسیار اندک خواهد بـود. بالعکس، داروهای ضـد بارداری خوراکی نقـش محافظتی در برابر تومورهای اپیتلیومی تخمدان و سرطان اندومتر رحم دارند. بررسی‌های سازمان بهداشت جهانی نشان می دهد که استفاده از داروی ضد بارداری تزریقی[1] تأثیری در افزایش ریسک سرطان سینه ندارد. هنوز مشخص نشده است که درمان جایگزینی هورمونی[2] (HRT)، در صورت وجود سوابق خانوادگی از سرطان سینه یا یافته‌های مثبت در بیوپسی قبلی، میزان خطر سرطان سینه را تغییر می‌دهد یا خیر. بعضی پژوهشگران، مصرف استروژن در سنین قبل و بعد از یائسگی را که به منظور جبران کمبود هورمونی صورت می گیرد، با افزایش مختصر در ریسک سرطان سینه همراه می‌دانند. به ویژه اگر سابقه ای از بیماری‌های خوش‌خیم سینه وجود داشته باشد. ولی در صورت مصرف استروژن به همراه پروژسترون، اطلاعات کافی از مقدار این تأثیر موجود نیست. در مجموع پیشنهاد شده که به طور متوسط زنان از HRT سود می‌برند.

❖ چاقی:

بنابر اکثر اطلاعات موجود، وزن بالا با بروز سرطان سینه ارتباط دارد. میزان ریسک در زنان چاق ۱/۵ تا ۲ برابر سایرین است که این ریسک مربوط به دوران یائسگی می‌باشد.

❖ سابقهٔ سرطان اولیه:

در زنانی که سابقه‌ی مثبت قبلی از بروز سرطان سینه دارند، ریسک ابتلای سینه سمت مقابل، ۳ تا ۴ برابر است. همچنین با وجود سابقهٔ قبلی از کارسینوم تخمدان یا رحم، احتمال بروز سرطان در سینه ۱/۳ تا ۱/۴ برابر افزایش می یابد.

---

[1] Depot MedroxyProgestrone Acetate (DMPA)

[2] Hormon Replacement Therapy (HRT)

❖ **پرتوتابی:**

در زنان جوانی که در معرض تابش اشعه به قفسهٔ سینه قرار گرفته‌اند، ریسک بالاتری از ابتلا به این بیماری گزارش شده است. ولی در سنین بالای ۳۰ سال، این میزان به کمترین مقدار خود می‌رسد. ریسک حاصله از پرتوتابی‌های متعدد با دوز پایین، با مورد مشابهی که یک نوبت پرتوتابی با دوز بالا انجام می‌گیرد، یکسان است. کمتر از ۱٪ موارد سرطان سینه از پروسه‌های تشخیصی رادیولوژیک ناشی می‌شود. احتمال دارد که رادیوتراپی برای درمان سرطان سینه نیز خطر درگیری طرف مقابل را افزایش دهد. اما به دنبال پرتوتابی جهت درمان سرطان دهانهٔ رحم خطر ابتلا به سرطان سینه کاهش می‌یابد که علت این امر، کاستن از میزان استروژن است.

۲-۱-۲-۳- نشانه‌های بیماری

علائمی که نشانه احتمال ابتلا به این نوع از سرطان می‌باشند عبارتند از:

۱. وجود توده و یا افزایش ضخامت در سینه و یا ناحیه‌ی زیر بغل
۲. ترشح شفاف یا خونی از سینه
۳. پوسته‌ریزی در نوک سینه
۴. فرو رفتن نوک سینه به داخل
۵. قرمزی و یا تورم سینه
۶. فرو رفتگی پوست سینه که در آن پوست، حالتی شبیه به پوست پرتقال پیدا می‌کند.
۷. وقوع تغییر شکل سینه که موجب متفاوت شدن آن با سینه‌ی دیگر شود.
۸. زخم در پوست سینه که بهبود پیدا نکند.

۲-۱-۲-۴- راه‌های درمان

درمان سرطان سینه به عوامل مختلفی بستگی دارد. ممکن است درمان جراحی به صورت برداشتن جزئی تا کامل سینه انجام شود و درمان‌های مکمل مانند درمان دارویی، درمان هورمونی، شیمی درمانی و رادیوتراپی انجام گیرند.

همانطور که در بالا اشاره شد هورمون‌ها نقش مهمی در ابتلا و پیشرفت سرطان سینه بر عهده دارند بنابراین در مواردی که فرد به سرطان مبتلا می‌شود سعی می‌گردد تا با مهار هورمون‌ها از پیشرفت و یا عود بیماری جلوگیری نمایند. تحقیقات نشان داده که در سرطان سینه استروژن نقش بسزایی را ایفا می‌نماید.

۲-۱-۳- استروژن

استروژن، هورمونی است که به طور اصلی درتخمدان خانم‌ها تولید شده، و سبب رشد قسمت‌هایی از سینه می‌شود (بخصوص آن قسمت‌هایی که گیرنده استروژن دارند).تحقیقات متعدد در درمان سرطان سینه با تجویز مهارکننده‌های استرروژن یا کم‌کننده‌های تولید استروژن صورت گرفته است. یکی از پرمصرف‌ترین داروهای هورمونی که به وسیله مهار دریافت‌کننده‌ی استروژن، سرعت رشد سرطان را کم می‌کند تاموکسی‌فن[۱] یا نولوادکس[۲] می‌باشد که معمولا برای ۵ سال و به صورت قرص تجویز می‌گردد. تحقیقات نشان داده است که استفاده از تاموکسی‌فن به عنوان درمان ادجوانت[۳] (درمان کمکی بعد از جراحی اصلی) احتمال عود بیماری را کمتر کرده است. بررسی‌های اخیر به طور روشن نشان داده در بیماران با سرطان سینه در مراحل اولیه صرف‌نظر از سن بیمار، مصرف این دارو به او کمک می‌کند. تعدادی از بررسی‌ها، افزایش خفیف در ریسک سرطان آندومتر رحم در مراحل اول (سرطان جدار رحم) در زنانی که تاموکسی‌فن مصرف می‌کنند را نشان داده است این احتمال در زنانی

[۱] Tamoxifen
[۲] Nolvadex
[۳] Adjuvant

که بیش از ۵ سال دارو مصرف کرده باشند افزایش بیشتری می‌یابد. اثرات جانبی دیگر این دارو شامل افزایش وزن، گر گرفتگی، تغییرات خلقی و آب مروارید است.

رالوکسی‌فن دارویی شبیه به تاموکسی‌فن می‌باشد که اثر استروژن را روی بافت سینه و سرطان سینه مهار می‌کند. در خانم‌هایی که ریسک بالای سرطان سینه دارند تاثیر رالوکسی‌فن در حال بررسی می‌باشد.

مگسترول استات[۱] یک داروی دیگر هورمونی در درمان سرطان سینه پیشرفته است و معمولا در بیمارانی به کار می‌رود که به تاموکسی‌فن جواب نداده‌اند و یا بعد از مدتی به آن مقاوم شده‌اند.

دسته دیگر داروهای هورمونی از طریق مهار کردن آنزیم لازم برای تولید استروژن عمل می‌کنند مثل آناسترازول[۲]. برعکس تاموکسی‌فن، که به وسیله مهار دریافت‌کننده، سرعت رشد سرطان را کم می‌کند، آناسترازول مهارکننده‌ی آنزیم لازم برای تولید استروژن است. آناستروزول یک انتخاب جدید در خانم‌هایی با سرطان سینه پیشرفته است که تحت درمان با تاموکسی‌فن یا بعد از آن، بیماری رشد می‌کند. در تحقیقات اخیر اثر تاموکسی‌فن و آناسترازول در درمان کمکی سرطان سینه بررسی و اثر آنها با یکدیگر مقایسه می‌شود.

طبق تحقیقات جدید درخانم‌هایی که بعد از یائسگی نیاز به درمان هورمونی پیدا می‌کنند، از دسته دیگر این‌گونه داروها به نام مهارکننده‌های آروماتاز استفاده می‌شود (مانند لتروزول[۳] یا فمارا). آروماتازها در تبدیل تستوسترون[۴] به استرودیول[۵] و همچنین آندروستندیون[۶] به استرون[۷] دخالت دارند (شکل‌های ۲-۱ و۲-۲). مکانیسم و نحوه عمل آروماتازها در شکل ۲-۳ نشان داده شده است.

---

[۱] Megestrol acetate (magace)

[۲] Anastrazole

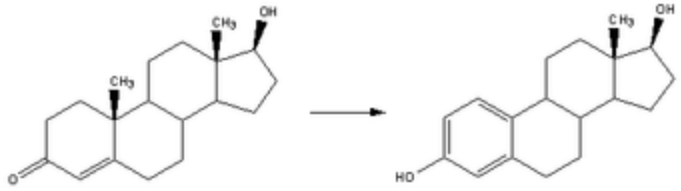
[۳] Letrozol (Femara)

[۴] Testosterone

[۵] Estradiol

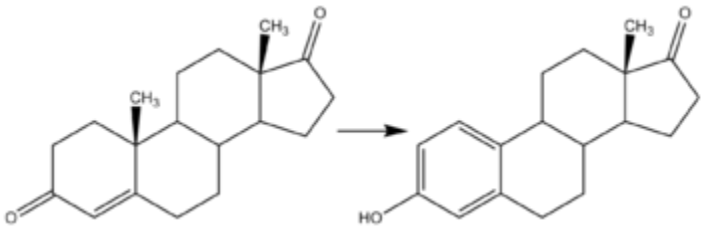
[۶] Androstendione

[۷] Estrone



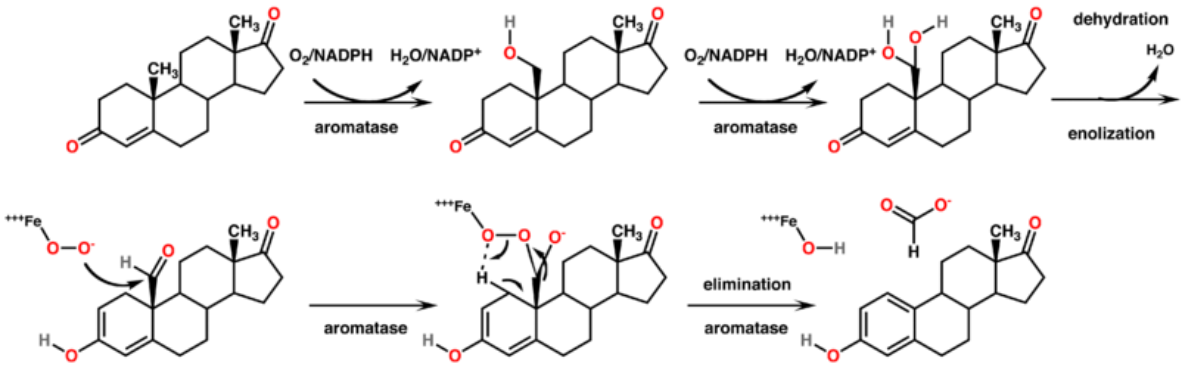
تستوسترون استرودیول

شکل ۲-۱- تبدیل تستوسترون به استرودیول بوسیله آروماتازها



استرون آندروستندیون

شکل ۲-۲- تبدیل آندروستندیون به استرون بوسیله آروماتازها



شکل ۲-۳- مکانیسم عمل آروماتازها در تبدیل آندروژن‌ها به استروژن‌ها

از داروهای هورمونی می‌توان جدید لتروزول اشاره نمود. روش‌های هورمونی دیگر در کنترل رشد سرطان سینه وجود دارد ولی تعداد آنها کم است. اضافه کردن پروژستین‌ها (هورمون دیگری که در سینه تولید می‌شود) یا آندروژن (هورمومن مردانه) ممکن است بعد از اینکه داروهای دیگر اثر نکند، در درمان سرطان سینه پیشرفته بکار رود. عوارض جانبی بر اساس داروی مصرفی است، اما با

پروژستین‌ها، شایع‌ترین عارضه جانبی احتباس مایعات است. آندروژن‌ها نیز سبب بروز بعضی از صفات مردانه مانند افزایش موهای صورت و صدای مردانه می‌شود.

با توجه به توضیحات ذکر شده، دانشمندان همچنان به دنبال یافتن ترکیباتی هستند که بتوانند به طور موثری در درمان سرطان سینه به کار گرفته شوند. همانطور که گفته شد استروژن‌ها در پیشرفت و عود مجدد سرطان سینه نقش مهمی دارند و در این موارد حضور آنها در بافت مطلوب نمی‌باشد. بنابراین باید به دنبال داروهایی بود که بتوانند به نحو موثری از فعالیت آنزیمی آروماتازها کاسته و در نتیجه موجب کاهش تولید استروژن‌ها در بدن فرد بیمار شده و از پیشرفت بیماری جلوگیری نمایند. یکی از جدیدترین ترکیباتی که در این زمینه مورد پژوهش و بررسی قرار گرفته‌اند سولفونانیلیدها هستند که بعنوان ترکیبات مهارکننده آروماتاز از آنها یاد می‌شود.

HIV -۲-۲

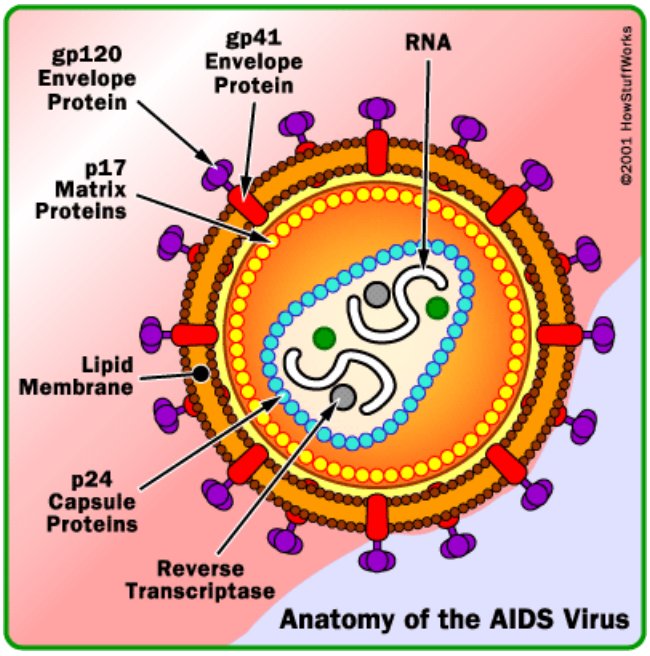
HIV -۱-۲-۲ چیست؟

HIV[۱] به معنی "ویروس نقص ایمنی انسان" می‌باشد. ویروس HIV از خانواده رترو ویروس‌ها[۲] و از زیر خانواده لنتی ویروس‌ها[۳] می‌باشد.

در میکروسکوپ الکترونی این ویروس به شکل دوازده وجهی با خارهای سطحی متعدد دیده می‌شود. دارای دو پوشش پروتئینی و اصلی ورای غشایی می‌باشد. ویریون[۴] (یک ذره ویروسی که توان آلوده کردن سلول را دارد) از سطح سلول‌های آلوده جوانه می‌زند و به محیط بیرون آزاد می‌شود. به همین علت است که ویروس HIV بطور مداوم پوشش خود را عوض می‌کند و به هنگام جوانه‌زدن از غشای سلول میزبان انواع متفاوتی پروتئین و گلیکوپروتئین جدا می‌کند که ترکیب آن منحصر به فرد

[۱] Human Immunodeficiency Virus
[۲] Retrovirus
[۳] lentivirus
[۴] Virion

می‌باشد. در شکل شماتیک از ویروس به طور واضح لایه‌ها و هسته‌ی ویروس قابل مشاهده است. دو قطعه RNA بصورت جدا از هم در مرکز ویروس مستقر هستند. این ویروس‌ها حاوی آنزیم ریورس ترانس کریپتاز معکوس هستند و به کمک همین آنزیم است که پس از ورود به داخل سلول می توانند از روی RNA ژنوم دورشته‌ای DNA را سنتز کنند. پس از این مرحله DNA سنتز شده به هسته سلول میزبان رفته و توسط آنزیم‌های خود در DNA سلول میزبان رخنه می‌کنند و سیستم همانندسازی‌کننده میزبان را وادار به همانندسازی از ژنوم خود کرده و در مراحل بعدی اجزای ویروس HIV به سرعت سنتز می‌شود و با پدیده جوانه‌زدن ازسلول میزبان سبب آسیب‌های جدی به غشای سلول میزبان شده و در نهایت سلول‌لیز می‌شود. و به این جهت که سلول‌های میزبان ویروس، گلبول‌های سفید می‌باشند سبب کاهش شدید این سلول‌ها شده و موجب تضعیف و ایجاد نقص در سیستم ایمنی بدن می‌شود.



شکل ۲-۴- تصویر شماتیک از ویروس ایدز

۲-۲-۱-۱- تفاوت بین HIV-1 و HIV-2 چیست ؟

در حال حاضر ۲ گونه[۱] از HIV وجود دارد:HIV-1 و HIV-2.

نوع غالب در جهان HIV-1 می‌باشد و هنگامی که بطور کلی و بدون مشخص کردن نوع و درباره HIV بحث می‌شود منظور HIV-1 می‌باشد. هر دو نوع HIV-1 و HIV-2 از طریق تماس جنسی، خون و محصولات خونی و از مادر و کودک منتقل می‌شوند و سبب ایدز با علائم بالینی غیر قابل افتراق از یکدیگر می‌شوند هر چند، HIV-2 مشکل‌تر از HIV-1 منتقل‌می‌شود و فاصله زمانی بین آلوده شدن با HIV-2 تا ایجاد بیماری طولانی‌تر است.

۲-۲-۱-۲- چرا HIV خطرناک است ؟

اگر سیستم ایمنی بدن به ویروس‌ها حمله می‌کند و آنها را می‌کشد پس مشکل چیست ؟

ویروس‌های مختلف به سلول‌ها و بافت‌های مختلفی از بدن حمله می‌کنند. برخی ویروس‌ها به پوست، برخی به دستگاه تنفسی و ... حمله می‌کنند. چیزی که HIV را این چنین خطرناک کرده این است که HIV به خود سیستم ایمنی حمله می‌کند. سیستم ایمنی، گروهی ازسلول‌ها هستند که بدن را در برابر انواع عفونت‌ها محافظت می‌کند و بدون آنها، توانایی بدن برای مبارزه با انواع عفونت‌ها تضعیف می‌شود. لذا وقتی ویروس HIV وارد بدن فردی شود به تدریج قدرت دفاعی بدن وی تضعیف می‌شود و این فرد در برابر انواع بیماری‌ها و عفونت‌ها حتی آنهایی که در حالت عادی بیماری‌زا نیستند آسیب پذیر می‌گردد. این روند قابل رؤیت نیست و راهی وجود ندارد تا با نگاه کردن به افراد بگوئیم که آیا به HIV مبتلا هستند یا خیر، ولی آزمایش خون می‌تواند پس از چند ماه از اولین تماس، ویروس را در خون آشکار کند.

افراد آلوده به HIV ممکن است سال‌ها کاملاً سالم بمانند و حتی خودشان ندانند که آلوده هستند. انسان می‌تواند توسط فرد دیگری که به HIV مبتلا است آلوده شود و او نیز می‌تواند بقیه افراد را

---

[۱] Type

آلوده کند و بر این اساس HIV منتشر می‌شود. به فردی که به HIV آلوده است ''HIV مثبت'' اطلاق می‌شود.

۲-۲-۲- ایدز چیست ؟

ایدز[۱] مخفف به معنی ''سندرم نقص ایمنی اکتسابی'' می‌باشد. ویروس ایدز سلول‌های ایمنی موجود در خون، لنفوسیت‌ها و سلول‌های ایمنی موجود در بافت‌ها مانند مغز استخوان، طحال، کبد و گره‌های لنفاوی را درگیر می‌سازد. این سلول‌ها در تولید پادتن برای مقابله با بیماری‌ها وسرطان‌ها نقش دارند. در مجموع باید گفت که ایدز یک نوع نقص ایمنی ثانویه است که در سیر عفونت با ویروس HIV ایجاد می‌شود.

۲-۲-۲-۱- چه مدت طول می‌کشد تا HIV به ایدز تبدیل شود؟

بدون درمان دارویی، بطور متوسط طی ۱۰ سال آلودگی به ویروس HIV به سمت ایدز پیش می‌رود که البته این مدت ۱۰ سال برای فردی است که تغذیه مناسبی دارد. اما فردی که در منقطه فقیرنشین است و بخوبی تغذیه نمی‌شود ممکن است بسیار سریع‌تر به سمت ایدز و نهایتاً مرگ پیش برود.

۲-۲-۳- راه‌های انتقال

HIV در خون، ترشحات جنسی فرد مبتلا و نیز شیر پستان زن آلوده، یافت می‌شود. HIV هنگامی منتقل می‌شود که مقدار کافی از این ترشحات به بدن فرد سالم برسد. راه‌های مختلفی که فرد سالم می‌تواند به HIV مبتلا گردد عبارتند از:

[۱] Acquired Immune Deficiency Syndrome (AIDS)

۱. رابطه جنسی بدون محافظت با فرد آلوده

۲. تماس با خون فرد آلوده

۳. سرایت از مادر باردار آلوده، به جنین در داخل رحم و یا از شیر مادر آلوده به کودک

ویروس HIV از مادر آلوده به کودک ممکن است در دوران حاملگی، هنگام وضع حمل یا کمی بعد از زایمان انتقال یابد. HIV می‌تواند در شیر پستان زن آلوده وجود داشته باشد و به نوزادش منتقل شود اما با وجود داروهای خاص، شانس این انتقال بسیار کاهش می‌یابد.

۴. استفاده از محصولات خونی آلوده

۵. استفاده از سرنگ‌های آلوده

افرادی که بصورت غیرقانونی داروی تزریقی استفاده می‌کنند نیز نسبت به عفونت HIV آسیب‌پذیر هستند (مثل معتادان تزریقی، که بصورت مخفیانه و غیرقانونی عمل کرده و معمولاً سرنگی برای تزریق در دسترس ندارند و از سرنگ مشترک استفاده می‌کنند).

۲-۲-۴- سیر بیماری و علایم آن

ابتلا به بیماری ایدز در اغلب موارد با عوارض ساده‌ای مانند اسهال، خارش پوست و افزایش خفیف حرارت بدن شروع می‌شود. این عوارض پس از چند هفته خودبخود برطرف می‌شود و شخص خیال می‌کند که به سرماخوردگی دچار بوده است. از آنجا که ساخته شدن پادتن‌های موثر حدود ۱۴ روز طول می‌کشد در این زمان جدال بین ویروس‌ها و پادتن‌ها در جریان است. پس از سپری شدن این مرحله دورانی فرا می‌رسد که حال شخص بیمار خوب می‌شود اما در خون او پادتن ضد ایدز را می‌توان تشخیص داد و از خون وی ویروس ایدز را بدست آورد.

لنفوسیتT_4 که مبتلا شده است شروع به تقسیم می‌کند همین امر باعث تورم غدد لنفاوی می‌شود. تورم غدد لنفاوی گاه ماه‌ها یا سال‌ها بعد از ابتلای اولیه پدیدار می‌شود و می‌تواند مدت‌ها به

همان حال باقی بماند. در بسیاری از موارد هم ناراحتی‌های جدی دیگری به آن اضافه می‌شود. بیماران اکثرا به شدت لاغر می‌شوند چون دیواره روده آنها مواد غذایی را بطور کامل جذب نمی‌کند.

عرق کردن در شب، احساس خستگی و التهاب بیضه‌ها از سایر علایم این بیماری است. مرحله بعدی که در آن تصویر نهایی بیماری ایدز به نمایش گذاشته می‌شود حاصل از درهم شکسته شدن کامل سیستم دفاعی بدن است. تعداد سلول‌های لنفوسیتT_4 به شدت کاهش یافته است و در این حالت انواع بیماری‌های باکتریایی، ویروسی و انگلی قادر هستند که فرد مبتلا را از پای بیاندازند.

۲-۲-۵- تشخیص بیماری

تشخیص این بیماری از طریق آزمایش خون امکان‌پذیر است. سه نوع کلی آزمایش (تست) تشخیصی HIV وجود دارد که عبارتند از:

۱. تست آنتی بادی HIV[۱]: این تست نشان می‌دهد که فرد به HIV آلوده شده است یا نه.

۲. تست آنتی ژن P24[۲]: این تست بطور اولیه برای غربالگری نمونه‌های خون استفاده می‌شود ولی در برخی مناطق آن را بعنوان تست تشخیصی HIV بکار می‌برند.

آنتی ژن P24، یک پروتئین است که جزئی از ساختمان HIV می‌باشد و در مراحل اولیه عفونت، به مقدار زیاد تولید می‌شود و بوسیلهٔ تست‌های تشخیصی می‌توان آن را در خون آشکار کرد. تست P24 می‌تواند آلودگی با HIV را قبل از تست آنتی بادی HIV آشکار کند. بنابراین، تست آنتی‌ژن P24 در تشخیص HIV در مراحل اولیه بکار می‌رود.

۳. تست میزان ویروس[۳]: این تست هنگامی استفاده می‌شود که شخص از آلوده بودن خود با HIV آگاه است و با این تست میزان ویروس در خون مشخص می‌شود.

---

[۱] HIV antibody test

[۲] P24 antigen testing

[۳] HIV Load test

۲-۲-۶- درمان

در حال حاضر داروها در معالجه عفونت ایدز مؤثر نیستند. برای مقابله با عفونت‌ها یا پیشگیری از آنها ممکن است آنتی‌بیوتیک تجویز شود.

داروهای ضد HIV در سه دسته قرار می گیرند:

۱. مهارکننده نوکلئوزیدی ترانس کریپتاز معکوس[۱]: زیدوودین[۲]، لامیوودین[۳]، دیدانوزین[۴]، زالسیتابین[۵] و استاوودین[۶].

۲. مهارکننده غیر نوکلئوزیدی ترانس کریپتاز معکوس[۷]: نویراپین، دلاوردین و افاویرنز.

۳. مهارکننده پروتئاز: ایندیناویر، ریتوناویر، ساکویی ناویر، نلفیناویر و آمپرناویر.

این داروها ممکن است پیشرفت بیماری را به تأخیر اندازند. تجویز آنها باید توسط پزشکان با تجربه انجام پذیرد. با این داروها (مثل زیدووین) می‌توان انتقال عفونت به نوزادان مادران باردار را کاهش داد. تحقیقات زیادی در رابطه با داروهای جدیدتر و واکسن آن از طریق مهندسی ژنتیک در دست انجام هستند.

۲-۳- پارامترهای معرفی خواص دارویی ترکیبات

جهت بیان هر خاصیت و ویژگی نیاز به استانداردهایی است که دارای تعریف کاملا مشخص و شناخته شده باشند تا به وسیله آن استانداردها بتوان ویژگی مورد نظر را در مواد و وسایل مختلف مورد مقایسه قرار داد. داروها نیز از این قاعده مستثنی نیستند و جهت مقایسه خواص دارویی آنها در برابر بیمای‌های مختلف از پارامترهای ویژه‌ای استفاده می‌گردد.

---

[۱] Nucleoside Reverse Transcriptase Inhibitors (NRTIs)

[۲] Zidovudine

[۳] Lamivudine

[۴] Didanosine

[۵] Zalcitabine

[۶] Stavudine

[۷] Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

عمل داروها در بدن انسان را فارماکودینامیک[۱] و آنچه را که بدن با دارو انجام می‌دهد فارماکوکینتیک[۲] می‌گویند. داروهایی که وارد بدن انسان می‌شوند تمایل دارند تا یک گیرنده خاص، کانال‌های یونی، عمل روی یک آنزیم یا پروتئین واسط را تحت تاثیر قرار دهند. در نتیجه آنها موجب می‌شوند که بدن انسان به طریق خاصی واکنش نشان دهدکه این واکنش را پاسخ می‌نامند.

داروها بطور کلی به دو دسته تقسیم می‌شوند:

۱. آگونیست‌ها[۳] مواد شیمیایی که به گیرنده یک سلول متصل شده و آن را تحریک و فعال می‌کنند.

۲. آنتاگونیست‌ها[۴]که از تحریک گیرنده‌ها بوسیله آگونیست‌ها جلوگیری می‌کنند.

ابتدا گیرنده‌ها فعال می‌شوند، سپس گیرنده‌ها یا به وسیله یک پاسخ ویژه مستقیما بر روی بدن اثر می‌گذارند یا موجب آزادسازی هورمون‌ها یا داروهای درونی بدن شده و از این طریق پاسخ ویژه را ایجاد می‌نمایند.

۲-۳-۱- توان دارویی[۵]

در مقوله داروشناسی[۶] توان (قدرت) دارویی میزان اثر بخشی دارو است که به صورت مقدار مورد نیاز آن برای اثر گذاری بر روی شدت داده شده بیان می‌شود. یک داوری بسیار قوی (مانند: مورفین، آلپرازولام و کلروپرومازین) در غلظت‌های خیلی کم پاسخ بزرگتری ایجاد می‌کند در حالیکه داروهای با توان کمتر (مانند: ایبوپروفن و استیل سالیسیلیک اسید) در غلظت‌های پایین پاسخ کوچکتری تولید می‌کنند. پاسخ تولید شده متناسب با میل‌ترکیبی[۷] و درجه‌تاثیردارو[۱] است.

---

[۱] Pharmacodynamics
[۲] Pharmacokinetics
[۳] Agonist
[۴] Antagonist
[۵] Potency
[۶] Pharmacology
[۷] Affinity

میل‌ترکیبی، توانایی دارو در اتصال به گیرنده است و درجه‌تاثیر، ارتباط بین اشغال‌کننده‌ی گیرنده و توانایی آن در شروع پاسخ در سطح مولکولی، سلولی، بافت یا دستگاه می‌باشد. پاسخ همان اثر یا (E) است و به میزان اتصال دارو و گیرنده‌ای که دارو به آن متصل شده وابسته است؛ با این توضیحات، قدرت دارو به میل‌ترکیبی و درجه‌تاثیر دارو وابسته می‌باشد. یک آگونیست، لیگاند، دارو یا هورمونی که به گیرنده متصل می‌شود و ایجاد پاسخ می‌کند را معمولاً بطور مختصر با **A** یا **D** نشان می‌دهند.

پایین‌تر از یک غلظت معین از آگونیست (**[A]**)، پاسخ یا اثر(**E**) برای اندازگیری خیلی کوچک است اما در غلظت‌های بالاتر قابل ارزیابی می‌شود و با افزایش غلظت آگونیست (**[A]**)، افزایش می‌یابد تا به غلظتی برسیم که دیگر با افزایش **[A]**، **E** نتواند افرایش یابد و آن را با E_{max} نشان می‌دهند.

E_{max} حداکثر اثر احتمالی آگونیست است. غلظتی از **A** که در آن **E** برابر ۵۰٪ E_{max} است را "نصف حداکثر غلظت موثر" گویند و به صورت $[A]_{50}$ یا بطور رایج‌تر با EC_{50} نشان می‌دهند. عبارت "توان دارویی" به مقدار $[A]_{50}$ برمی‌گردد. در مقادیر پایین‌تر $[A]_{50}$، غلظت کمتری از دارو مورد نیاز است تا ۵۰٪ از اثر بیشینه دارو را ایجاد نماید.

باید توجه داشت که دوز دارویی بیشتر لزوما به معنای اثرات جانبی بیشتر نیست.

۲-۳-۲- دوز متوسط کشنده[۲]

در سم شناسی، دوز متوسط کشنده، از یک ماده سمی یا پرتو عبارتست از دوز مورد نیاز جهت کشتن نیمی از اعضای جمعیت نمونه در طول مدت یک آزمایش خاص و با علائم LD_{50}[۳]، LC_{50}[۴] یا LCt_{50}[۵] نمایش داده می‌شود. نمودارهای LD_{50} خیلی اوقات به عنوان شاخص برای مواد با سمیت بالا

---

[۱] Efficacy

[۲] Median lethal dose

[۳] Lethal Dose, 50% (LD_{50})

[۴] Lethal Concentration, 50% (LC_{50})

[۵] Lethal Concentration & Time (LCt_{50})

مورد استفاده قرار می‌گیرند. این آزمون بوسیله جی دابلیو تروان[۱] در سال ۱۹۲۷ پیشنهاد شد. هرچند این روش بوسیله برخی آزمایش‌ها مانند روش دوز ثابت به صورت علمی رد می‌شود ولی این مفهوم و محاسبه دوز متوسط کشنده جهت اهداف مقایسه‌ای بطور گسترده مورد بهره‌برداری قرار می‌گیرد.

اندازه‌گیری سمیت، LD_{50}، در برخی مواقع غیر قابل اتکا است و نتایج بدست آمده بوسیله وسایل و ابزار مختلف به دلایلی از قبیل تنوع خصوصیات ژنتیکی جمعیت نمونه، گونه‌های جانوری مورد آزمایش، عوامل محیطی و نحوه اجرا بسیار متغیر می‌باشند. نقطه ضعف دیگر این روش این است که فقط سمیت حاد را اندازگیری می‌کند (در مقابل سمیت مزمن در دوزهای کمتر) و اثرات سمیتی را که منجر به مرگ نمی‌شوند اما با این وجود مهم هستند در بر نمی‌گیرد.

۲-۳-۳- بازدارندگی

نصف حداکثر غلظت بازدارندگی[۲] (**IC_{50}**) میزان تاثیر یک ترکیب در بازدارندگی فعالیت بیولوژیکی یا بیوشیمی است. این کمیت نشان می‌دهد که چه مقدار از یک داروی خاص یا یک ماده دیگر (بازدارنده) نیاز است تا یک فرآیند ویژه بیولوژیکی (یا بخشی از یک فرآیند بیولوژیکی به عنوان مثال آنزیم، سلول، گیرنده سلولی یا میکرواورگانیسم) را به نصف کاهش دهد. به عبارت دیگر نصف حداکثر(۵۰٪) غلظت بازدارندگی (IC) یک ماده (50% IC یا IC_{50}) است. از این کمیت عموما به عنوان معیاری از توان داروهای آنتاگونیست در تحقیقات داروشناسی استفاده می‌شود. بعضی مواقع نیز آن را به تابع لگاریتمی و بصورت تابع pIC_{50} (که برابر با $-\log IC_{50}$ است) تبدیل می‌نمایند، که مقادیر بزرگتر آن به صورت نمایی بیانگر توان بیشتر دارو می‌باشند. بر اساس تعریف FDA[۳]، IC_{50} نماینده غلظتی از دارو است که جهت ۵۰٪ اثر بازدارندگی در محیط کشت مصنوعی مورد نیاز

[۱] J. W. Tervan

[۲] Half maximal inhibitory concentration

[۳] Food and Drug Administration

می‌باشد. این پارامتر با مقدار EC_{50} برای داروهای آگونسیت قابل مقایسه است که نماینده غلظت پلاسمای مورد نیاز برای رسیدن به ۵۰٪ از اثر بیشینه در محیط طبیعی است.

۲-۳-۳-۱- تعیین مقدار IC_{50} برای یک دارو

IC_{50}یک دارو را می‌توان بوسیله رسم منحنی دوز برحسب پاسخ و آزمایش اثر غلظت‌های مختلف از آنتاگونیست روی فعالیت آگونیست بدست آورد. مقادیر IC_{50} را می‌توان با تعیین غلظت مورد نیاز آنتاگونیست مورد نظر جهت بازدارندگی نیمی از پاسخ بیولوژیکی آگونیست مربوطه محاسبه نمود.

مقادیر IC_{50} به شدت به شرایط اندازگیری وابسته هستند. بطور کلی، در غلظت بالاتر بازدارنده، فعالیت آگونیست کاهش بیشتری می‌یابد. مقدار IC_{50} با افزایش غلظت آنزیم افزایش می‌یابد. همچنین با توجه به نوع بازدارنده سایر عوامل نیز ممکن است روی مقدار IC_{50} اثر گذار باشند بعنوان مثال در مورد آنزیم‌های وابسته به ATP، مقدار IC_{50} نیز به غلظت ATP وابسته خواهد بود به ویژه هنگامی که بازدارندگی کاملا رقابتی باشد. مقادیر IC_{50} را می‌توان برای مقایسه توان دو آنتاگونیست بکار برد.

۲-۴- پیشینه کارهای انجام شده در QSAR ترکیبات دارویی

۲-۴-۱- سرطان

- در سال ۲۰۰۴ تارون جها[1] وهمکارانش پس از سنتز دسته‌ای از مشتقات اسید آمینه غیر ضروری اسید گلوتامیک به مطالعه QSAR بروی آنها در زمینه فعالیت ضد تومور پرداختند. پارامترهای فیزیکو-شیمیایی و توپولوژیکی ترکیبالت مورد پژوهش مورد بررسی قرار گرفت و با استفاده از رگرسیون خطی مرحله‌ای بهترین توصیف‌کننده‌ها انتخاب شدند. سپس به کمک رگرسیون خطی چندگانه رابطه بین فعالیت بیولوژیکی و توصیف‌کننده‌های انتخاب شده،

[1] Tarun Jha

بصورت یک معادله خطی ارائه شده و پارامترهای آماری برای مدل خطی بدست آمده، PRESS=۰/۷۹۳، R=۰/۸۲۱ و Q^2=۰/۶۲۵، محاسبه گردید[۱۹].

- در سال ۲۰۰۵ ماریلین ای. موریس[۱] و همکارانش با استفاده از مطالعات QSAR اثر بازدارندگی گروهی از ترکیبات به نام فلاونوئیدها را بر روی پروتئین‌های BCRP[۲] مورد بررسی قرار دادند. BCPRها دسته جدیدی از انتقال دهنده‌های ABC هستند که نقش مهمی را در صورت‌بندی دارو بازی می‌کنند. فلاونوئیدها دسته مهمی از ترکیبات طبیعی هستند که بطور گسترده‌ای در غذاها و محصولات گیاهی حضور دارند و نشان داده شده است که جزء بازدارنده‌های BCRP می‌باشند. در این تحقیق با استفاده از سه توصیف‌کننده ساختاری مدلی خطی برای پیش‌بینی مقادیر pEC_{50} ایجاد شد. نتایج حاصل از این مدل برای سری آموزش (R^2=۰/۸۵۲ و Q^2=۰/۷۸۴) و برای سری تست (R^2=۰/۹۲۲) گویای توانایی مدل ارائه شده در پیش‌گویی مقادیر مورد نظر می‌باشد[۲۰].
- در سال ۲۰۰۶ کاتریتزکی[۳] و همکارانش مقاله‌ای تحت عنوان فعالیت ضد هجوم ترکیبات آلی با استفاده از توصیف‌کننده‌های ساختاری ارائه دادند. هجوم یک معیار اندازگیری فعالیت سلول تومور می‌باشد که نهایتا منجر به سرطان می‌شود. فعالیت ۱۳۹ ترکیب بوسیله شبکه عصبی مصنوعی پیش‌گویی شد. مدل QSAR غیر خطی (ANN) برای ۷۱٪ از ۶۰ ترکیب سری آموزش و ۷۰٪ از ۳۲ ترکیب سری تست بطور صحیح ترکیبات را به دو دسته فعال و غیرفعال کلاسه‌بندی کرده است[۲۱].
- در سال ۲۰۰۷کاترینا ام نیکولیک مدل‌های QSAR را جهت پیدا کردن ارتباط میان خواص فیزیکوشیمیایی α-توکوفرول و مشتقاتش با فعالیت‌های بیولوژیکی آنها اجرا نمود. این ترکیبات سلول‌های سرطان سینه MCF7 انسان را از طریق ناپایدار کردن غشای لیزوزوم و

[۱] Marilyn E. Morris
[۲] Breast Cancer Resistance Protein
[۳] Katritzky

میتوکندری دستخوش مرگ زمانبندی شده می‌کنند. نتایج رگرسیون خطی چندگانه برای دو، سه و چهار توصیف‌کننده محاسبه گردید و مقدار R^2 برای آنها بزرگتر از ۰/۸۳ و پارامتر اعتبارسنجی متقابل (Q^2) بزرگتر از ۰/۶۶ بدست آمد[۲۲].

- در سال ۲۰۰۸ نیز تحقیقات QSAR توسط کی. ام .نیکولیک بر روی خواص دارویی دو دسته از ترکیبات α-توکوفرول در درمان سرطان سینه انجام گرفت. توصیف‌کننده‌های هندسی، فیزیکوشیمیایی و الکترونی این ترکیبات محاسبه گردید و توصیف‌کننده‌هایی که بیشترین ارتباط را با خاصیت مورد نظر (IC_{50}) داشتند انتخاب شدند. سپس مدل رگرسیون خطی چندگانه با چهار توصیف‌کننده ساخته شد که برای آن مقدار R^2 برابر با ۰/۹۸ و مقدار پارامتر اعتبارسنجی متقابل Q^2 برابر با ۰/۹۱ بدست آمد[۲۳].
- در سال ۲۰۰۹ کاترینا نیکولیک و همکارانش تحقیقات QSAR را بر روی ترکیبات α-توکوفرول و کلسترول، در زمینه درمان سرطان سینه ادامه دادند. در این پژوهش توصیف‌کننده‌های ساختاری، هندسی، فیزیکوشیمیایی و الکترونی محاسبه و آنالیز شدند. سپس با انتخاب مناسب‌ترین توصیف‌کننده‌ها، مدل‌های رگرسیون خطی ساخته شدند. بهترین مدل‌ها با سه و چهار توصیف‌کننده دارای مقادیر R^2 بزرگتر از ۰/۹۵ و پارامتر اعتبارسنجی متقابل (Q^2) بزرگتر از ۰/۸۸بودند[۲۴].
- در سال ۲۰۱۰ افشین فصیحی و همکارانش در یک تحقیق QSAR ۴۷ ترکیب از خانواده ایساتین به عنوان داروهای ضد سرطان مورد بررسی قرار دادند. در این پژوهش دو روش رگرسیون خطی چندگانه (MLR) و الگوریتم ژنتیک همراه با حداقل مربعات جزئی (GA-PLS) به کار گرفته شدند. بین دو روش بکار گرفته شده، MLR نتایج بهتری را به همراه پارامترهای آماری بالاتر (Q^2=۰/۹۰ و R^2=۰/۹۲) در پیش‌بینی فعالیت این ترکیبات نشان داد[۲۵].

HIV-1 -۲-۴-۲

- در سال ۱۹۹۷ ترونکت[۱] با استفاده از مطالعات QSAR برهمکنش HETP[۲]را با HIV، با استفاده از شبکه عصبی مصنوعی پیش‌بینی نمود.[۲۶]
- جلالی هروی و همکارانش در سال ۲۰۰۰ در یک مطالعه QSAR، فعالیت ترکیبات ضد ایدز را برای یک گروه بزرگ از مشتقات HETP بررسی کردند. در این کار روش رگرسیون خطی با شبکه عصبی مصنوعی مقایسه شده است که در نتیجه مدل شبکه عصبی مصنوعی به عنوان مدل پیش‌گویی کننده‌ی بهتر نسبت به مدل رگرسیون خطی معرفی گردیده است[۲۷].
- در سال ۲۰۰۲ مولفتا[۳] و همکارانش از روش نیمه تجربی AM1 برای محاسبه ویژگی‌های مولکولی ۱۴ ترکیب بی‌فلاونوئید با خواص ضد HIV-1 استفاده کردند. سپس از سه روش PCA[۴]، HCA[۵] و SDA[۶] برای تعیین ویژگی‌هایی که بیشترین اثر را بر روی فعالیت ضد HIV-1 این ترکیبات دارند استفاده نمودند. این تحقیقات نشان داد که متغیرهای HOMO[۷]، LUMO[۸] و مساحت سطح[۹] بیشترین اثر را بر روی فعالیت ضد HIV-1 دارند.[۲۸].
- در سال ۲۰۰۳ با استفاده از مطالعات QSAR فعالیت ضد ایدز برای مشتقات HETP بررسی شد. در این مطالعه سه روش شامل رگرسیون خطی، شبکه عصبی پس انتشار و بهینه‌سازی تکاملی ادغام شده با شبکه عصبی، مقایسه گردید. در روش بهینه‌سازی تکاملی ادغام شده با شبکه عصبی، جهت بهینه‌سازی وزن‌ها و بایاس‌های شبکه عصبی بوسیله حداقل نمودن خطای خروجی‌ها، از روش تکاملی استفاده شده است. نتایج تأیید می‌کنند که محاسبات

---

[۱] Tronchet

[۲] 1-(2-hydroxyethoxymethyl)-6-(phenylthio)thymine

[۳] F. A. Molfetta

[۴] Principal Component Analysis (PCA)

[۵] Hierarchical Cluster Analysis (HCA)

[۶] Stepwise Discriminant Analysis (SDA)

[۷] Highest Occupied Molecular Orbital (HOMO)

[۸] Lowest Unoccupied Molecular Orbital (LUMO)

[۹] Surface area

تکاملی می‌تواند به عنوان یک روش آموزش شبکه عصبی اجرا شود. همچنین نتایج حاکی از این است که بایاس‌ها در آموزش اولیه و داده‌های تست روی عملکرد محاسبات کامپیوتر می-توانند اثر بگذارند. شبکه بهینه‌شده ساختار ۱-۱۱-۹ دارد که نتایج شبکه با محاسبات تکاملی بهتر از نتایج شبکه پس‌انتشار می‌باشد[۲۹].

- در سال ۲۰۰۴ تارون جها[۱] و همکارانش با استفاده از پارامترهای فیزیکوشیمیایی و توپولوژیکی، به بررسی رابطه ساختار-کمیت مشتقات HEPT پرداختند. این ترکیبات به عنوان بازدارنده‌های ریورز ترانسکریپتاز HIV-1 عمل می‌کنند و می‌توانند در درمان بیماری ایدز مورد استفاده قرار گیرند. آنها به کمک روش MLR، ۱۶ مدل متفاوت را برای پیش‌گویی مقادیر $\log EC_{50}$ برای ۸۴ ترکیب از این خانواده ارائه کردند. در بهترین مدل پارامترهای آماری PRESS، R و R_{cv}^2 به ترتیب ۳۰/۶۲۲، ۰/۹۰۰ و ۰/۷۸۳ گزارش شدند[۳۰].
- در سال ۲۰۰۵ حیات‌شاهی و همکارانش رابطه ساختار-فعالیت را برای بازدارنده‌های رقابتی آدنوزین دآمیناز بوسیله دو مدل غیر خطی رگرسیون لوجستیک[۲] و شبکه عصبی مصنوعی مورد بررسی قرار دادند. آنزیم آدنوزین دآمیناز یا به اختصار ADA یکی از آنزیم‌های مهم در مکانیسم پورین‌هاست. این آنزیم در همه بافت‌های بدن انسان فعالیت دارد، البته بیشترین فعالیت آن در بافت‌های ایمنی از جمله تیموس و لنفوسیت‌هاست. بیش فعالی این آنزیم در بیماری‌های پارکینسون و ایدز دیده شده است. علاوه بر این اختلالات این آنزیم در برخی آنمی‌ها و سرطان‌ها نیز مشاهده شده است. بازدارنده‌های آنزیم آدنوزین دآمیناز علاوه بر کاهش برخی بیماری‌های یاد شده، در سرکوب دستگاه ایمنی بدن در پیوند اعضا برای جلوگیری از دفع آنها نیز اهمیت دارند. مدل‌ها برای طبقه‌بندی ترکیبات به دو کلاس به کار

[۱] Tarun Jha
[۲] logistic

گرفته شد. رگرسیون لوجستیک برای ۹۱/۶۶٪ ترکیبات بدرستی پیش‌بینی کرده بود در حالی که با استفاده از شبکه عصبی ۱-۳-۲ و ۱-۴-۳ این مقدار به ۹۵/۸۳٪ رسید[۳۱].

- در سال ۲۰۰۶ پابلو آر. دوچوویکز[۱] و همکاانش ۱۵۴ ترکیب بازدارنده غیرنوکلئوزیدی ریورز ترانس کریپتاز برای ویروس‌های وحشی HIV-1 و ۵۶ ترکیب بازدارنده ریورز ترانس کریپتاز برای نوع جهش یافته آن (K-103N) را مورد بررسی قرار دادند. توصیف‌کننده‌های مولکولی بوسیله نرم‌افزار DRAGON 5 محاسبه شدند. سپس سه روش رگرسیون مرحله‌ای، روش تعویضی و الگوریتم ژنتیک جهت انتخاب بهترین توصیف‌کننده‌ها مورد استفاده قرار گرفتند. بهترین مدل بدست آمده شامل ۷ توصیف‌کننده بود و مقدار R برای ویروس وحشی و جهش یافته به ترتیب برابر با ۰/۷۹۹۱ و ۰/۹۲۶۱ گزارش شد[۳۲].
- در سال ۲۰۰۷ عرب چم‌جنگلی و همکارانش ارتباط ساختار- فعالیت ترکیبات ضد ایدز را بوسیله شبکه عصبی مصنوعی با الگوریتم آموزشی لونبرگ-مارکورت پیش‌بینی کردند. توصیف‌کننده‌ها بوسیله نرم‌افزار DRAGON محاسبه شد و توصیف‌کننده‌های مهم بوسیله رگرسیون مرحله‌ای انتخاب گردید. ترکیبات بطور تصادفی به دو مجموعه شامل ۳۱ ترکیب سری آموزش و ۱۱ ترکیب سری ارزیابی تقسیم شدند. ارزیابی مدل توسط سری ارزیابی و روش رد مرحله‌ای تک تک انجام گرفت. خطای جذر مجذور میانگین (RMSE) و خطای مطلق میانگین (MAE) برای سری ارزیابی به ترتیب ۰/۰۴۲ و ۰/۰۲۴ می‌باشد. نتایج نشانگر توانایی مدل برای پیش‌بینی داده‌های روابط ساختار- فعالیت داروهای آنتی HIV می‌باشد[۴].
- در سال ۲۰۰۸ گری بی. فوگل[۲] و همکارانش طی یک پژوهش QSAR اثر بازدارندگی مشتقات پیریمتامین را بر روی دی هیدروفولات ریداکتاز (DHFR) مورد بررسی قرار دادند. توصیف‌کننده‌های مولکولی توسط MOE و QikProp استخراج شده و در ساخت مدل‌های

[۱] Pablo R. Duchowicz
[۲] Gary B. Fogel

خطی و غیرخطی به کار گرفته شدند. بهترین مدل بدست آمده شامل توصیف‌کننده‌های حاصل از MOE بود که بوسیله شبکه عصبی مصنوعی با محاسبات تکاملی بدست آمد[۳۳].

- در سال ۲۰۰۹ فن-اِر چن[۱] و همکارش برهمکنش میان HIV-1 وحشی و سه نوع جهش یافته آن را با ۳۴ ترکیب دی‌آریل پیریمیدین بوسیله رگرسیون خطی چندگانه پیش‌بینی نمودند. چهار مدل خطی برای برهمکنش هر یک از انواع ویروس با ترکیبات یاد شده ارائه شد که دارای مقایر R، ۰/۸۲۱۱ (N=۳۴)، ۰/۸۵۹۹ (N=۳۳)، ۰/۸۷۱۱ (N=۳۰) و ۰/۹۰۷۹ (N=۲۹) برای به ترتیب ویروس وحشی و انواع جهش یافته L100I، Y181C و Y188N بودند[۳۴].
- در سال ۲۰۰۹ عرب چم‌جنگلی به منظور توسعه یک مدل غیرخطی جهت پیش‌بینی خواص ضدHIV ۲۷ ترکیب از مشتقات ۱-[(۵-کلروفنیل)سولفونیل]-۱-H-پیرول، شبکه عصبی مصنوعی با الگوریتم لونبرگ-مارکورت را به کار گرفت. وی پس از محاسبه توصیف‌کننده‌های مولکولی، به کمک روش رگرسیون خطی مرحله‌ای از میان تعداد زیادی توصیف‌کننده ۸ توصیف‌کننده را برگزید. در ادامه با تقسیم تصادفی ترکیبات به دوسری آموزش (شامل ۲۰ ترکیب) و ارزیابی (شامل ۷ ترکیب) به بهینه‌سازی پارامترهای شبکه پرداخت.مقادیر MSE و MAE حاصل از مدل بدست آمده برای سری ارزیابی به ترتیب برابر ۰/۰۰۶۷ و ۰/۰۰۶۶ و برای نتایج حاصل از رد مرحله‌ای تک‌تک به ترتیب برابر۰/۰۱۳ و ۰/۰۸۷ بود[۳۵].

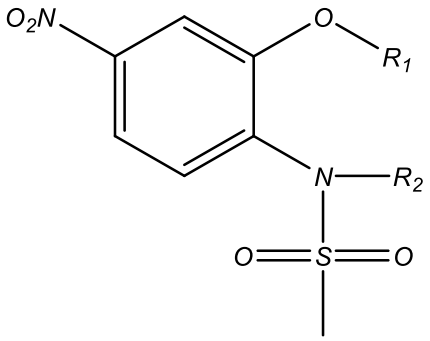
[۱] Fen-Er Chen

فصل سوم

محاسبات و مدل‌سازی QSAR اثر برخی از ترکیبات سولفونانیلید در کاهش فعالیت آروماتاز در بافت‌های سرطانی با استفاده از روش رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی(ANN)

۳-۱- سری داده‌ها

سری داده‌ها شامل ۲۱ ترکیب از مشتقات سولفونانیلید[۱] می‌باشد که از مرجع [۳۶] جمع‌آوری شده است. همانطور که در ساختار پایه این ترکیبات (شکل ۳-۱) نشان داده شده است، ترکیباتی که در این تحقیق مورد بررسی قرار گرفته‌اند دارای دو گروه R_1 و R_2 متفاوت بر روی ساختار اصلی هستند. جزئیات این ترکیبات به همراه مقادیر $\log(1/IC_{50})$ (بر حسب مولار) در جدول ۳-۱ آورده شده است. همانطور که در فصل دوم گفته شد IC_{50} یا غلظت مهارکنندگی نصف بیشینه[۲]، غلظتی از دارو است که برای بازدارندگی ۵۰٪ از محیط کشت مصنوعی مورد نیاز است.



شکل ۳-۱- ساختار اصلی ترکیبات مورد بررسی

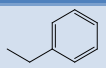
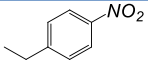
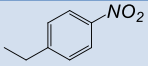
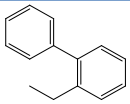
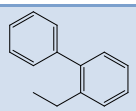
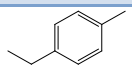
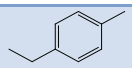
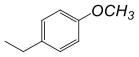
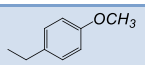
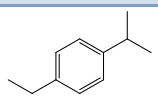
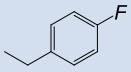
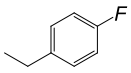
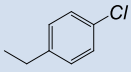
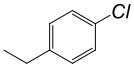
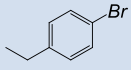
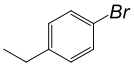
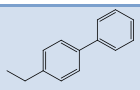
در این تحقیق با توجه به تعداد کم داده‌ها جهت بدست آوردن مدل مناسب از روش ارزیابی تقاطعی[۳] استفاده شده است و سعی شده تا با استفاده از روش یاد شده مدلی ارائه گردد که بتواند با دقت بالایی قدرت بازدارندگی ترکیبات مورد بررسی را پیش‌بینی نماید.

[۱] Sulfonanilide

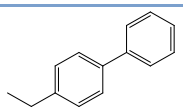
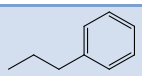
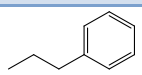
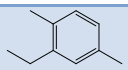
[۲] Half maximal inhibitory concentration

[۳] Cross validation

جدول ۳-۱- ترکیبات مورد بررسی در تحقیق

شماره	R_1	R_2	$\log(1/IC_{50})(M)$
۱		CH_3	۶/۰۹
۲		H	۶/۴۰
۳		CH_3	۶/۳۱
۴		H	۶/۵۹
۵		CH_3	۶/۴۸
۶		H	۵/۷۶
۷		CH_3	۵/۶۳
۸		H	۵/۱۱
۹		CH_3	۶/۳۳
۱۰		H	۵/۱۳
۱۱		H	۶/۲۴
۱۲		CH_3	۵/۷۵
۱۳		H	۶/۵۴
۱۴		CH_3	۵/۶۲
۱۵		H	۵/۲۳
۱۶		CH_3	۶/۴۸
۱۷		H	۶/۰۷

ادامه جدول (۳-۱)

۱۸		CH_3	۶/۱۹
۱۹		H	۶/۶۴
۲۰		CH_3	۶/۲۸
۲۱		H	۶/۴۳

۳-۲- نرم‌افزارهای مورد استفاده

در این تحقیق از نرم‌افزارهای زیر استفاده شده است که در ادامه به طور مختصر به توضیح و بررسی هر یک از آنها خواهیم پرداخت:

- Hyperchem v.8
- Dragon v.2.1
- PASW Statistics 18
- MATLAB v.2008b

۳-۲-۱- Hyperchem

این نرم‌افزار جهت رسم و بهینه‌سازی اولیه ساختار ترکیبات شیمیایی، مورد استفاده قرار می‌گیرد. در این نرم‌افزار ساختار ترکیبات می‌تواند با چهار روش آغازین، نیمه تجربی، مکانیک کوانتومی و تابع دانسیته بهینه‌سازی شود. با توجه به ساختارهای ارائه شده در این پروژه از روش نیمه تجربی AM1 جهت بهینه‌سازی ساختارها استفاده گردید و بهینه‌سازی تا زمانی ادامه یافت که گرادیان انرژی به ۰/۰۰۱ کیلوکالری بر مول برسد.

Dragon -۲-۲-۳

از این نرم‌افزار جهت استخراج توصیف‌کننده‌های مولکولی استفاده می‌شود. نرم‌افزار Dragon توسط گروه QSAR و کمومتریکس میلانو [۳۷] طراحی شده است. این نرم‌افزار قادر به محاسبه بیش از ۱۴۰۰ توصیف‌کننده می‌باشد که به ۱۸ دسته اصلی تقسیم می‌شوند که در جدول ۳-۲ نام این ۱۸ دسته اصلی آورده شده است.

جدول ۳-۲- توصیف کننده های محاسبه شده توسط نرم‌افزار Dragon

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

علاوه بر محاسبه ساده‌ترین نوع اتم‌ها، گروه‌های عاملی و شمارش اجزا، می‌توان تعداد زیادی از توصیف‌کننده‌های توپولوژیکی و هندسی را نیز توسط این نرم‌افزار محاسبه نمود. برای اجرای این نرم‌افزار به فایل‌های ساختار مولکولی ایجاد شده به وسیله سایر نرم‌افزارهای مدل‌سازی مولکولی مانند

---

۱ Constitutional descriptors
۲ Geometrical descriptors
۳ Topological descriptors
۴ 3D-MoRSE descriptors
۵ Burden eigenvalues
۶ Charge indices
۷ 2D autocorrelation
۸ Functional group descriptors
۹ Charge descriptors
۱۰ Atom-centered fragments
۱۱ Aromaticities indices
۱۲ Randic molecular profiles
۱۳ Molecular properties

Hyperchem نیاز می‌باشد. اکثر فایل‌های مولکولی با فرمت‌های معمول، در این نرم‌افزار قابل قبول هستند. برای استفاده بهتر از محاسبات این نرم‌افزار باید بهینه‌سازی سه بعدی ساختارها با هیدروژن‌هایشان به کاررود. Dragon به عنوان یک نرم‌افزار QSAR طراحی نشده است، بلکه تنها توصیف‌کننده‌های مولکولی را نمایش می‌دهد و هیچ‌گونه آنالیز QSAR را انجام نمی‌دهد. با این حال، Dragon یک فایل خروجی کامل را که به سادگی توسط هر نرم‌افزار آنالیز همبستگی قابل کاربرد است، فراهم می‌سازد. اطلاعات کامل در مورد توصیف‌کننده‌های به کار رفته در این نرم‌افزار (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[۳۸].

PASW Statistics -۳-۲-۳

SPSS یا بسته‌ی آماری برای علوم اجتماعی[۱] نام نرم‌افزاری است که برای تحلیل‌های آماری به کار می‌رود و نخستین بار در سال ۱۹۶۸ ارائه و منتشر شد. پس از ۲۸ جولای ۲۰۰۹ که شرکت سازنده این نرم‌افزار توسط IBM خریداری شد این نرم‌افزار با نام PASW[۲] است منتشر گردید.

SPSS ازجمله نرم‌افزارهایی است که برای تحلیل‌های آماری در علوم اجتماعی، به صورت بسیار گسترده‌ای استفاده می‌شود. این نرم‌افزار توسط پژوهشگران بازار و داد و ستد، پژوهشگران سلامتی، شرکت‌های نقشه‌برداری، دولت‌ها، دانشجویان، سازمان‌های بازاریابی و غیره به کار می‌رود. افزون بر تحلیل‌های آماری، مدیریت داده‌ها و مستندسازی داده‌ها نیز از ویژگی‌های این نرم‌افزار هستند.

SPSS دارای توانایی‌های بسیار بالایی است که به تعدادی از آنها اشاره می‌شود[۳۹]:

- 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، درصدفراوانی و...

[۱] Statistical Package for Social Science

[۲] Predictive Analytics SoftWare

- انواع توزیع‌های آماری شامل توزیع‌های گسسته و پیوسته
- تهیه انواع طرح‌های آماری
- انجام آنالیز واریانس یکطرفه، دوطرفه، چندطرفه وآنالیز 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 تحلیل داده ها با استفاده از SPSS، سه مرحله‌ی اساسی وجود دارد. نخست باید داده‌های خام وارد شده در یک پرونده ذخیره گردند. سپس باید تحلیل مورد نیاز تعیین و انتخاب شود و در مرحله سوم برونداد حاصله وارسی می‌گردد.

در این تحقیق از این نرم‌افزار جهت کاهش توصیف‌کننده‌ها و مدل‌سازی MLR استفاده گردید.

MATLAB -۴-۲-۳

نرم‌افزار MATLAB برنامه‌ای کامپیوتری است که برای کسانی که با محاسبات عددی و به ویژه خطی سر و کار دارند تهیه شده است. نام این نرم‌افزار از عبارت انگلیسی MATrix LABoratory اقتباس شده و هدف اولیه آن قادر ساختن مهندسین و دانشمندان به حل مسایل شامل عملیات ماتریسی بدون نیاز به نوشتن برنامه در زبان‌های برنامه نویسی متداول نظیر فورترن[۱] و سی[۲] بود. با گذشت زمان قابلیت‌های بیشتری به این نرم‌افزار افزوده شده است. به طوری که در حال حاضر

[۱] FORTRAN
[۲] C

MATLAB به ابزار پرقدرتی برای ترسیم داده‌ها، برنامه نویسی و انجام محاسبات مهندسی و پژوهشی تبدیل شده است[۴۰].

در این پروژه از این نرم‌افزار جهت انجام محاسبات رگرسیون خطی چندگانه و شبکه عصبی مصنوعی استفاده شده است و تمام برنامه‌های استفاده شده در آن توسط نگارنده در محیط این نرم‌افزار نوشته شده است.

۳-۳- انتخاب توصیف‌کننده‌های مهم یا کاهش متغیرهای مستقل

از آنجا که تعداد توصیف‌کننده‌هایی که توسط نرم‌افزار دراگون ارائه می‌شود زیاد است، باید به دنبال روشی جهت کاهش تعداد توصیف‌کننده‌ها بود تا از انجام محاسبات وقت‌گیر و بیهوده جلوگیری کرده و از پیچیدگی کار پرهیز شود. روش مورد استفاده باید قادر باشد تا توصیف‌کننده‌هایی که بیشترین ارتباط را با متغیر وابسته (در اینجا $\log(1/IC_{50})$) دارند انتخاب کرده و سایر توصیف‌کننده‌ها را حذف نماید. این کار در سه مرحله انجام شد که در ادامه به شرح مراحل انجام گرفته پرداخته می‌شود:

- مرحله اول: در مرحله اول توصیف‌کننده‌هایی که دارای مقادیر یکسانی برای حداقل ۹۰٪ از ترکیبات سری داده‌ها بودند حذف شدند. این کار با استفاده از نرم‌افزار MATLAB انجام شد. بدین منظور برنامه‌ای کوتاه در نرم‌افزار یاد شده نوشته شد که قادر بود داده‌های مربوط به هر ستون از توصیف‌کننده‌ها را با هم مقایسه کرده و اگر تعداد داده‌های یکسان یک ستون بیش از ۹۰٪ تعداد کل داده‌های آن ستون را تشکیل می‌داد آن ستون را حذف نماید. استفاده از این برنامه موجب می‌شود تا از خطای انسانی جلوگیری شده و در زمان نیز به مقدار زیادی صرفه‌جویی شود.
- مرحله دوم: با توجه به اینکه نبایستی متغیرهای ورودی در روش رگرسیون خطی چندگانه با یکدیگر همبستگی داشته باشند، در این مرحله همبستگی میان متغیرها مورد بررسی قرار

گرفت و از میان دو متغیر که دارای ضریب همبستگی بالای ۰/۹ بودند متغیری که دارای همبستگی کمتری با متغیر وابسته بود، حذف گردید. برای انجام این فرآیند از نرم‌افزار MATLAB استفاده شد، بدین ترتیب که برنامه‌ای در این نرم‌افزار نوشته شد که ابتدا ضریب همبستگی را برای متغیرهای مستقل محاسبه می‌نمود. سپس اگر دو متغیر دارای همبستگی بالای ۰/۹ بودند، همبستگی هر دو متغیر با متغیر وابسته را محاسبه کرده و متغیر دارای ضریب همبستگی کوچکتری را حذف می‌کرد.

- مرحله سوم: در پایان جهت بدست آوردن توصیف‌کننده‌هایی که بیشترین ارتباط را با مقادیر $\log(1/IC_{50})$ داشتند، با استفاده از نرم‌افزار SPSS برای تمام مولکول‌ها محاسبات رگرسیون مرحله‌ای بر روی تمام توصیف‌کننده‌های باقیمانده از مراحل قبل انجام گردید. در روش رگرسیون مرحله‌ای ابتدا متغیری که بالاترین میزان همبستگی را با متغیر وابسته دارد وارد مدل می‌شود، سپس با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله مورد بررسی قرار می‌گیرند و اگر متغیری سطح معنی داری خود را از دست بدهد قبل از ورود متغیر جدید از مدل کنار گذاشته می‌شود. بدین منظور مقادیر $\log(1/IC_{50})$ بعنوان متغیر وابسته و داده‌های توصیف‌کننده‌ها بعنوان متغیر مستقل به نرم‌افزار SPSS داده شدند و از میان تمام توصیف‌کننده‌های ارائه شده، تنها ۲۰ توصیف‌کننده بعنوان توصیف‌کننده‌های مهم انتخاب شدند. این توصیف‌کننده‌ها به همراه کلاس مربوطه در جدول ۳-۳ آورده شده‌اند.

جدول ۳-۳- کل توصیف‌کننده‌های انتخاب شده به همراه کلاس آنها

ردیف	نشانه*	کلاس	توضیحات
۱	R6m[۴]	GETAWAY	R autocorrelation of lag 6
۲	BELm7	Burden eigenvalues	Lowest eigenvalue n. 7 of Burden matrix
۳	L2m[۴]	WHIM	2nd component size directional WHIM index
۴	Mor15e[۱]	3D-MoRSE	3D-MoRSE signal 15
۵	RDF035u[۵]	RDF	Radial Distribution Function-3.5
۶	GGI3	Topological	Topological charge index of order 3
۷	Mor31v[۳]	3D-MoRSE	3D-MoRSE signal 31
۸	Mor27m[۴]	3D-MoRSE	3D-MoRSE signal 27
۹	Mor16m[۴]	3D-MoRSE	3D-MoRSE signal 16
۱۰	MATS7p[۲]	2D autocorrelation	Moran autocorrelation - lag 7
۱۱	RDF020e[۱]	RDF	Radial Distribution Function-2.0
۱۲	Mor15m[۴]	3D-MoRSE	3D-MoRSE signal 15
۱۳	ATS8m[۴]	2D autocorrelation	Broto-Moreau autocorrelation of a topological structure - lag 8
۱۴	Mor19p[۲]	3D-MoRSE	3D-MoRSE signal 19
۱۵	Mor21e[۱]	3D-MoRSE	3D-MoRSE signal 21
۱۶	BEHe8	Burden eigenvalues	Highest eigenvalue n. 8 of Burden matrix
۱۷	HATS8p[۲]	GETAWAY	Leverage-weighted autocorrelation of lag 8
۱۸	Hy	Molecular properties	Hydrophilic factor
۱۹	piPC04	Walk and path counts	Molecular multiple path count of order 4
۲۰	BELv3	Burden eigenvalues	Lowest eigenvalue n. 3 of Burden matrix

*معنی حروف قرار گرفته در انتهای علامت توصیف‌کننده‌ها در این ستون:

۱. e: وزن‌دار شده به وسیله الکترونگاتیویته سندرسون

۲. p: وزن‌دار شده به وسیله قطبش‌پذیری اتمی

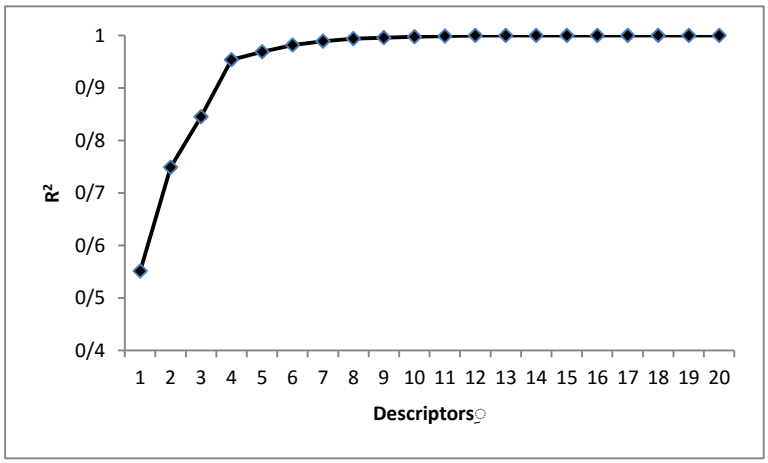
۳. v: وزن‌دار شده به وسیله حجم واندروالس اتمی

۴. m: وزن‌دار شده به وسیله جرم اتمی

u: با هیچ ویژگی اتمی وزن‌دار نشده است.

۳-۴- انتخاب مدل مناسب

در این پروژه جهت انتخاب بهترین مدل از میان مدل‌های بدست آمده، مقدار R^2 محاسبه شده برای مدل‌های مختلف و همچنین تعداد توصیف‌کننده‌های موجود در هر مدل به عنوان معیار در نظر گرفته شده است. بدین ترتیب مدلی که دارای کمترین تعداد توصیف‌کننده و بیشترین مقدار R^2 باشد انتخاب شده است. همان طور که در شکل ۳-۲ مشاهده می‌شود بعد از مدل ششم روند تغییرات R^2 تقریباً ثابت است. با توجه به این موضوع و نیز کم بودن تعداد ترکیبات سری آموزش، مدل شماره ۶ که شامل ۶ توصیف‌کننده است انتخاب و در مدل‌سازی به کار گرفته شد. در جدول ۳-۴ توصیف‌کننده‌های انتخاب شده به همراه کلاس مربوط به آنها نشان داده شده‌اند. همچنین ماتریس همبستگی این توصیف‌کننده‌ها در جدول ۳-۵ آورده شده است. همانطور که مشاهده می‌شود همبستگی معناداری میان این توصیف‌کننده‌ها وجود ندارد.



شکل ۳-۲-نمودار مقادیر R^2 بر حسب تعداد توصیف‌کننده‌ها

جدول ۳-۴- توصیف‌کننده‌های مورد استفاده در مدل‌سازی شبکه عصبی مصنوعی به همراه کلاس مربوطه

ردیف	علامت اختصاری	کلاس	توضیحات
۱	R6m	GETAWAY	R autocorrelation of lag 6
۲	BELm7	Burden eigenvalues	Lowest eigenvalue n. 7 of Burden matrix
۳	L2m	WHIM	2nd component size directional WHIM index
۴	Mor15e	3D-MoRSE	3D-MoRSE signal 15
۵	RDF035u	RDF	Radial Distribution Function-3.5
۶	GGI3	Topological	Topological charge index of order 3

جدول ۳-۵- ماتریس همبستگی برای توصیف‌کننده‌های انتخاب شده

	R6m	BELm7	L2m	Mor15e	RDF035u	GGI3
R6m	۱					
BELm7	-۰/۴۲۴	۱				
L2m	-۰/۶۲۲	۰/۱۰۶	۱			
Mor15e	-۰/۲۹۳	۰/۰۲۰	۰/۰۱۶	۱		
RDF035u	-۰/۳۵۰	۰/۵۵۱	۰/۰۸۶	-۰/۲۳۷	۱	
GGI3	-۰/۶۹۰	۰/۵۴۷	۰/۴۵۱	۰/۱۲۹	۰/۶۹۸	۱

۳-۵- فرآیند مدل‌سازی QSAR

جهت بدست آوردن مدل‌های مناسب برای توضیح رابطه ساختار-فعالیت، ابتدا ترکیبات به دو گروه آموزش و تست تفکیک شدند. سری تست شامل ۵ ترکیب از ۲۱ ترکیب سری داده‌ها بود که بصورت تصادفی به نحوی انتخاب شدند که تمام طیف ترکیبات و تغییرات $\log(1/IC_{50})$ را پوشش دهند. شماره این ترکیبات بر اساس جدول ۳-۱ عبارتند از ۶، ۱۰، ۱۲، ۱۷ و ۲۱. ترکیبات باقیمانده نیز سری آموزش را تشکیل دادند.

۳-۵-۱- مدل رگرسیون خطی چندگانه (MLR)

به منظور اندازه‌گیری قدرت پیش‌بینی هر یک از مدل‌ها با استفاده از توصیف‌کننده‌های مربوط به هر یک از آنها و به کمک برنامه‌ای که در نرم‌افزار MATLAB نوشته شده بود ضرایب رگرسیون محاسبه و به روش ارزیابی تقاطعی مدل مورد ارزیابی قرار گرفت.

در روش ارزیابی تقاطعی با توجه به اینکه سری ارزیابی وجود ندارد از خود سری آموزش جهت ارزیابی مدل مورد بررسی، استفاده می‌گردد. این روش مشابه روش رد مرحله‌ای تک‌تک[۱] است، بدین ترتیب که ابتدا یکی از مولکول‌های سری آموزش حذف شده و ضرایب رگرسیون با استفاده از ترکیبات باقیمانده محاسبه می‌شوند. سپس با استفاده از ضرایب بدست آمده مقدار $\log(1/\mathrm{IC}_{50})$ برای مولکول کنار گذاشته شده محاسبه می‌گردد. این کار برای تمام مولکول‌های سری آموزش انجام می‌گیرد و با استفاده از نتایج بدست آمده پارامترهای آماری مورد نیاز جهت ارزیابی مدل، محاسبه می‌گردند. سپس این پارامترهای بدست آمده برای مدل‌های متفاوت با هم مورد مقایسه قرار می‌گیرند و مدلی که دارای توانایی پیش‌بینی بالاتری باشد انتخاب می‌شود. در این پروژه روش ازریابی تقاطعی برای ساختن مدل‌هایی حاوی دو تاشش توصیف‌کننده به کار گرفته شد. مدل‌های ایجاد شده به همراه روابط مربوطه در زیر آورده شده‌اند.

$$\log\left(\frac{1}{\mathrm{IC}_{50}}\right) = 17.523 - 28.092\mathrm{R6m} - 8.4922\mathrm{BELm7} \qquad (۱-۳)$$

$$\log\left(\frac{1}{\mathrm{IC}_{50}}\right) = 18.709 - 31.85\mathrm{R6m} - 8.5793\mathrm{BELm7} - 0.1689\mathrm{L2m} \qquad (۲-۳)$$

$$\log\left(\frac{1}{\mathrm{IC}_{50}}\right) = 18.402 - 32.071\mathrm{R6m} - 7.4151\mathrm{BELm7} - 0.2607\mathrm{L2m} - 0.63735\mathrm{Mor15e} \qquad (۳-۳)$$

[۱] Leave- One –Out (LOO)

$$\log\left(\frac{1}{IC_{50}}\right) = 17.743 - 34.07R6m - 5.5712BELm7 - 0.3442L2m - 0.9498Mor15e - 0.0242RDF035u \quad (۴-۳)$$

$$\log\left(\frac{1}{IC_{50}}\right) = 17.948 - 34.614R6m - 5.282BELm7 - 0.433L2m - 1.139Mor15e - 0.057RDF035u + 0.541GGI3 \quad (۵-۳)$$

مقادیر میانگین مربعات خطا (MSE) و ضریب همبستگی (R^2) حاصل از ارزیابی تقاطعی برای مدل‌های ارائه شده در بالا محاسبه شده و در جدول ۳-۶ آورده شده‌اند.

جدول ۳-۶- مقادیر MSE و R^2 بدست آمده برای ارزیابی تقاطعی سری آموزش از مدل‌های ۲ تا ۶

مدل	MSE	F	R^2
۲	۰/۰۳۵	۲۶/۸۲۷	۰/۸۴
۳	۰/۰۳۳	۳۰/۷۸۲	۰/۸۵
۴	۰/۰۲۴	۸۲/۲۳۳	۰/۹۰
۵	۰/۰۲۱	۹۴/۳۶۳	۰/۹۱
۶	۰/۰۱۲	۱۲۷/۹۶۱	۰/۹۵

براساس نتایج بدست آمده مدل شماره ۶ که دارای کمترین مقدار MSE و بیشترین مقدار R^2 می‌باشد به عنوان بهترین مدل انتخاب گردید.

۳-۵-۲- مدل‌سازی با استفاده از شبکه عصبی مصنوعی(ANN)

در این پروژه، شبکه عصبی مصنوعی برای مدل‌سازی غیرخطی جهت توصیف رفتار سری داده‌ها مورد استفاده قرار گرفت. در این تحقیق برنامه‌نویسی و اجرای یک شبکه عصبی پیش خور سه‌لایه با الگوی آموزش پس‌انتشار در محیط نرم‌افزار MATLAB انجام شد. در این روش تابع کارایی شبکه، میانگین مربعات خطای استاندارد (MSE) می‌باشد. برای بدست آوردن بهترین مدل و کمترین خطا، پارامترهای شبکه (تابع آموزش، تابع انتقال، تعداد متغیرهای ورودی شبکه، تعداد گره‌ها در لایه مخفی وتعداد دورهای آموزش) بهینه‌سازی شدند.

۳-۵-۲-۱- بهینه‌سازی پارامترهای شبکه

برای بهینه‌سازی پارامترهای شبکه و شناسایی بهترین پارامترهای آن، شبکه‌هایی با ورودی‌های از ۲تا ۶ توصیف‌کننده ایجاد شدند. هرشبکه با الگوریتم آموزشی لونبرگ-مارکورت و تنظیم بایزین با تعداد گره‌های متفاوت از ۲ تا ۱۰ گره در لایه پنهان آموزش داده شد. این در حالی بود که برای بدست آوردن بهترین تابع انتقال در لایه‌ی پنهان مدل‌های شبکه عصبی مصنوعی طراحی شده، از توابع لگاریتم زیگموئید و تانژانت هایپربولیک به عنوان تابع فعالیت استفاده گردید و تعداد دورهای آموزش نیز از ۲ تا ۳۰ تغییر داده شد. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطای استاندارد (MSE) به عنوان معیار انتخاب شد. با توجه به اینکه خروجی‌های شبکه داده‌های کمّی $\log(1/IC_{50})$ می‌باشند، لذا تابع انتقال خطی برای لایه خروجی برگزیده شد.

برای رسیدن به این هدف همانطور که قبلا گفته شد سری داده‌ها به دو سری آموزش و تست تقسیم‌بندی شد. در اینجا نیز مانند روش MLR، با استفاده از سری آموزش، از روش ارزیابی تقاطعی برای بدست آوردن بهترین مدل بر اساس کمترین مقدار مجذور میانگین خطای استاندارد استفاده شد. در این روش در هر بار، داده‌های مربوط به یکی از ساختارها حذف شد و شبکه بوسیله اطلاعات ساختارهای باقیمانده آموزش داده شد؛ پس از آموزش، شبکه برای پیش‌بینی مقدار مورد نظر (در اینجا $\log(1/IC_{50})$) برای ترکیب حذف شده، بکار برده شد. این کار برای هر شبکه منحصر بفرد و برای تک تک ساختارهای سری آموزش انجام گرفت. در پایان با استفاده از مقادیر پیش‌بینی شده توسط شبکه و مقادیر تجربی، مقدار MSE برای هر شبکه محاسبه شده و جهت مقایسه شبکه‌های مورد بررسی مورد استفاده قرار گرفت و مدل با کمترین MSE به عنوان مدل مناسب انتخاب گردید.

۳-۵-۲-۲- انتخاب تعداد متغیرهای ورودی شبکه، الگوریتم آموزشی، نوع تابع تبدیل، تعداد گره‌های لایه پنهان و تعداد دورهای آموزش

در این پروژه بهینه‌سازی و انتخاب بهترین تعداد توصیف‌کننده‌ها و الگوریتم آموزشی، نوع تابع تبدیل، تعداد گره‌ها یا نرون‌های لایه پنهان و تعداد دورهای آموزش به طور همزمان انجام گرفته است. این کار موجب می‌شود بتوانیم نقش تمام پارامترها را به طور همزمان و یک‌جا بررسی نماییم. جهت نیل به این هدف به شبکه‌ای کاملا تکرارپذیر نیازمندیم که بتواند به طور کاملا خودکار پارامترهای شبکه را تغییر داده و مقادیر MSE را، که معیار انتخاب بهترین شبکه در این پروژه می‌باشد، برای هر شبکه خاص محاسبه نماید. از سوی دیگر با توجه به اینکه در این تحقیق از روش ارزیابی تقاطعی استفاده شده است هر شبکه خاص برای رسیدن به مقدار MSE باید ۱۶ بار اجرا شود. بنابراین می‌بایست الگوریتمی برای شبکه نوشته می‌شد که تکرارپذیر باشد؛ بدین معنی که الگوریتم شبکه برای یک شبکه خاص با پارامترهای ثابت، هرچند بار، بدون نیاز به خارج شدن از محیط نرم‌افزار MATLAB، اجرا گردد نتایج یکسانی ارائه دهد. جهت نشان دادن اهمیت تکرارپذیری شبکه کافی است به تعداد شبکه‌هایی که در این‌جا مورد بررسی قرار گرفته‌اند نگاهی بیاندازیم.

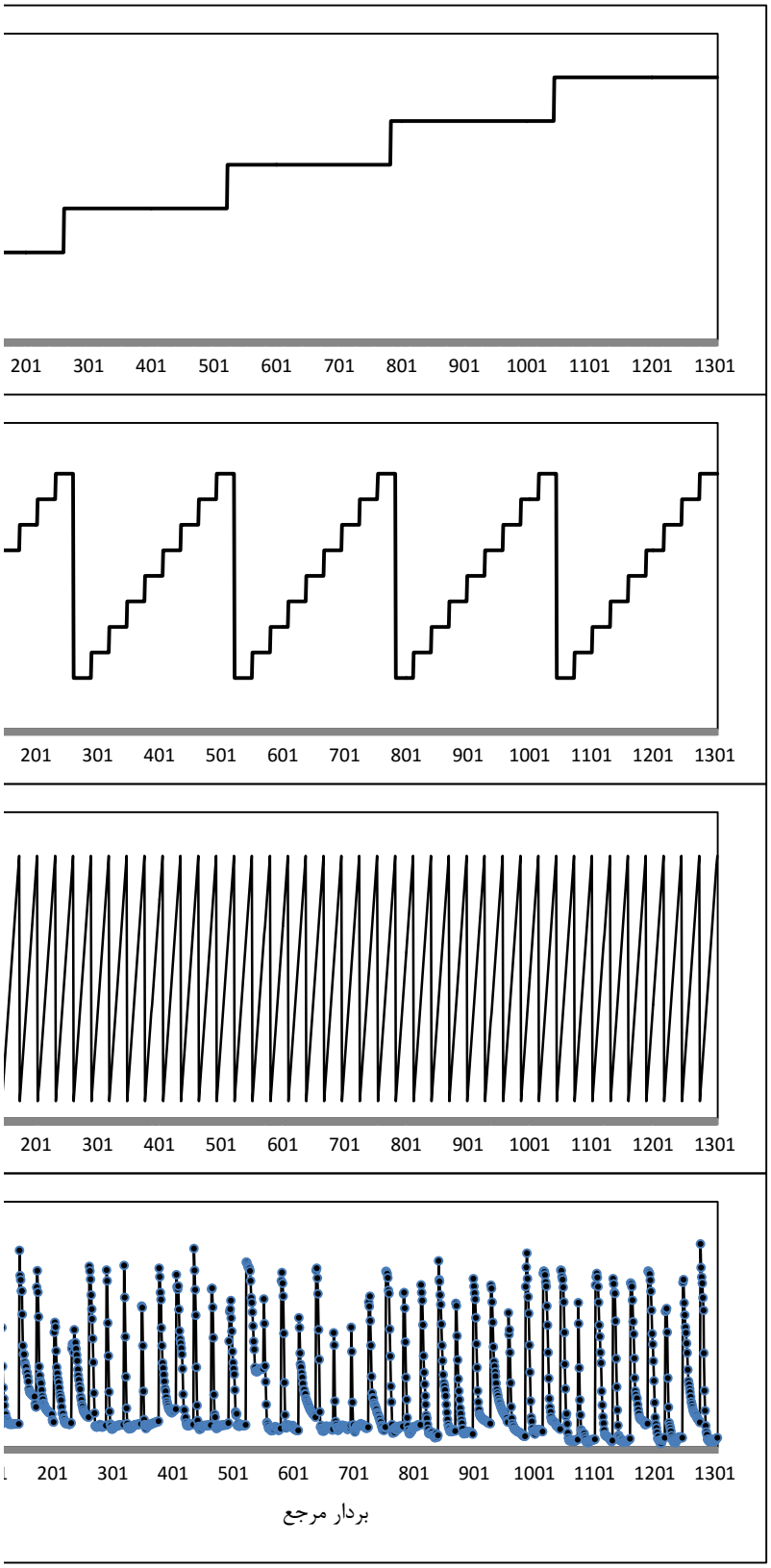
با توجه به پارامترهایی که باید برای شبکه بهینه می‌شد جهت بهینه‌سازی همزمان تمام پارامترها در مجموع ۵۲۲۰ شبکه مورد ارزیابی قرار گرفته است که در زیر به تعداد موارد مورد بررسی اشاره شده است.

۵۲۲۰=(دورهای آموزش)۲۹×(گره)۹×(تابع تبدیل)۲×(تابع آموزش)۲×(توصیف‌کننده)۵

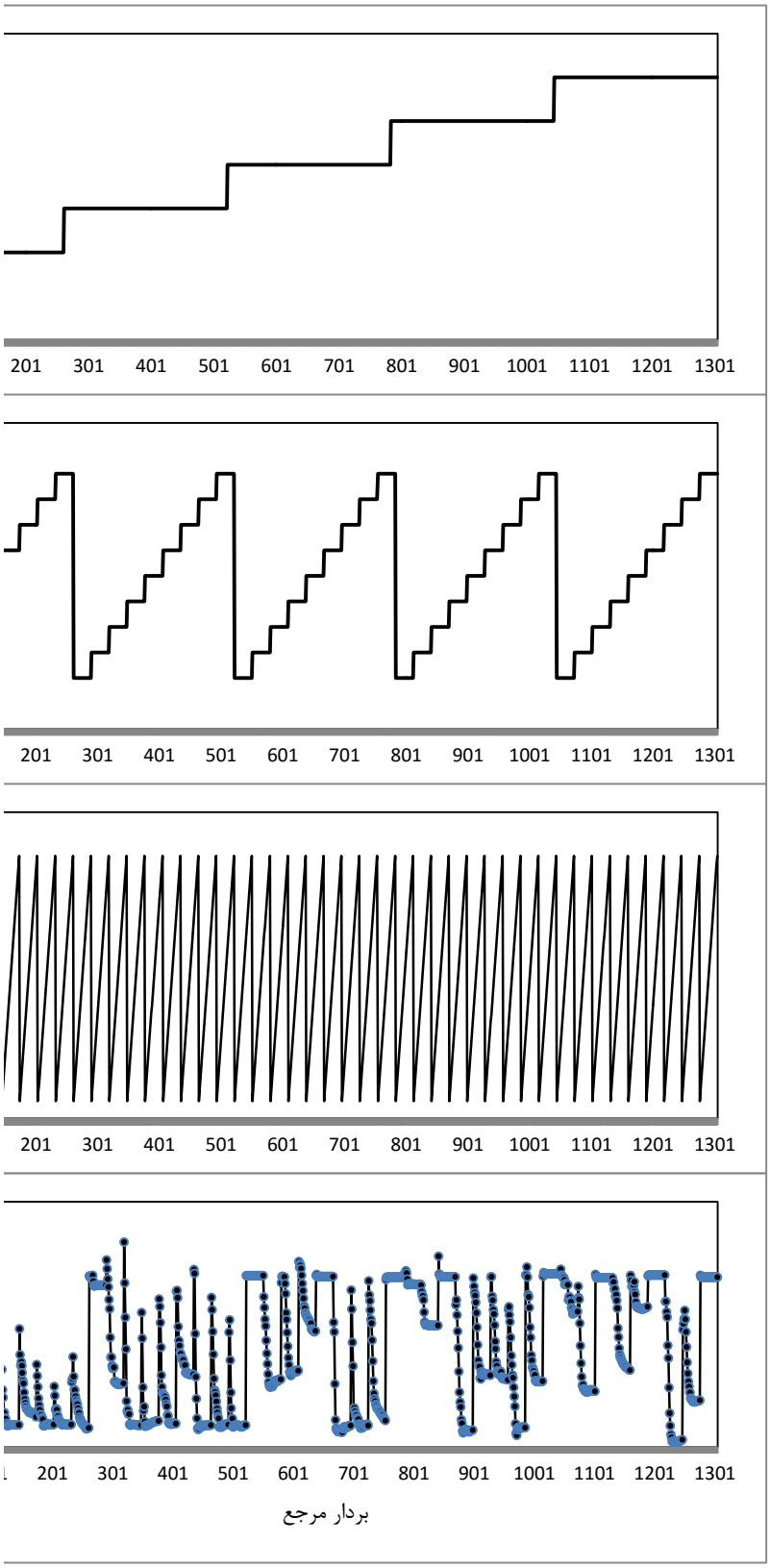
همچنین با در نظر گرفتن این مطلب که هریک از ۵۲۲۰ شبکه، درروش ارزیابی تقاطعی ۱۶ بار تکرار شده اند، تعداد دفعات اجرای الگوریتم شبکه‌ها به ۸۳۵۲۰ مرتبه می‌رسد. بنابراین خودکار شدن شبکه برای بدست آوردن نتایج مربوط به هر شبکه منحصر بفرد می‌تواند به میزان زیادی زمان مورد نیاز جهت بهینه‌سازی شبکه را کاهش دهد.

بخشی از روند تغییرات پارامترهای شبکه در حین بهینه‌سازی همزمان پارامترها به همراه مقادیر MSE به دست آمده به روش ارزیابی تقاطعی به صورت نموداری بر حسب یک بردار مرجع فرضی در شکل‌های ۳-۳، ۳-۴، ۳-۵ و ۳-۶ آمده است.

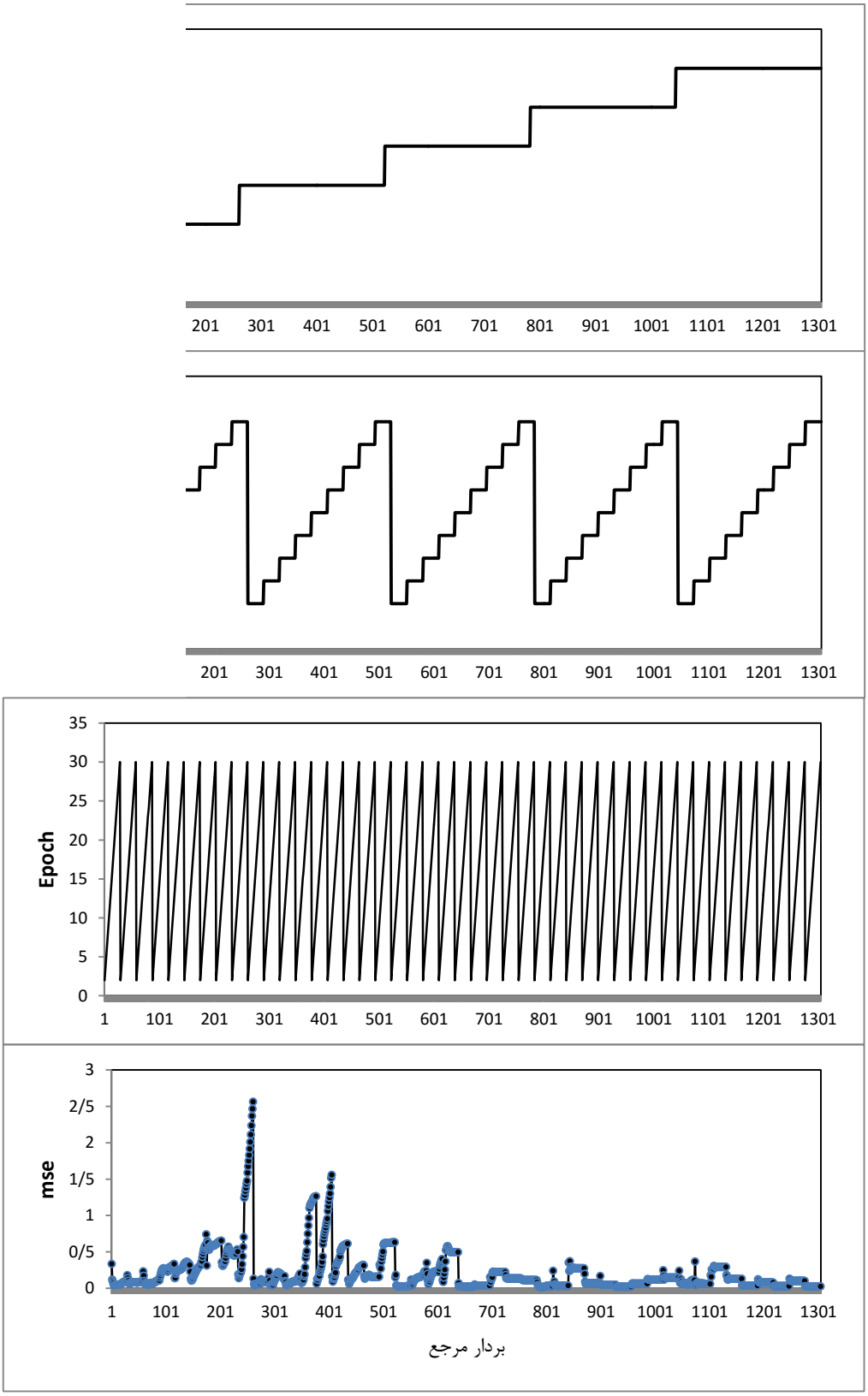
با توجه به نتایج حاصله برای تعداد مختلف توصیف‌کننده‌ها و گره‌ها و دورهای آموزش و همچنین توابع متفاوت آموزش و تبدیل، بهترین شبکه‌های به دست آمده بر اساس کمترین مقدار MSE در جدول ۳-۷ خلاصه شده‌اند. با توجه به جدول ۳-۷ شبکه عصبی سه‌لایه با ۶ توصیف‌کننده در لایه ورودی، هفت نرون در لایه مخفی و تعداد دور آموزشی ۲۳ با تابع آموزشی تنظیم بایزین و تابع تبدیل تانژانت هایپربولیک کمترین MSE را نسبت به سایر شبکه‌ها نشان می‌دهد. بنابراین شبکه‌ای با این ساختار به عنوان بهترین مدل ANN برای مدل‌سازی داده‌های بازدارندگی آروماتاز ترکیبات سولفونانیلید در نظر گرفته شد.



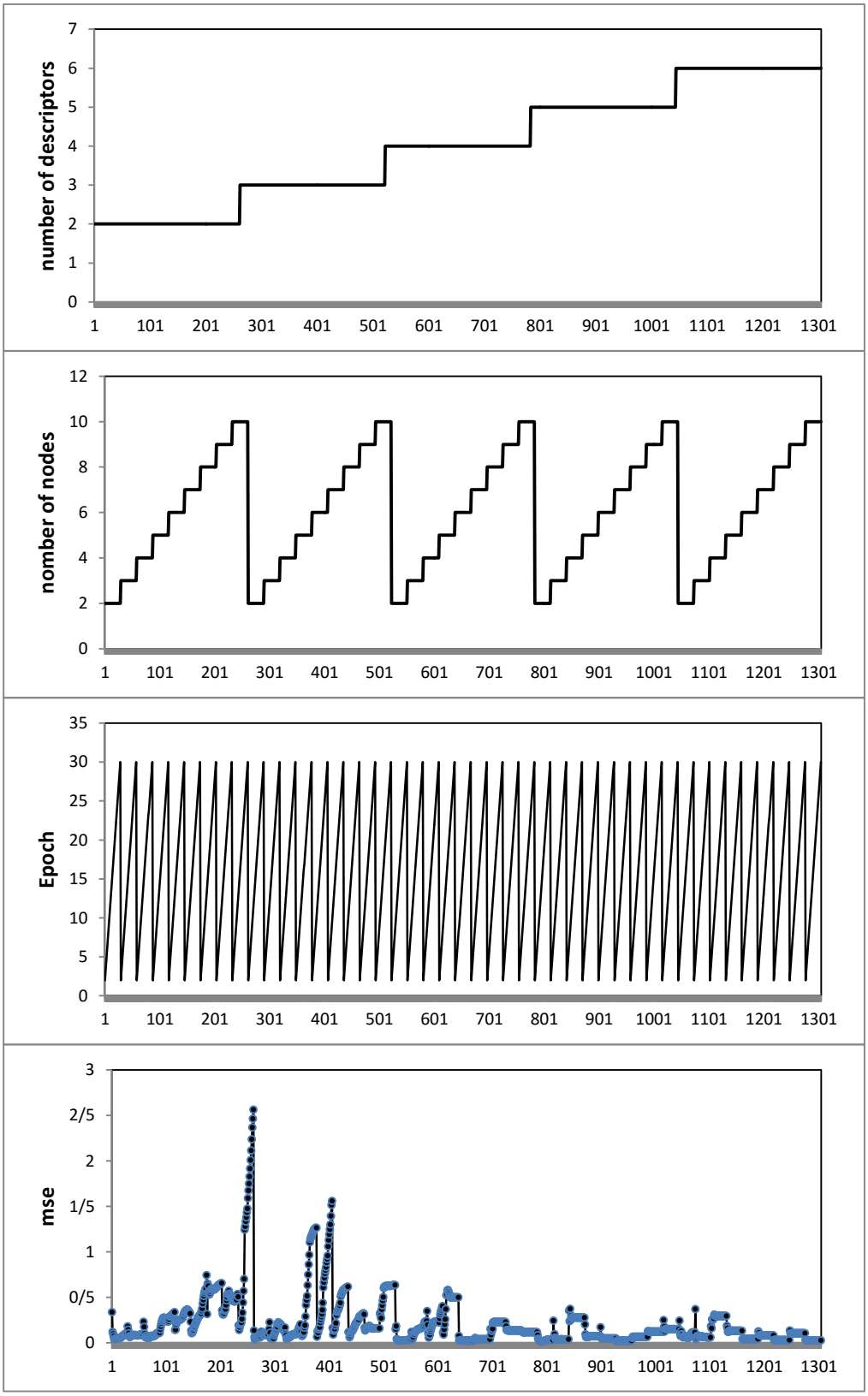
شکل ۳-۳- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین به همراه تابع تبدیل تانژانت هایپربولیک



شکل ۳-۴- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین به همراه تابع تبدیل لگاریتم زیگموئید



شکل ۳-۵- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش الگوریتم لونبرگ-مارکورت به همراه تابع تبدیل تانژانت هایپربولیک



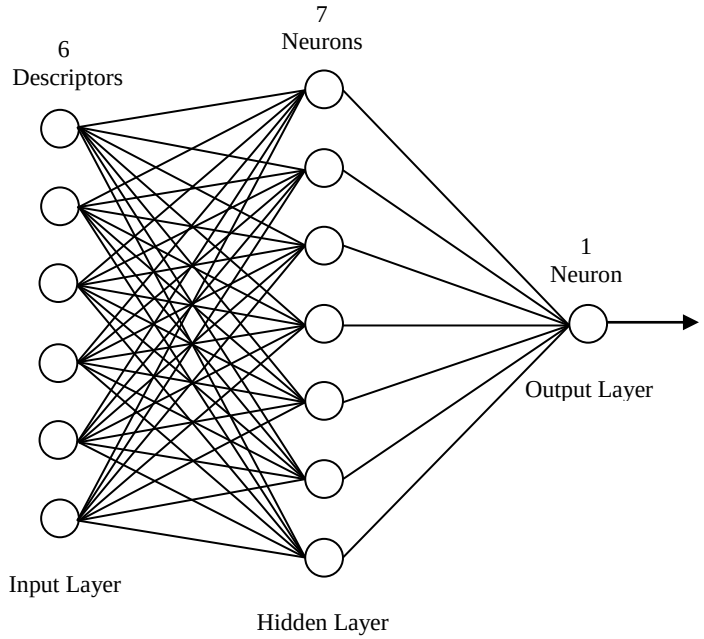
شکل ۳-۶- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش الگوریتم لونبرگ مارکورت به همراه تابع تبدیل لگاریتم زیگموئید

جدول ۳-۷- پارامترهای شبکه‌های بهینه بدست آمده بر اساس مقادیر میانگین مربعات خطای استاندارد (MSE)

تعداد توصیف‌کننده‌ها	تابع آموزش	تابع تبدیل	تعداد گره‌ها (نرون‌ها)	تعداد دورهای آموزش	MSE
۶	تنظیم بایزین	تانژانت هایپربولیک	۷	۲۳	۰/۰۰۵
۶	تنظیم بایزین	لگاریتم زیگموئید	۸	۲۳	۰/۰۰۷
۵	لونبرگ-مارکورت	تانژانت هایپربولیک	۲	۱۰	۰/۰۱۴
۶	لونبرگ-مارکورت	لگاریتم زیگموئید	۱۰	۴	۰/۰۱۲

۳-۵-۳- معماری شبکه عصبی مصنوعی بهینه شده

با توجه به نتایج حاصل در مرحله بهینه‌سازی، همان‌طور که قبلا گفته شد، شبکه عصبی پیش‌خور سه‌لایه با الگوی آموزش پس انتشار با الگوریتم تنظیم بایزین و تابع انتقال تانژانت هایپربولیک انتخاب گردید. ساختار هندسی شبکه بهینه شده در شکل ۳-۷ نشان داده شده است.



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

۳-۶- ارزیابی و مقایسه مدل‌ها

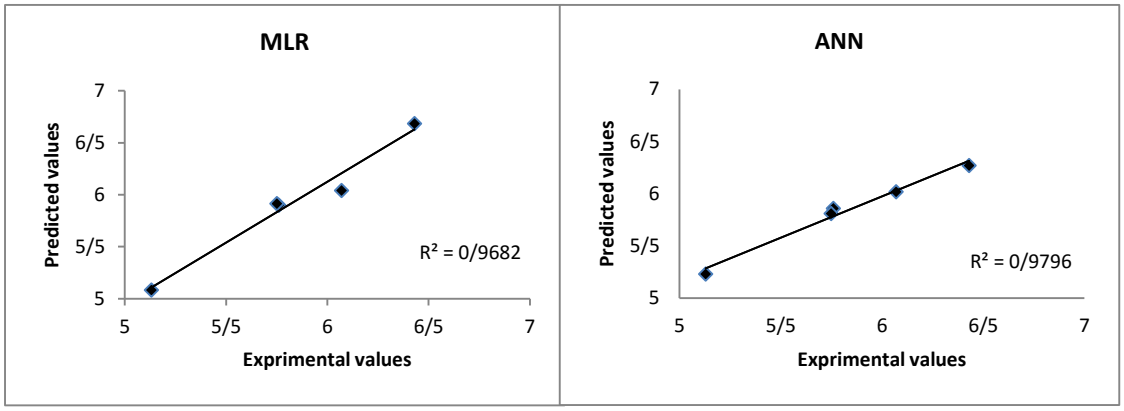
اساسی‌ترین مرحله در مطالعات QSAR ارزیابی قدرت پیش‌بینی مدل ارائه شده می‌باشد. قدرت پیش‌بینی مدل، به صورت توانایی آن در ارائه خروجی‌های رضایت‌بخش برای آن دسته از مولکول‌هایی که در سری آموزش وجود نداشته‌اند، تعریف می‌شود. جهت ارزیابی توانایی بهترین مدل‌های ارائه شده (که دارای کمترین مقدار MSE حاصل از ارزیابی تقاطعی سری آموزش بودند) در پیش‌بینی داده‌های $\log(1/IC_{50})$، در اینجا چندین روش به کار گرفته شده است.

۳-۶-۱- ارزیابی مدل‌های به دست آمده با استفاده از سری تست

پس از مدل‌سازی به روش‌های خطی و غیرخطی، جهت ارزیابی مدل‌های به دست آمده با استفاده از نتایج حاصل از ارزیابی تقاطعی سری آموزش و بررسی قدرت تعمیم این مدل‌ها و جلوگیری از خطای برازش و همچنین مقایسه دو روش، قدرت پیش‌بینی مدل‌ها با استفاده از ۵ ترکیب که در مرحله آموزش از آنها استفاده نشده است و از آنها به نام سری تست یاد می‌شود مورد بررسی قرار گرفت. مقادیر محاسبه شده $\log(1/IC_{50})$ برای سری تست به روش رگرسیون خطی چندگانه با استفاده از رابطه (۳-۵) و همچنین شبکه عصبی بهینه به همراه مقادیر خطای نسبی آنها در جدول ۳-۸ آمده است. در شکل ۳-۸ نیز نمودار پراکندگی این مقادیر رسم شده است.

جدول ۳-۸- نتایج حاصل از مدل رگرسیون خطی چندگانه و شبکه عصبی برای سری تست

شماره	مقدار واقعی	مقدار پیش‌بینی		خطای نسبی (٪)	
		MLR	ANN	MLR	ANN
۶	۵/۷۶	۵/۹۰	۵/۸۶	۲/۴۳	۱/۷۳
۱۰	۵/۱۳	۵/۰۸	۵/۲۳	-۰/۹۷	۱/۹۵
۱۲	۵/۷۵	۵/۹۱	۵/۸۱	۲/۷۸	۱/۰۴
۱۷	۶/۰۷	۶/۰۴	۶/۰۲	-۰/۴۹	-۰/۸۲
۲۱	۶/۴۳	۶/۶۸	۶/۲۷	۳/۸۹	-۲/۴۹



الف) ب)

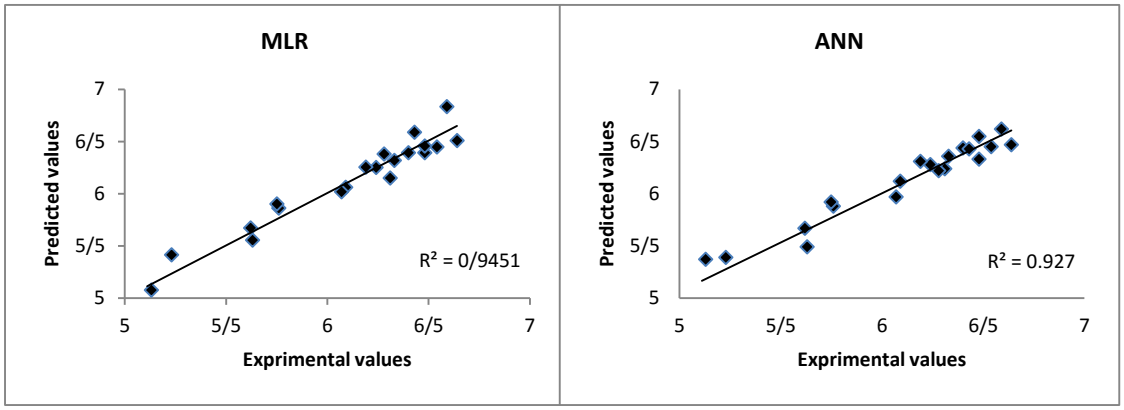
شکل ۳-۸- نمودار مقادیر پیش‌بینی شده $\log(1/IC_{50})$ توسط الف) روش MLR و ب) شبکه عصبی مصنوعی بهینه بر حسب مقادیر تجربی برای سری تست

۳-۶-۱-۱- ارزیابی مدل‌ها به روش حذف مرحله‌ای تک‌تک

به منظور ارزیابی مدل‌های به دست آمده علاوه بر استفاده از سری تست، روش حذف مرحله‌ای تک‌تک داده‌ها نیز به‌کار گرفته شد. در این روش برای هر یک از مدل‌ها هر بار یکی از ملکول‌ها از سری داده‌ها حذف و سپس مدل‌سازی با ۲۰ ملکول باقیمانده صورت گرفت و مدل بدست آمده برای پیش‌بینی $\log(1/IC_{50})$ ملکول حذف شده به کار گرفته شد. در جدول ۳-۹ مقادیر پیش‌بینی‌شده به همراه مقادیر تجربی در کنار مقادیر خطای نسبی آمده است. نمودار پراکندگی مقادیر محاسبه شده بر حسب مقادیر تجربی نیز در شکل ۳-۹ نشان داده شده است. نزدیک بودن نتایج تجربی به نتایج به دست آمده از مدل‌های منتخب نشان دهنده صحت مدل‌ها می‌باشد.

جدول ۳-۹- نتایج حاصل از ارزیابی مدل رگرسیون خطی چندگانه و شبکه عصبی مصنوعی با استفاده از روش رد مرحله‌ای تک‌تک

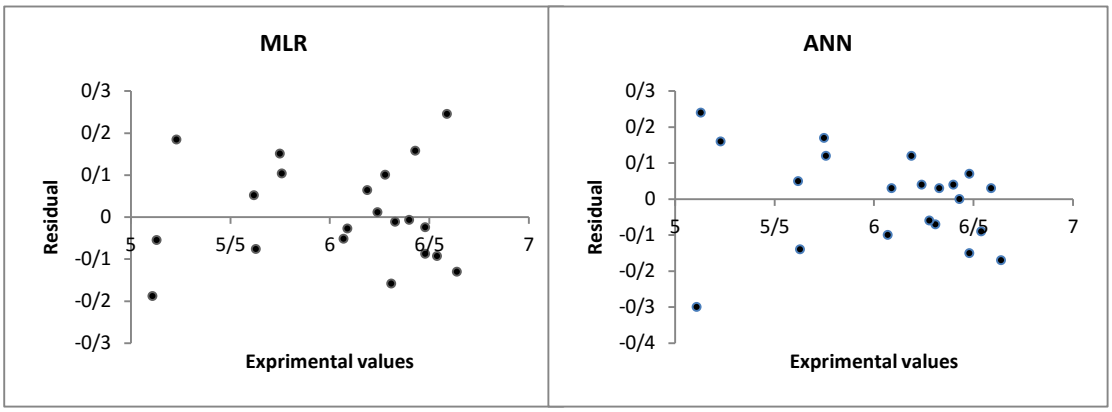
شماره	مقدار واقعی	مقدار پیش‌بینی شده		خطای نسبی (٪)	
		MLR	ANN	MLR	ANN
۱	۶/۰۹	۶/۰۶	۶/۱۲	-۰/۴۹	۰/۴۹
۲	۶/۰۴	۶/۳۹	۶/۴۴	۵/۷۹	۶/۶
۳	۶/۳۱	۶/۱۵	۶/۲۴	-۲/۵۴	-۱/۱۱
۴	۶/۵۹	۶/۸۴	۶/۶۲	۳/۷۹	۰/۴۵
۵	۶/۴۸	۶/۳۹	۶/۳۳	-۱/۳۹	-۲/۳۱
۶	۵/۷۶	۵/۸۶	۵/۸۸	۱/۷۴	۲/۰۸
۷	۵/۶۳	۵/۵۵	۵/۴۹	-۱/۴۲	-۲/۴۸
۸	۵/۱۱	۴/۹۲	۴/۸۱	-۳/۷۲	-۵/۸۷
۹	۶/۳۳	۶/۳۲	۶/۳۶	-۰/۱۶	۰/۴۷
۱۰	۵/۱۳	۵/۰۸	۵/۳۷	-۰/۹۷	۴/۶۸
۱۱	۶/۲۴	۶/۲۵	۶/۲۸	۰/۱۶	۰/۶۴
۱۲	۵/۷۵	۵/۹۰	۵/۹۲	۲/۶۱	۲/۹۶
۱۳	۶/۵۴	۶/۴۵	۶/۴۵	-۱/۳۸	-۱/۳۸
۱۴	۵/۶۲	۵/۶۷	۵/۶۷	۰/۸۹	۰/۸۹
۱۵	۵/۲۳	۵/۴۱	۵/۳۹	۳/۴۴	۳/۰۶
۱۶	۶/۴۸	۶/۴۶	۶/۵۵	-۰/۳۱	۱/۰۸
۱۷	۶/۰۷	۶/۰۲	۵/۹۷	-۰/۸۲	-۱/۶۵
۱۸	۶/۱۹	۶/۲۵	۶/۳۱	۰/۹۷	۱/۹۴
۱۹	۶/۶۴	۶/۵۱	۶/۴۷	-۱/۹۶	-۲/۵۶
۲۰	۶/۲۸	۶/۳۸	۶/۲۲	۱/۵۹	-۰/۹۶
۲۱	۶/۴۳	۶/۵۹	۶/۴۳	۲/۴۵	۰/۰۰



الف ب

شکل ۳-۹- نمودار مقادیر پیش‌بینی شده $\log(1/IC_{50})$ با استفاده از الف) مدل رگرسیون خطی چندگانه و ب) شبکه عصبی مصنوعی بهینه برحسب مقادیر واقعی به روش رد مرحله‌ای تک‌تک

همچنین نمودار مقادیر باقیمانده محاسبه شده در ارزیابی مدل MLR و شبکه عصبی بر حسب مقادیر تجربی $\log(1/IC_{50})$ نیز در شکل ۳-۱۰ رسم شد. تقارن نسبی داده‌ها حول خط صفر نشان‌دهنده عدم وجود خطای معین در هر یک از مدل‌های ارائه شده می‌باشد.



الف ب

شکل ۳-۱۰- نمودار مقادیر خطای مطلق در ارزیابی مدل الف) رگرسیون خطی چندگانه و ب) شبکه عصبی مصنوعی بهینه به روش رد مرحله‌ای تک‌تک در مقابل مقادیر تجربی $\log(1/IC_{50})$

۳-۶-۱-۲- ارزیابی مدل‌های ارائه شده با استفاده از پارامترهای آماری

برای بررسی میزان کارآیی مدل رگرسیون خطی چندگانه ارائه شده و شبکه عصبی مصنوعی بهینه چهار پارامتر آماری MSE، MAE[۱]، REP[۲] و R^2[۳] بوسیله روابط زیر محاسبه گردید که نتایج بدست آمده در جدول ۳-۱۰ آورده شده است.

$$MAE = \frac{\sum_{i=1}^{n}\left|\frac{X_i - Y_i}{X_i}\right|}{n} \times 100 \qquad (۶-۳)$$

$$REP(\%) = \frac{100}{\bar{X}} \times \sqrt{\frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{n}} \qquad (۷-۳)$$

$$R^2 = \left(\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right)\right)^2 \qquad (۸-۳)$$

در روابط فوق X_i مقدار تجربی بازدارندگی ترکیب iام، $\bar{X}$ میانگین کل مقادیر تجربی بازدارندگی، Y_i مقدار پیش‌بینی شده‌ی بازدارندگی iامین ترکیب، $\bar{Y}_i$ میانگین کل مقادیر پیش‌بینی شده، s_X انحراف استاندارد مقادیر تجربی بازدارندگی، s_Y انحراف استاندارد مقادیر پیش‌بینی شده و n تعداد ترکیبات مورد بررسی می‌باشند.

جدول ۳-۱۰- پارامترهای آماری محاسبه شده برای مدل رگرسیون خطی چندگانه و شبکه عصبی مصنوعی

پارامتر	سری آموزش (n=۱۶) (ارزیابی تقاطعی)		سری تست (n=۵)		کل داده‌ها (n=۲۱)	
	MLR	ANN	MLR	ANN	MLR	ANN
MSE	۰/۰۱۲۱	۰/۰۰۴۷	۰/۰۲۲۷	۰/۰۰۹۸	۰/۰۱۳۲	۰/۰۱۶۱
MAE	۰/۰۸۶۵	۰/۰۵۸۴	۰/۱۲۶۷	۰/۰۹۰۹	۰/۰۹۴۲	۰/۱۰۳۱
REP	۱/۷۹۲۸	۱/۱۲۰۲	۲/۵۸۴	۱/۶۹۵۷	۱/۸۹۲۴	۲/۰۹۵۵
R^2	۰/۹۴۷۲	۰/۹۷۸۷	۰/۹۶۸۲	۰/۹۸۱۱	۰/۹۴۵۱	۰/۹۲۸۳

[۱] Mean Absolute Error

[۲] Relative Error of Prediction

[۳] Coefficient of determination

۳-۶-۱-۳- آزمون Y- تصادفی[۱]

به منظور اجتناب از همبستگی‌های تصادفی و تضمین قدرت شبکه عصبی مصنوعی ارائه شده، آزمون Y- تصادفی انجام شد. در این مرحله، با استفاده از کدی که در نرم‌افزار MATLAB نوشته و اجرا شد، مقادیر پاسخ به صورت تصادفی در محدوده پاسخ‌ها (مقادیر تجربی $\log(1/IC_{50})$) ایجاد گردید. سپس مدل شبکه عصبی بهینه با استفاده از ماتریس متغیرهای مستقل و مقادیر تصادفی پاسخ توسعه یافت. در نهایت با توجه به پاسخ شبکه، همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از شاخص R^2 مورد بررسی قرار گرفت. این کار ده بار تکرار شد. در جدول ۳-۱۱ مقادیر R^2 برای سری آموزش و تست آورده شده است. مقادیر پایین R^2 نشان‌دهنده عدم وجود همبستگی شانسی در مدل به دست آمده از شبکه هستند.

جدول ۳-۱۱- مقادیر R^2 برای سری‌های آموزش و تست پس از چندین آزمون Y- تصادفی

تعداد دفعات تکرار آزمون	R^2 سری آموزش(n=۱۶) (ارزیابی تقاطعی)	R^2 سری تست (n=۵)
۱	۰/۳۲	۰/۱۲
۲	۰/۰۶	۰/۴۲
۳	۰/۱۴	۰/۰۱
۴	۰/۴۵	۰/۳۳
۵	۰/۱۷	۰/۱۰
۶	۰/۰۶	۰/۶۶
۷	۰/۱۸	۰/۰۲
۸	۰/۰۶	۰/۰۹
۹	۰/۰۱	۰/۰۱
۱۰	۰/۱۰	۰/۰۵

[۱] Y- radomization

۳-۶-۲- نتیجه‌گیری نهایی

با توجه به نتایج حاصل از رد مرحله‌ای تک‌تک برای مدل‌های به دست آمده می‌توان گفت که هر دو مدل از قدرت پیش‌بینی بالایی برای مقادیر بازدارندگی ترکیبات سولفونانیلید می‌باشند. اما با توجه به نتایج حاصل از سری تست که شامل ترکیباتی است که در مرحله آموزش از آنها استفاده نشده است و در فرایند مدل‌سازی شرکت نداشته‌اند شبکه عصبی نتایج بسیار بهتری ارائه می‌دهد که با در نظر گرفتن این مطلب که توانایی مدل‌های QSAR در پیش‌بینی مقادیر مورد نظر در مواجهه با ترکیباتی است که قبلا آنها را ندیده است، می‌توان گفت شبکه عصبی در پیش‌بینی میزان بازدارندگی برای ترکیباتی که قبلا آنها را ندیده است موفق‌تر از مدل رگرسیون خطی عمل می‌نماید و توانمندتر می‌باشد.

فصل چهارم

محاسبات و مدل‌سازی QSAR اثر بازدارندگی آنتاگونیست‌های CCR5 در درمان HIV-1

۴-۱- مقدمه

CCR5[۱] پروتئینی است که در بدن انسان بوسیله ژن CCR5 کدگذاری شده است. CCR5 عضوی از خانواده گیرنده‌های بتا کموکین[۲] است که عضو ثابت غشای سلول‌ها هستند. ویروس HIV اغلب با استفاده از CCR5 یا CXCR4[۳] وارد سلول هدف می‌شود. در مراحل اولیه عفونت، در سلول‌هایی که بوسیله HIV مورد تهاجم قرار گرفته‌اند، ویروس‌ها با استفاده از CCR5 به صورت گونه‌های غالبا جدا شده[۴] درمی‌آیند. پیشنهاد شده است که این ویروس‌ها ممکن است در طول انتقال و یا مرحله حاد بیماری دارای مزیت گزینش‌پذیری باشند. همچنین در طول مراحل تهاجم، حداقل نیمی از سلول‌هایی که مورد تهاجم قرار گرفته‌اند، فقط ویروس‌هایی را که از CCR5 استفاده می‌کنند تولید می‌نمایند. تعدادی از داروهای آزمایشی جدید HIV، که تحت عنوان بازدارنده‌های ورود نامیده می‌شوند، برای دخالت در برهمکنش میان CCR5 و HIV طراحی شده‌اند. یکی از این دسته داروها مشتقات ۵-اکسوپیرولیدین-۳-کربوکسامید هستند که بعنوان آنتاگونیست CCR5 عمل می‌نمایند و از تحریک این گیرنده‌ها توسط ویروس‌ها ممانعت بعمل می‌آورند.

در این تحقیق اثر بازدارندگی این مشتقات مورد بررسی قرار گفته و سعی شده است تا به کمک مطالعات QSAR مدلی با توانایی پیش‌بینی توان بازدارندگی ترکیبات یاد شده ارائه گردد تا بتوان از این مدل در طراحی و ساخت ترکیبات دارویی جدید از این دسته استفاده نمود.

۴-۲- نرم‌افزارهای مورد استفاده

همانطور که در فصل گذشته ذکر شد در این پروژه نیز از نرم‌افزار Hyperchem 8 جهت رسم و بهینه‌سازی ساختار ترکیبات به روش نیمه تجربی AM1 استفاده گردید. جهت محاسبه‌ی

---

[۱] C-C chemokine receptor type 5 (CCR5)

[۲] Beta chemokine receptors

[۳] C-X-C chemokine receptor type 4

[۴] Acute isolated species

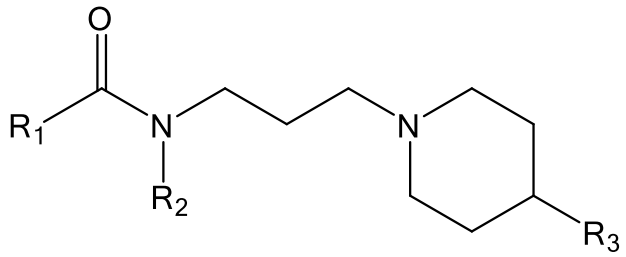
توصیف‌کننده‌های مولکولی نرم‌افزار Dragon 2.1 که توسط یک گروه تحقیقاتی ایتالیایی نوشته شده است، به کار گرفته شد. بدین منظور هریک از ساختارها بصورت تکی به این نرم‌افزار داده شد و برای هریک از آنها تمامی توصیف‌کننده‌ها محاسبه گردید. به کمک نرم‌افزار SPAW 18 و به روش رگرسیون خطی مرحله‌ای توصیف‌کننده‌های مولکولی مناسب انتخاب و مدل‌های آماری ایجاد گردید. برنامه‌ها و الگوریتم شبکه عصبی مصنوعی که در این پروژه به کار گرفته شدند نیز در محیط نرم‌افزار MATLAB 7.7 (2008b) نوشته و اجرا شده‌اند.

۴-۳- محاسبات و مدل‌سازی

در این پروژه فعالیت دارویی (IC_{50}) ۱۰۴ ترکیب از مشتقات ۵-اکسوپیرولیدین-۳-کربوکسامید که (در شرایط عملی یکسانی به دست آمده بودند) از مراجع [۴۱، ۴۲ و ۴۳] گردآوری شده‌اند، مورد بررسی قرار گرفت. در مرحله اول ساختار مولکولی ترکیبات رسم و بهینه شد، سپس مدل مناسب به کمک شبکه عصبی مصنوعی طراحی شد.

۴-۳-۱- سری داده ها

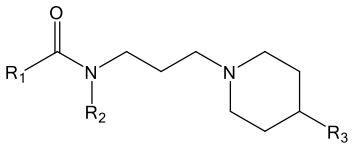
مقادیر منفی لگاریتم برای داده‌های IC_{50} (بر حسب مولار) که از مراجع بالا گردآوری شده‌اند، به عنوان متغیر وابسته در مدل‌سازی مورد استفاده قرار گرفت. ساختار ترکیبات مورد بررسی به همراه مقادیر $\log(1/IC_{50})$ آنها (که از این به بعد با pIC_{50} شناخته می‌شوند) در جداول ۴-۱، ۴-۲ و ۴-۳ نشان داده شده است. جهت تسهیل در تشخیص ساختار ترکیبات مورد بررسی، این ترکیبات در سه جدول متفاوت ارائه شده‌اند؛ اما با توجه به اینکه این ترکیبات از یک خانواده هستند دارای ساختار کلی شکل ۴-۱می‌باشند.



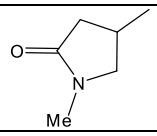
شکل ۴-۱- ساختار پایه مشتقات ۵-اکسوپیرولیدین-۳-کربوکسامید

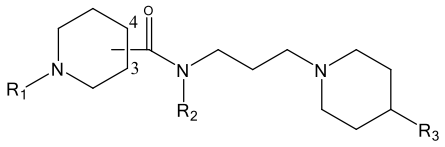
۴-۳-۲- رسم و بهینه‌سازی ساختار مولکول‌ها

در مرحله دوم از انجام مطالعه QSAR، باید ساختار ترکیبات مورد بررسی رسم و بهینه‌سازی شوند. جهت انجام این کار از نرم‌افزار Hyperchem 8 استفاده گردید. ساختارها ابتدا در این نرم‌افزار رسم شدند و سپس ساختار نهایی به روش نیمه تجربی AM1 بهینه گردید. بهینه‌سازی تا رسیدن به گرادیان انرژی ۰/۰۰۱ کیلوکالری بر مول ادامه یافت. روش AM1 در کنار روش PM3 یکی از دقیق‌ترین روش‌های بهینه‌سازی ساختارهای مولکولی است. با توجه به اینکه روش PM3 بخوبی قادر به بهینه‌سازی برهمکنش‌های بین اتم‌ها و پیوندهای رزونانسی و توزیع بار در ساختار مولکول نمی‌باشد اطلاعات حاصل از بهینه‌سازی به وسیله این روش در مورد مکانیک میدان‌های نیرو در مولکول غیر قابل اعتماد هستند. به همین دلیل در این پروژه از روش AM1 که در این زمینه عملکرد بهتری ارائه می‌نماید استفاده شده است.



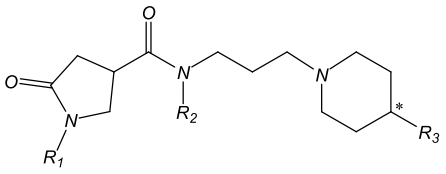
جدول ۴-۱-جزئیات ساختاری ترکیبات ۱ تا ۳۰

ردیف	R_1	R_2	R_3	$log(1/IC_{50})$
1		Ph	$PhCH_2$	۶/۳۲
۲	PhNH	Ph	$PhCH_2$	۷/۷۴
۳	$PhCH_2NH$	Ph	$PhCH_2$	۶/۸۲
۴	*c*-HexNH	Ph	$PhCH_2$	۷/۶۲
۵	*n*-PrNH	Ph	$PhCH_2$	۷/۲۱
۶	PhN(Me)	Ph	$PhCH_2$	۵/۱۸
۷	Ph	Ph	$PhCH_2$	۵/۶۴
۸	$PhCH_2$	Ph	$PhCH_2$	۶/۳۵
۹	$Ph(CH_2)_2$	Ph	$PhCH_2$	۶/۰۴
۱۰	$PhCH_2O$	Ph	$PhCH_2$	۵/۷۰
۱۱	(4-Cl-Ph)NH	Ph	$PhCH_2$	۸/۲۳
۱۲	(4-Cl-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۸/۱۱
۱۳	(4-F-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۷/۸۹
۱۴	(4-Br-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۷/۷۴
۱۵	(4-Me-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۸/۱۸
۱۶	(4-*i*-Pr-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۷/۱۲
۱۷	$(4\text{-}CF_3\text{-Ph})NH$	Ph	$(4\text{-F-Ph})CH_2$	۷/۸۵
۱۸	(4-NC-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۷/۸۲
۱۹	(4-EtOCOPh)NH	Ph	$(4\text{-F-Ph})CH_2$	۶/۵۵
۲۰	(4-HOCO-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۶/۴۶
۲۱	$(4\text{-}H_2NCO\text{-Ph})NH$	Ph	$(4\text{-F-Ph})CH_2$	۷/۵۴
۲۲	(4-MeO-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۷/۵۲
۲۳	(4-MeS-Ph)NH	Ph	$(4\text{-F-Ph})CH_2$	۷/۸۲
۲۴	$(4\text{-}MeSO_2\text{-Ph})NH$	Ph	$(4\text{-F-Ph})CH_2$	۷/۵۸
۲۵	(4-Cl-Ph)NH	4-Me-Ph	$(4\text{-F-Ph})CH_2$	۷/۹۶
۲۶	(4-Cl-Ph)NH	3-Cl-Ph	$(4\text{-F-Ph})CH_2$	۷/۷۲
۲۷	(4-Cl-Ph)NH	3,4-diCl-Ph	$(4\text{-F-Ph})CH_2$	۷/۲۱
۲۸	(4-Cl-Ph)NH	Ph	$(4\text{-}SO_2Me\text{-Ph})CH_2$	۸/۹۲
۲۹	(4-Cl-Ph)NH	Ph	$(4\text{-(Morpholino)}SO_2\text{-Ph})CH_2$	۹/۰۰

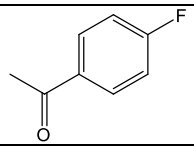
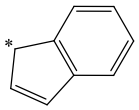
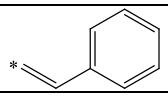


جدول ۴-۲- جزئیات ساختاری ترکیبات ۳۰ الی ۶۳

ردیف	**Position**	**R_1**	**R_2**	**R_3**	$log(1/IC_{50})$
۳۰	۴	Cbz	Ph	$PhCH_2$	۷/۳۲
۳۱	۴	H	Ph	$PhCH_2$	۵/۳۷
۳۲	۴	Ac	Ph	$PhCH_2$	۷/۸۰
۳۳	۴	*i*-PrCO	Ph	$PhCH_2$	۷/۴۱
۳۴	۴	Ms	Ph	$PhCH_2$	۷/۴۰
۳۵	۴	Ac	3-Cl-Ph	$(4\text{-F-Ph})CH_2$	۸/۵۲
۳۶	۴	Ac	4-Me-Ph	$(4\text{-F-Ph})CH_2$	۸/۵۵
۳۷	۴	Ac	3,4-diCl-Ph	Ph	۸/۷۲
۳۸	۴	Ac	3,4-diCl-Ph	$(4\text{-F-Ph})CH_2$	۸/۹۲
۳۹	۴	Ac	3-Cl,4-Me-Ph	Ph	۹/۲۱
۴۰	۴	Ac	3-Cl,4-Me-Ph	$(4\text{-F-Ph})CH_2$	۹/۵۴
۴۱	۴	Ac	3-Cl,4-*i*-Pr-Ph	$(4\text{-F-Ph})CH_2$	۷/۷۴
۴۲	۴	Ac	3-Cl,4-MeO-Ph	$(4\text{-F-Ph})CH_2$	۷/۵۲
۴۳	۴	Ms	3-Cl,4-Me-Ph	$(4\text{-F-Ph})CH_2$	۹/۳۰
۴۴	۴	Ms	3,4-diCl-Ph	$(4\text{-F-Ph})CH_2$	۸/۴۸
۴۵	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}CF_3\text{Ph})CH_2$	۸/۰۶
۴۶	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}NO_2\text{Ph})CH_2$	۸/۶۲
۴۷	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}NH_2\text{Ph})CH_2$	۸/۰۴
۴۸	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-(AcNH)Ph})CH_2$	۸/۲۳
۴۹	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-(MsNH)Ph})CH_2$	۸/۶۶
۵۰	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-(Morpholino)Ph})CH_2$	۸/۵۵
۵۱	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-(MeO)Ph})CH_2$	۸/۱۰
۵۲	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-(MeS)Ph})CH_2$	۸/۵۱
۵۳	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}(MeSO_2)\text{Ph})CH_2$	۸/۶۶
۵۴	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}(EtSO_2)\text{Ph})CH_2$	۸/۷۲
۵۵	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}(i\text{-}PrSO_2)\text{Ph})CH_2$	۸/۸۲
۵۶	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}(NH_2SO_2)\text{Ph})CH_2$	۸/۴۷
۵۷	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}(MeNHSO_2)\text{Ph})CH_2$	۸/۸۲
۵۸	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}(Me_2NSO_2)\text{Ph})CH_2$	۸/۹۲
۵۹	۴	$MeSO_2$	3,4-diCl-Ph	$(4\text{-}((morpholino)SO_2)\text{Ph})CH_2$	۸/۸۹
۶۰	۳	Cbz	Ph	$PhCH_2$	۵/۹۲
۶۱	۳	H	Ph	$PhCH_2$	۶/۱۱
۶۲	۳	Ac	Ph	$PhCH_2$	۶/۱۷
۶۳	۳	Ms	Ph	$PhCH_2$	۶/۷۴



جدول ۴-۳-جزئیات ترکیبات ۶۴ الی ۱۰۴

ردیف	$\mathbf{R_1}$	$\mathbf{R_2}$	$\mathbf{R_3}$	$log(1/IC_{50})$
۶۴	Me	Ph		۵/۷۲
۶۵	Me	Ph	Ph	۵/۴۴
۶۶	Me	Ph	(4-Cl-Ph)C*OH	۵/۲۸
۶۷	Me	Ph		۵/۶۸
۶۸	Me	Ph		۵/۹۲
۶۹	Me	Ph	$Ph(CH_2)_2$	۵/۹۶
۷۰	Me	Ph	PhO	۵/۸۰
۷۱	Me	Ph	$PhCH_2O$	۵/۸۲
۷۲	Me	Ph	$(4\text{-F-Ph})CH_2$	۶/۵۱
۷۳	*n*-Bu	Ph	$PhCH_2$	۶/۸۹
۷۴	*c*-Hex	Ph	$PhCH_2$	۶/۹۶
۷۵	*c*-HexCH_2	Ph	$PhCH_2$	۶/۶۴
۷۶	Ph	Ph	$PhCH_2$	۶/۵۱
۷۷	$PhCH_2$	Ph	$PhCH_2$	۷/۴۲
۷۸	$(2\text{-Cl-Ph})CH_2$	Ph	$PhCH_2$	۷/۴۸
۷۹	$(3\text{-Cl-Ph})CH_2$	Ph	$PhCH_2$	۷/۰۶
۸۰	$(4\text{-Cl-Ph})CH_2$	Ph	$PhCH_2$	۶/۶۶
۸۱	$(4\text{-Me-Ph})CH_2$	Ph	$PhCH_2$	۶/۴۸
۸۲	$Ph(CH_2)_2$	Ph	$PhCH_2$	۶/۴۴
۸۳	$(\text{furan-2-yl})CH_2$	Ph	$PhCH_2$	۷/۰۹
۸۴	$(\text{pyridine-4-yl})CH_2$	Ph	$PhCH_2$	۶/۶۲
۸۵	H	Ph	$PhCH_2$	۶/۲۴
۸۶	CF_3CH_2	Ph	$PhCH_2$	۷/۱۲
۸۷	$(2\text{-Me-Ph})CH_2$	Ph	$PhCH_2$	۷/۴۷
۸۸	$PhCH_2$	3-Cl-Ph	$PhCH_2$	۷/۳۶
۸۹	$PhCH_2$	3,4-diCl-Ph	$PhCH_2$	۷/۳۷
۹۰	Me	2-Me-Ph	$PhCH_2$	۵/۵۳
۹۱	Me	3-Me-Ph	$PhCH_2$	۶/۸۰

ادامه جدول ۴-۳

۹۲	Me	4-Me-Ph	$PhCH_2$	۷/۰۵
۹۳	Me	4-*t*-Bu-Ph	$PhCH_2$	۶/۸۰
۹۴	Me	3,4-$(CH_2)_3$-Ph	$PhCH_2$	۶/۷۴
۹۵	Me	4-MeO-Ph	$PhCH_2$	۵/۸۲
۹۶	Me	3-Cl-Ph	$PhCH_2$	۶/۹۲
۹۷	Me	4-Cl-Ph	$PhCH_2$	۵/۹۲
۹۸	Me	3,4-diCl-Ph	$PhCH_2$	۷/۲۴
۹۹	Me	3-Cl,4-F-Ph	$PhCH_2$	۶/۲۲
۱۰۰	Me	3,4-diF-Ph	$PhCH_2$	۵/۶۸
۱۰۱	Me	3-CF_3-Ph	$PhCH_2$	۶/۲۹
۱۰۲	Me	3-CN-Ph	$PhCH_2$	۵/۳۹
۱۰۳	Me	$PhCH_2$	$PhCH_2$	۶/۵۱
۱۰۴	Me	3,4-diCl-Ph	(4-F-Ph)CH_2	۷/۳۰

*: در استخلاف‌هایی که کربن محل اتصال با علامت (*) نشاندار شده است کربن محل اتصال و کربن متصل به شاخه R_3 مشترک می‌باشد.

۴-۳-۳- محاسبه توصیف‌کننده‌های مولکولی

پس از رسم و بهینه‌سازی ساختارها، جهت بدست آوردن توصیف‌کننده‌های ساختاری از نرم‌افزار Dragon 2.1 استفاده شد. ساختارهای بهینه شده به صورت تک تک به این نرم‌افزار داده شدند و سپس توصیف‌کننده‌های محاسبه شده برای هر یک از آنها در یک فایل مجزا ذخیره گردید. در پایان فایل‌های ذخیره شده به یکدیگر پیوست شده و در قالب یک فایل درآورده شدند.

۴-۳-۴- انتخاب توصیف‌کننده‌های مهم

به منظور انتخاب توصیف‌کننده‌های مهم از میان تمام توصیف‌کننده‌های محاسبه شده توسط نرم‌افزار DRAGON، ابتدا به وسیله برنامه‌ای که در محیط نرم‌افزار MATLAB نوشته شده بود از میان تمام توصیف‌کننده‌ها، توصیف‌کننده‌هایی که بیش از ۹۰٪ از مقادیر آنها یکسان بودند حذف شدند. سپس جهت اطمینان از عدم وجود همبستگی معنادار بین توصیف‌کننده‌های منتخب، با استفاده از برنامه‌ی دیگری که در محیط نرم‌افزار MATLAB نوشته و اجرا شد، ضریب همبستگی

میان تمام توصیف‌کننده‌های باقیمانده از مرحله قبل به صورت دو به دو محاسبه گردید و از بین دو توصیف‌کننده که دارای همبستگی بیشتر از ۰/۹ با یکدیگر بودند، توصیف‌کننده‌ای که دارای همبستگی کمتری با متغیر وابسته بود حذف شد. در پایان برای انتخاب توصیف‌کننده‌های مهم از میان توصیف‌کننده‌های باقیمانده از دو مرحله قبل، نرم‌افزار PASW 18 به کار گرفته شد. به کمک این نرم‌افزار و با استفاده از روش رگرسیون خطی مرحله‌ای توصیف‌کننده‌های مهم انتخاب شدند. در این مرحله ۲۶ توصیف‌کننده ارتباط معناداری با داده‌های مورد بررسی داشتند که جهت بدست آوردن بهترین مدل، از آنها به عنوان متغیر مستقل در کنار مقادیر pIC_{50} به عنوان متغیر وابسته در مدل‌سازی استفاده گردید. جدول ۴-۴ توصیف‌کننده‌های بدست آمده در این مرحله را به همراه کلاس مربوطه نشان می‌دهد.

جدول ۴-۴- توصیف کننده‌های انتخاب شده توسط روش رگرسیون خطی مرحله‌ای به همراه کلاس مربوطه

ردیف	نشانه*	کلاس	توضیحات
۱	GATS5p[۲]	2D autocorrelation	Geary autocorrelation – lag5
۲	C001	Atom-centered fragments	CH3R/CH4
۳	R8v[۳]	GETAWAY	R autocorrelation of lag 8
۴	RBF	Constitutional	Rotatable bond fraction
۵	Mor16v[۳]	3D-MoRES	3D-MoRSE – signal 16
۶	Hy	Molecular properties	Hydrophilic factor
۷	R3u[۵]	GETAWAY	R autocorrelation of lag 3
۸	GATS2e[۱]	2D autocorrelation	Geary autocorrelation –lag 2
۹	Mor18m[۴]	3D-MoRES	3D-MoRES – signal 18
۱۰	JGI5	Topological	Mean topological charge index of order 5
۱۱	Mor16m[۴]	3D-MoRES	3D-MoRES – signal 16
۱۲	BEHe6	Burden eigenvalue	Highest eigenvalue n. 6 of Burden matrix
۱۳	BEHp8	Burden eigenvalue	Highest eigenvalue n. 8 of Burden matrix
۱۴	X3sol	Connectivity indices	Salvation connectivity index chi-3
۱۵	GATS8p[۲]	2D autocorrelation	Geary autocorrelation –lag 8
۱۶	ATS8p[۲]	2D autocorrelation	Broto-Moreau autocorrelation of topological structure - lag 8
۱۷	GATS1m[۴]	2D autocorrelation	Geary autocorrelation – lag 1
۱۸	MATS3e[۱]	2D autocorrelation	Moran autocorrelation – lag 3
۱۹	GATS6m[۴]	2D autocorrelation	Geary autocorrelation – lag 6
۲۰	HATS8m[۴]	GETAWAY	Leverage-weighted autocorrelation of lag 8
۲۱	R3p[۲]	GETAWAY	R autocorrelation of lag 3
۲۲	E2e[۱]	WHIM	2nd component accessibility directional WHIM index
۲۳	BEHv1	Burden eigenvalue	Highest eigenvalue n. 8 of Burden matrix
۲۴	H0p[۲]	GETAWAY	H autocorrelation of lag 0
۲۵	nHAcc	Functional group counts.	Number of acceptor atoms for H-bonds (N, O,F)
۲۶	Mor10p[۲]	3D-MoRES	3D-MoRES – signal 10

*معنی حروف قرار گرفته در انتهای علامت توصیف‌کننده‌ها در این ستون:

۱. e: وزن‌دار شده به وسیله الکترونگاتیویته سندرسون

۲. p: وزن‌دار شده به وسیله قطبش‌پذیری اتمی

۳. v: وزن‌دار شده به وسیله حجم واندروالس اتمی

۴. m: وزن‌دار شده به وسیله جرم اتمی

۵. u: با هیچ ویژگی اتمی وزن‌دار نشده است.

۴-۴- ایجاد مدل QSAR

برای ایجاد و توسعه مدل‌های توضیح دهنده‌ی رابطه ساختار- فعالیت، سری داده‌ها به سه گروه آموزش، ارزیابی و تست تقسیم شدند. سری ارزیابی و تست هرکدام شامل ۲۱ ترکیب هستند که بصورت تصادفی و به گونه‌ای انتخاب شده‌اند که تمام طیف داده‌ها را دربر بگیرند. بر اساس جداول ۴-۱، ۴-۲ و ۴-۳ سری ارزیابی شامل ترکیبات شماره ۳، ۸، ۱۳، ۱۸، ۲۳ ،۲۸، ۳۳، ۳۸، ۴۳، ۴۸، ۵۳، ۵۸، ۶۳، ۶۸، ۷۳، ۷۸، ۸۳، ۸۸، ۹۳، ۹۸، ۱۰۳ می‌باشد که از آنها برای ارزیابی خطای مدل و بهینه‌سازی آن استفاده شد. ترکیبات شماره ۲، ۷، ۱۲، ۱۷، ۲۲، ۲۷، ۳۲، ۳۷، ۴۲، ۴۷، ۵۲، ۵۷، ۶۲، ۶۷، ۷۲، ۷۷، ۸۲، ۸۷، ۹۲، ۹۷، ۱۰۲ نیز سری تست را تشکیل می‌دهند که از ابتدا در مدل‌سازی از آنها استفاده نشد و کاملا کنار گذاشته شدند. ۶۲ ترکیب باقیمانده نیز سری آموزش را تشکیل می‌دهند که از آنها برای ایجاد مدل استفاده گردید.

۴-۴-۱- مدل‌سازی با استفاده از شبکه عصبی مصنوعی(ANN)

در این پروژه سعی شده تا به کمک روش‌های غیرخطی مدلی ارائه شود که بتواند به خوبی و با قدرت بالا خواص بازدارندگی سری ترکیبات مورد بررسی را پیش‌بینی نماید. بدین منظور برای ایجاد مدل غیرخطی، که بتواند ارتباط مناسبی بین ساختار و فعالیت دارویی ترکیبات مورد نظر فراهم نماید، شبکه عصبی مصنوعی انتخاب گردید. شبکه عصبی انتخاب شده یک شبکه عصبی پیش‌خور سه‌لایه با الگوی آموزش پس‌انتشار می‌باشد. برای آموزش موفق شبکه و در نتیجه بدست آوردن شبکه‌ای با قدرت پیش‌بینی بالا باید عوامل موثر در آموزش شبکه عصبی مصنوعی مورد بررسی و بهینه‌سازی قرار گیرند. پارامترهای موثر و مهمی که در اینجا برای بهینه‌سازی شبکه مورد بررسی قرار گرفتند عبارتند از:

- تعداد متغیرهای ورودی شبکه

- تعداد نرون‌های لایه پنهای
- نوع تابع تبدیل
- نوع الگوریتم آموزشی
- تعداد دورهای آموزش

همانطور که در ابتدای فصل گفته شد ترکیبات ۵-اکسوپیرولیدین-۳-کربوکسامید به سه گروه شامل سری آموزش، سری ارزیابی و سری تست تقسیم شدند. معیار بهینه‌سازی و انتخاب بهترین شبکه، به حداقل رساندن میانگین مربعات خطای استاندارد (MSE) حاصل از شبکه برای سری ارزیابی درنظر گرفته شد. پس از تکمیل آموزش، جهت آزمایش قابلیت تعمیم و یادگیری شبکه و همچنین جلوگیری از خطای برازش اضافی، شبکه بهینه‌شده توسط سری تست مورد ارزیابی قرار گرفت. بدین ترتیب قابلیت یادگیری شبکه توسط ایجاد خروجی برای گروهی از ترکیبات که در مرحله آموزش مورد استفاده واقع نشده بودند، آزموده شد. زیرا در مرحله آموزش امکان حفظ الگوها بوسیله شبکه وجود دارد که در نتیجه، شبکه قادر نخواهد بود در مرحله‌ی تست پاسخ درستی ارائه نماید.

۴-۱-۴-۱- بهینه‌سازی پارامترهای مهم در آموزش شبکه عصبی مصنوعی

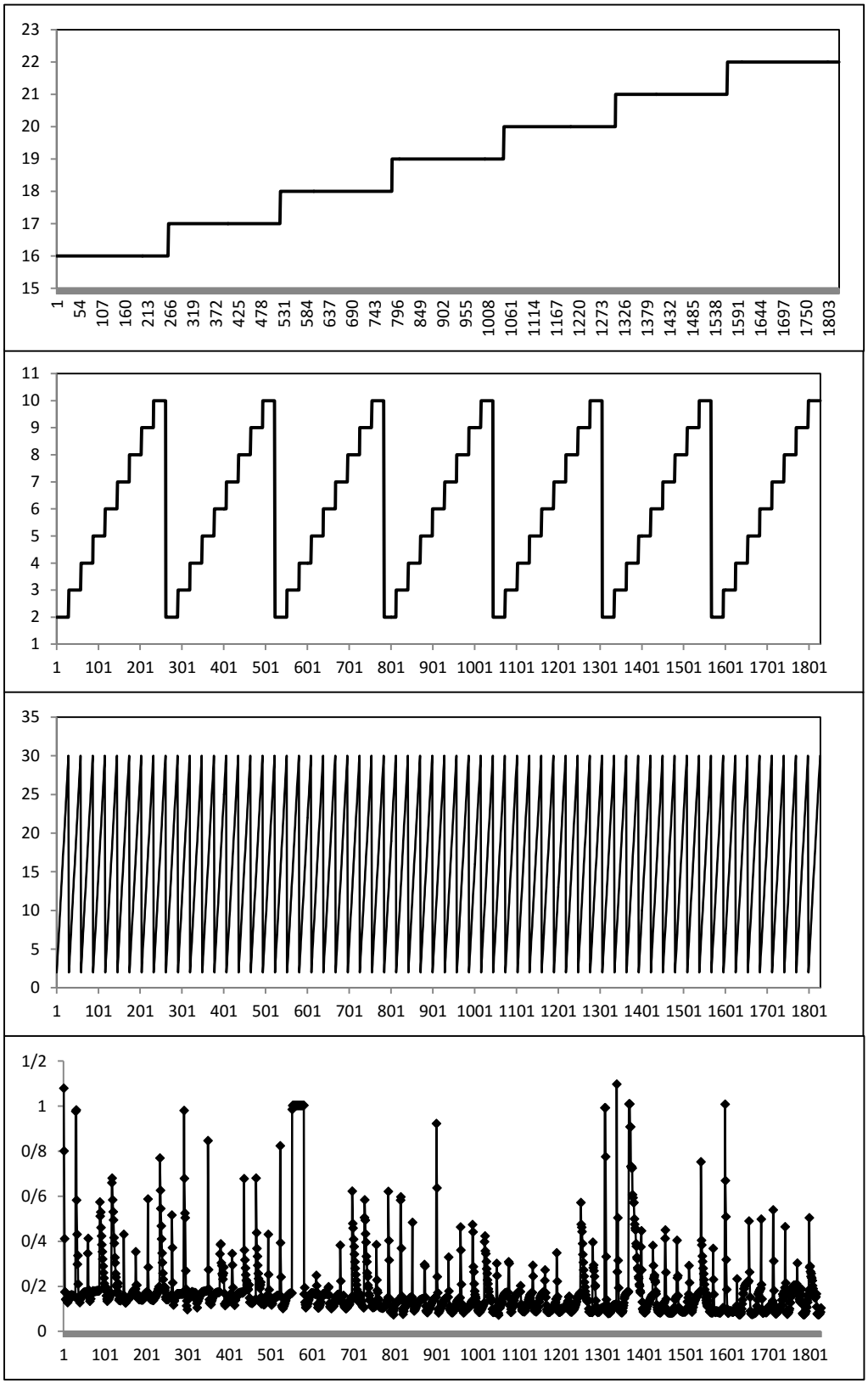
شبکه عصبی مصنوعی با استفاده از نرم‌افزار MATALB طراحی و بهینه شد. پس از طراحی اولیه شبکه، برای اینکه شبکه قادر به ایجاد پاسخ‌های مطلوب و صحیحی باشد باید آموزش ببیند و پارامترهای موثر در آن بهینه‌سازی شوند. الگوریتم‌های مختلفی در جعبه ابزار[۱] نرم‌افزار MATLAB وجود دارد. در این تحقیق دو الگوریتم آموزشی لونبرگ-مارکورت و تنظیم بایزین مورد بررسی قرار گرفتند. با توجه به اینکه خروجی‌های شبکه عصبی داده‌های pIC_{50} می‌باشند، لذا از تابع انتقال خطی برای لایه خروجی استفاده گردید.

[۱] Toolbox

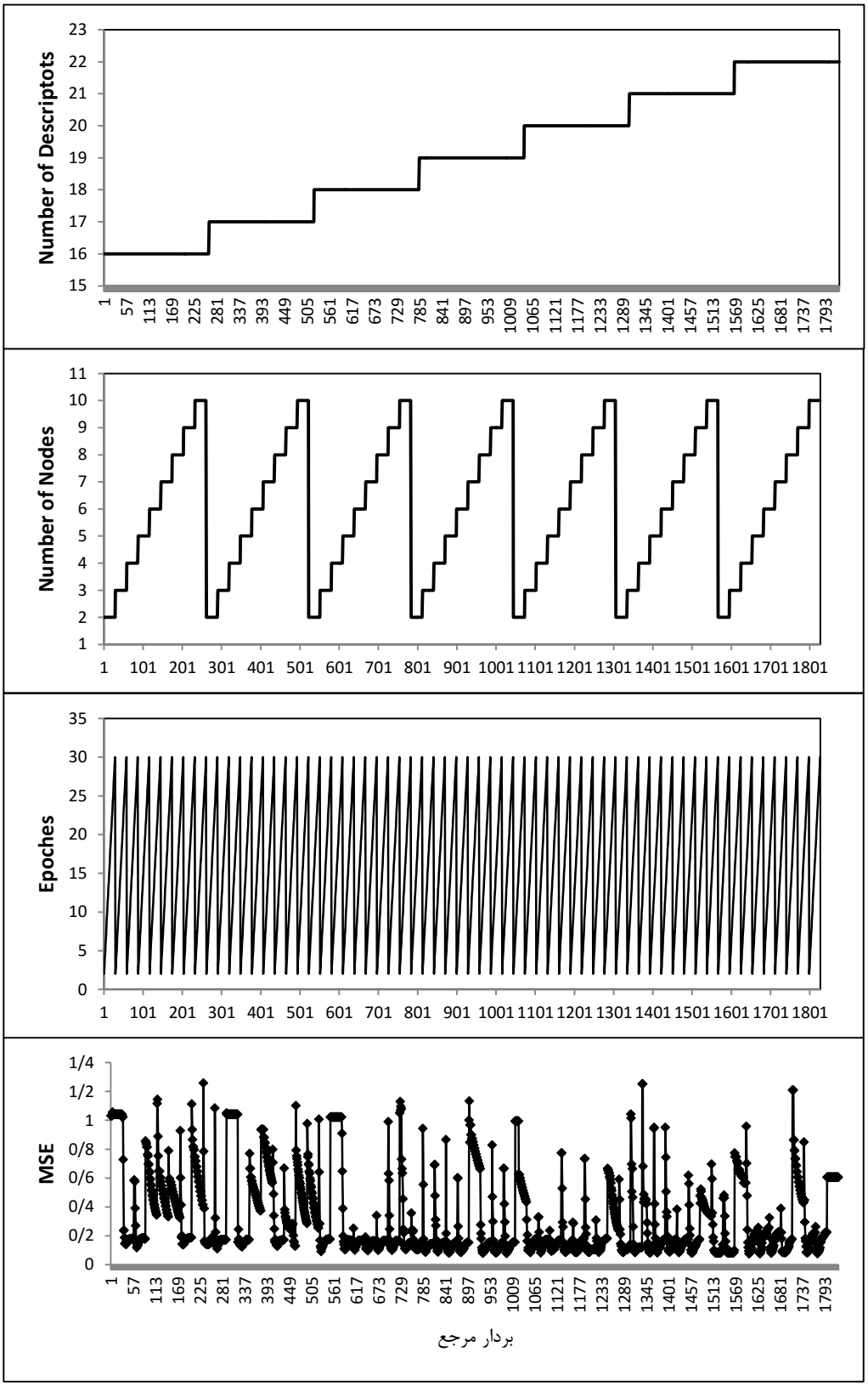
برای بهینه‌سازی و شناسایی بهترین مقادیر پارامترهای شبکه، شبکه‌هایی با ورودی‌های از ۲ تا ۲۶ توصیف‌کننده ایجاد شدند. هر یک از شبکه‌ها با دو الگوریتم آموزشی لونبرگ-مارکورت و تنظیم بایزین و با تعداد ۲ تا ۱۰ گره در لایه پنهان، آموزش داده شدند. برای بدست آوردن بهترین تابع انتقال در لایه پنهان مدل‌های شبکه عصبی، از توابع لگاریتم زیگموئید و تانژانت هایپربولیک به عنوان تابع تبدیل استفاده گردید و تعداد دورهای آموزش نیز از ۲ تا ۳۰ تغییر داده شد. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطای استاندارد به عنوان معیار انتخاب گردید. شبکه‌های آموزش یافته برای پیش‌بینی مقادیر pIC_{50} برای سری ارزیابی به کار گرفته شد و کوچکترین مقدار میانگین مربعات خطای استاندارد برای این سری به عنوان معیار انتخاب بهترین شبکه در نظر گرفته شد. به این ترتیب به منظور بهینه‌سازی همزمان تمام پارامترها ۲۶۱۰۰ شبکه سه‌لایه آموزش داده شدند و سپس هر یک از آنها برای پیش‌بینی خاصیت بازدارندگی سری ارزیابی مورد استفاده قرار گرفتند. برای مشخص شدن چگونگی روند بهینه‌سازی و محاسبات انجام شده برای توابع مختلف آموزش و تبدیل، بخشی از نتایج به صورت نمودار در شکل‌های ۴-۲، ۴-۳، ۴-۳ و ۴-۵ آورده شده‌اند. در این نمودارها بردار مرجع بیانگر تعداد الگوریتم‌هایی است که با تغییر پارامترهای شبکه در راستای بهینه‌سازی مورد ارزیابی قرار گرفته‌اند.

بر اساس نتایج بدست آمده، شبکه عصبی با الگوریتم لونبرگ-مارکورت با تابع تبدیل تانژانت هایپربولیک و مشخصاتی که در جدول ۴-۵ ذکر شده است به عنوان بهترین شبکه انتخاب گردید.

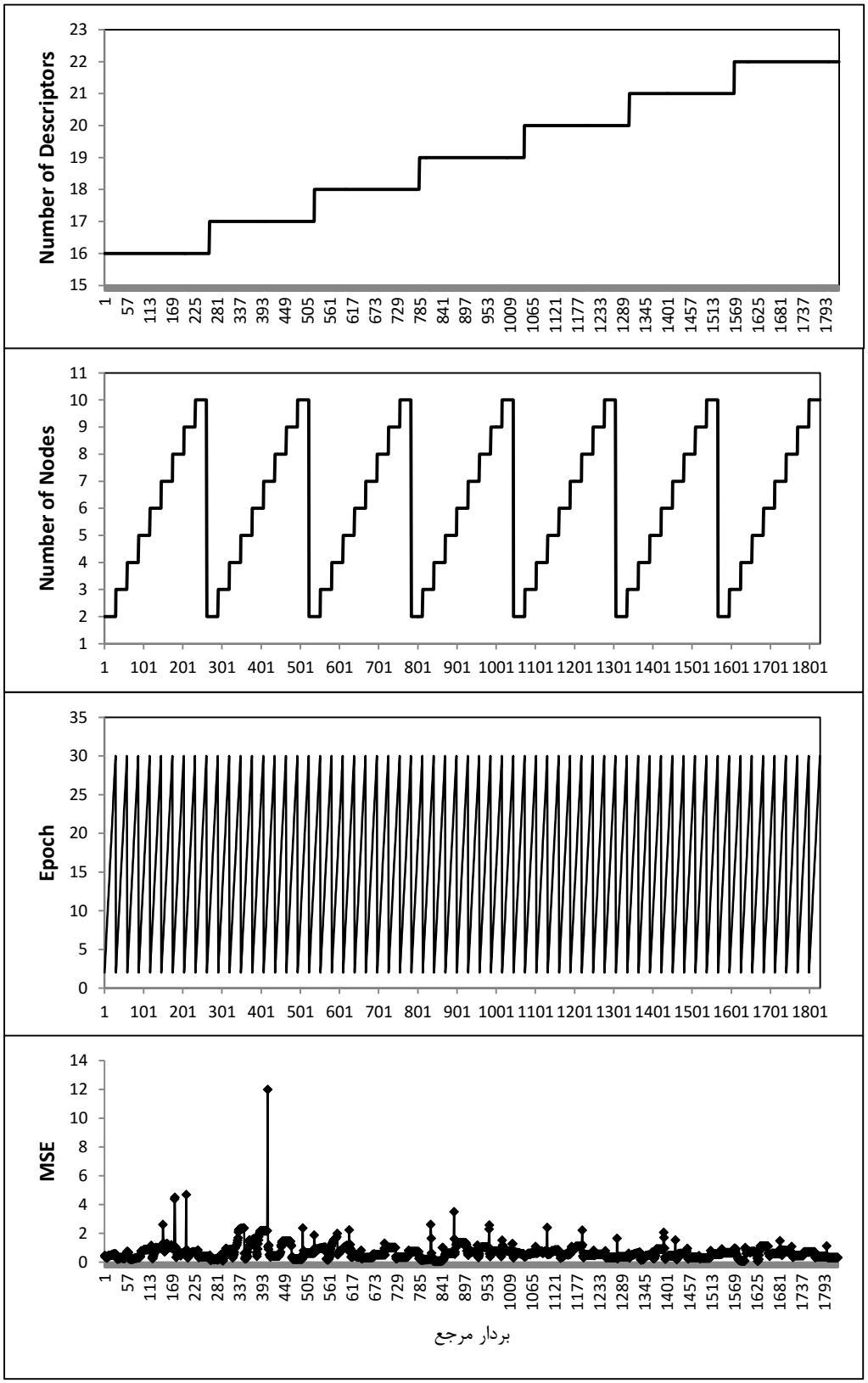
توصیف‌کننده‌های بکارگرفته شده در مدل شبکه عصبی به همراه کلاس مربوط به آنها در جدول ۴-۶ آورده شده‌اند. همچنین جهت بررسی میزان همبستگی میان این توصیف‌کننده‌ها ضرایب همبستگی آنها در جدول ۴-۷ نشان داده شده است که با توجه به مقادیر این ضرایب، همبستگی معناداری بین توصیف‌کننده‌های به کار گرفته شده وجود ندارد.



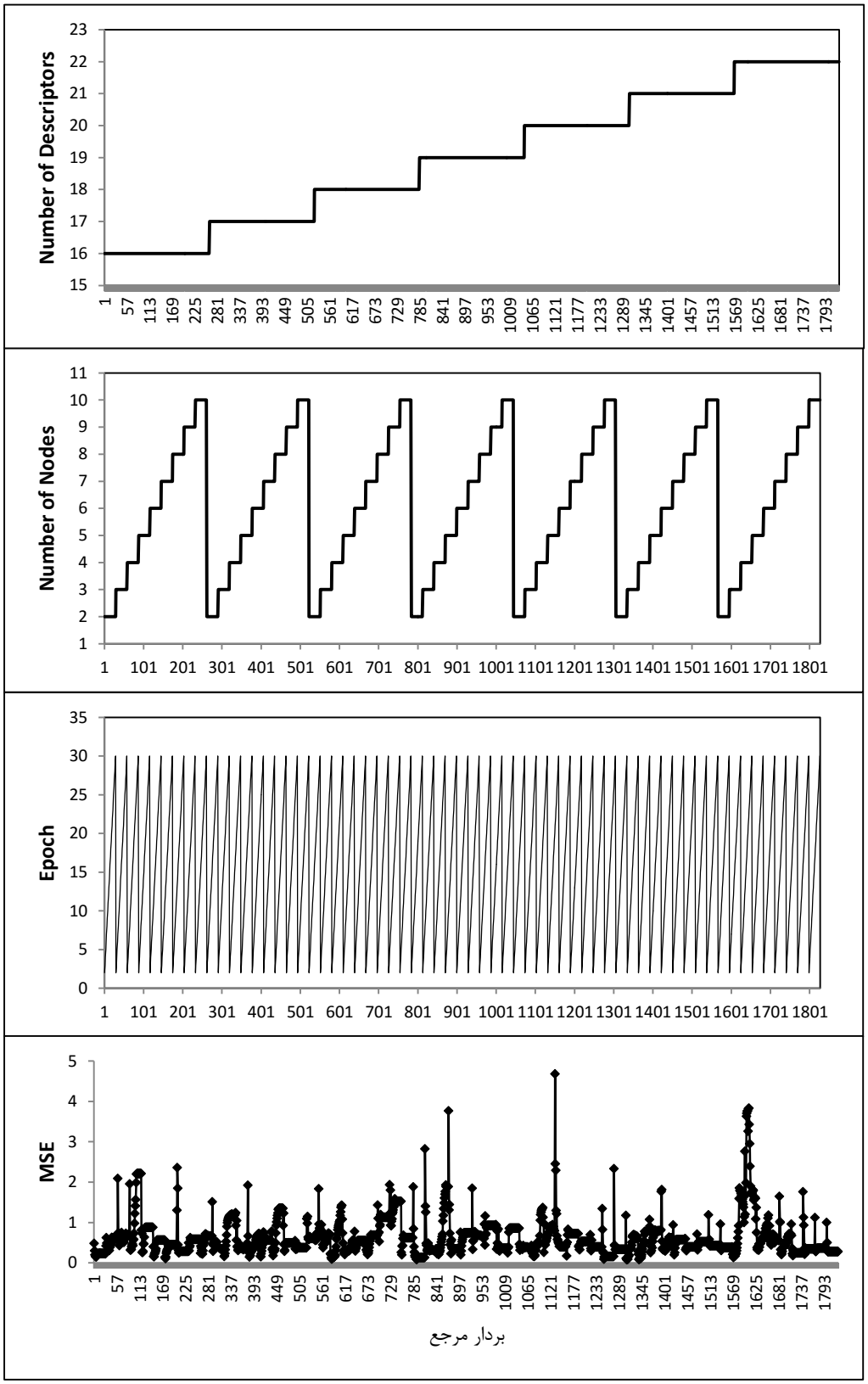
شکل ۴-۲- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین به همراه تابع تبدیل تانژانت هایپربولیک



شکل ۴-۳- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین به همراه تابع تبدیل لگاریتم زیگموئید



شکل ۴-۴- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش الگوریتم لونبرگ-مارکورت به همراه تابع تبدیل تانژانت هایپربولیک



شکل ۴-۵- نمودارهای الف) تعداد توصیف‌کننده‌ها، ب)تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش الگوریتم لونبرگ مارکورت به همراه تابع تبدیل لگاریتم زیگموئید

جدول ۴-۵- پارامترهای شبکه بهینه بدست آمده بر اساس مقادیر میانگین مربعات خطای استاندارد (MSE)

MSE	تعداد دورهای آموزش	تعداد گره‌های لایه پنهان	تابع تبدیل	تابع آموزش	تعداد توصیف‌کننده‌ها
۰/۰۷۷۷	۷	۴	لگاریتم زیگموئید	تنظیم بایزین	۲۲
۰/۰۷۳۴	۷	۲	تانژانت هایپربولیک	تنظیم بایزین	۲۰
۰/۰۸۴۲	۴	۲	لگاریتم زیگموئید	لونبرگ-مارکورت	۲۱
۰/۰۷۰۹	۱۲	۳	تانژانت هایپربولیک	لونبرگ-مارکورت	۱۹

جدول ۴-۶- توصیف‌کننده‌های مورد استفاده در مدل‌سازی شبکه عصبی مصنوعی بهینه به همراه کلاس مربوطه

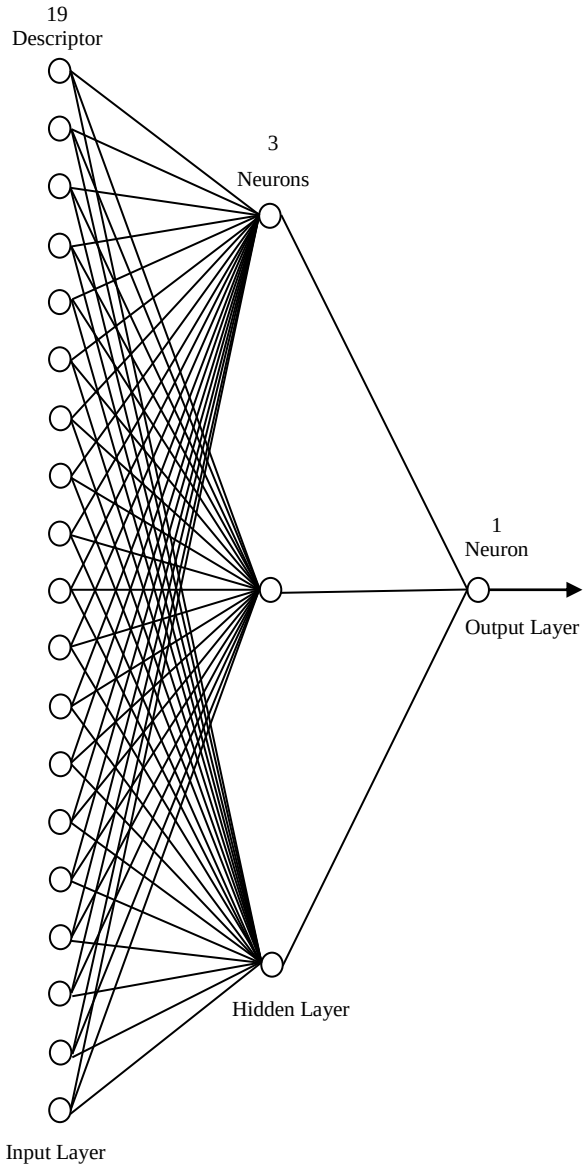
ردیف	نشانه	کلاس	توضیحات
۱	GATS5p	2D autocorrelation	Geary autocorrelation – lag5
۲	C001	Atom-centered fragments	CH3R/CH4
۳	R8v	GETAWAY	R autocorrelation of lag 8
۴	RBF	Constitutional	Rotatable bond fraction
۵	Mor16v	3D-MoRES	3D-MoRSE – signal 16
۶	Hy	Molecular properties	Hydrophilic factor
۷	R3u	GETAWAY	R autocorrelation of lag 3
۸	GATS2e	2D autocorrelation	Geary autocorrelation –lag 2
۹	Mor18m	3D-MoRES	3D-MoRES – signal 18
۱۰	JGI5	Topological charge indices	Mean topological charge index of order 5
۱۱	Mor16m	3D-MoRES	3D-MoRES – signal 16
۱۲	BEHe6	Burden eigenvalue	Highest eigenvalue n. 6 of Burden matrix
۱۳	BEHp8	Burden eigenvalue	Highest eigenvalue n. 8 of Burden matrix
۱۴	X3sol	Connectivity indices	Salvation connectivity index chi-3
۱۵	GATS8p	2D autocorrelation	Geary autocorrelation –lag 8
۱۶	ATS8p	2D autocorrelation	Broto-Moreau autocorrelation of topological structure - lag 8
۱۷	GATS1m	2D autocorrelation	Geary autocorrelation – lag 1
۱۸	MATS3e	2D autocorrelation	Moran autocorrelation – lag 3
۱۹	GATS6m	2D autocorrelation	Geary autocorrelation – lag 6

جدول ۴-۷- ماتریس همبستگی برای توصیف‌کننده‌های انتخاب شده

	GATS5p	C001	R8v	RBF	Mor16v	Hy	R3u	GATS2e	Mor18m	JGI5	Mor16m	BEHe6	BEHp8	X3sol	GATS8p	ATS8p	GATS1m	MATS3e	GATS6m
GATS5p	۱																		
C001	۰/۰۷۸	۱																	
R8v	۰/۳۴۸	۰/۳۵۲	۱																
RBF	۰/۲۶۶	۰/۴۱۸	۰/۲۵۵	۱															
Mor16v	۰/۱۴۰	۰/۲۲۴	۰/۰۵۹	۰/۰۶۴	۱														
Hy	۰/۰۷۵	۰/۰۹۶	۰/۰۲۰	۰/۱۱۹	۰/۰۸۲	۱													
R3u	۰/۳۷۳	۰/۱۵۷	۰/۲۱۴	۰/۰۹۸	۰/۱۵۶	۰/۱۲۳	۱												
GATS2e	۰/۵۲۷	۰/۰۶۸	۰/۲۹۲	۰/۲۱۶	۰/۰۷۶	۰/۰۵۹	۰/۲۰۶	۱											
Mor18m	۰/۳۱۴	۰/۲۳۰	۰/۱۰۸	۰/۰۶۴	۰/۴۱۸	۰/۰۲۲	۰/۳۶۲	۰/۲۶۱	۱										
JGI5	۰/۵۵۵	۰/۱۶۴	۰/۳۸۷	۰/۳۳۵	۰/۱۱۷	۰/۲۲۱	۰/۲۷۹	۰/۵۸۲	۰/۰۲۳	۱									
Mor16m	۰/۵۳۲	۰/۲۱۶	۰/۱۴۵	۰/۲۰۳	۰/۷۰۷	۰/۰۴۶	۰/۲۱۰	۰/۳۴۲	۰/۴۲۹	۰/۳۷۰	۱								
BEHe6	-/۵۳۹	۰/۱۰۵	۰/۶۶۰	۰/۲۱۳	۰/۱۸۴	۰/۰۰۵	۰/۳۹۳	۰/۳۰۵	۰/۲۷۷	۰/۶۲۴	۰/۴۲۴	۱							
BEHp8	۰/۴۱۴	۰/۱۵۰	۰/۵۵۰	۰/۱۷۲	۰/۰۴۴	۰/۰۴۴	۰/۳۶۷	۰/۱۵۹	۰/۱۵۱	۰/۵۵۷	۰/۲۶۹	۰/۸۲۷	۱						
X3sol	۰/۶۱۶	۰/۰۲۵	۰/۴۴۷	۰/۲۰۷	۰/۰۱۵	۰/۱۱۱	۰/۳۶۴	۰/۲۸۰	۰/۰۹۲	۰/۷۸۲	۰/۴۲۵	۰/۸۴۰	۰/۸۳۸	۱					
GATS8p	۰/۳۰۸	۰/۱۰۱	۰/۰۳۵	۰/۳۵۵	۰/۱۴۰	۰/۰۹۴	۰/۰۵۱	۰/۲۲۵	۰/۱۱۷	۰/۲۷۹	۰/۱۳۶	۰/۰۱۷	۰/۰۴۲	۰/۱۸۱	۱				
ATS8p	۰/۷۶۷	۰/۰۷۹	۰/۲۶۲	۰/۱۷۴	۰/۲۶۶	۰/۰۱۹	۰/۳۷۱	۰/۴۵۹	۰/۳۵۳	۰/۴۷۱	۰/۵۵۸	۰/۵۲۶	۰/۳۵۶	۰/۵۹۵	۰/۴۶۳	۱			
GATS1m	۰/۱۵۶	۰/۲۸۴	۰/۰۹۲	۰/۲۶۶	۰/۱۰۰	۰/۰۴۲	۰/۰۶۷	۰/۳۲۶	۰/۱۱۸	۰/۰۲۳	۰/۰۰۲	۰/۱۸۸	۰/۱۸۹	۰/۱۸۷	۰/۰۳۲	۰/۰۵۵	۱		
MATS3e	۰/۲۵۰	۰/۱۱۹	۰/۱۴۰	۰/۰۸۲	۰/۰۵۲	۰/۲۰۶	۰/۲۱۵	۰/۰۰۱	۰/۰۲۸	۰/۲۵۵	۰/۱۹۵	۰/۲۱۰	۰/۱۴۳	۰/۲۳۶	۰/۰۱۵	۰/۱۸۳	-۰/۲۶۴	۱	
GATS6m	۰/۳۲۷	۰/۲۹۹	۰/۱۳۲	۰/۱۱۵	۰/۲۲۳	۰/۱۶۵	۰/۲۹۴	۰/۵۲۱	۰/۳۴۹	۰/۴۶۴	۰/۴۶۹	۰/۳۰۵	۰/۱۶۶	۰/۳۱۸	۰/۱۵۵	۰/۴۲۲	۰/۵۵۳	۰/۰۸۰	۱

۴-۴-۲- معماری شبکه عصبی بهینه شده

با توجه به نتایج حاصل در مرحله بهینه‌سازی، شبکه عصبی پیش‌خور سه‌لایه با الگوی آموزش پس انتشار با الگوریتم لونبرگ-مارکورت و تابع انتقال تانژانت هیپربولیک انتخاب گردید. ساختار هندسی شبکه بهینه شده در شکل ۴-۶ نشان داده شده است.



شکل ۴-۶- تصویر شماتیک ساختار هندسی شبکه عصبی مصنوعی به دست آمده پس از بهینه سازی

۴-۴-۳- ارزیابی مدل شبکه عصبی مصنوعی

جهت ارزیابی مدل شبکه عصبی ارائه شده در این پروژه برای پیش‌بینی داده‌های IC_{50}، خواص بازدارندگی پروتئین CCR5 در درمان HIV-1 از روش‌های مختلفی استفاده گردید که در ادامه به آنها پرداخته می‌شود.

۴-۴-۳-۱- ارزیابی مدل با استفاده از سری ارزیابی

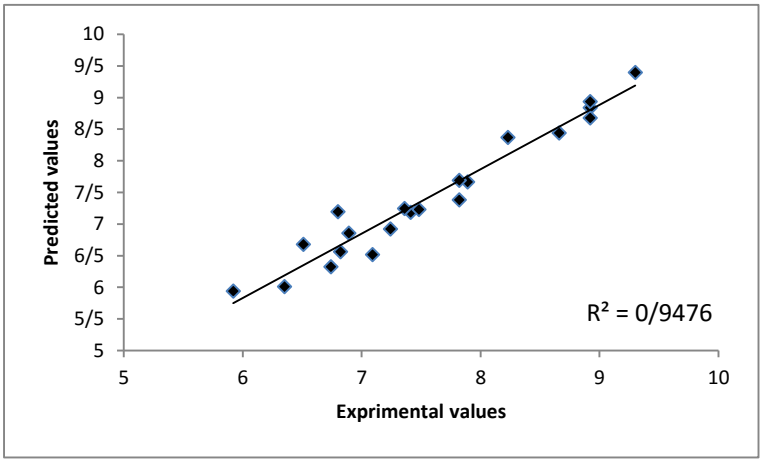
به منظور ارزیابی مدل با استفاده از سری ارزیابی، ابتدا به کمک ترکیبات سری آموزش، شبکه آموزش داده شد. سپس شبکه آموزش دیده برای پیش‌بینی داده‌های pIC_{50} برای سری ارزیابی به کار گرفته شد و درصد خطای نسبی برای مقادیر پیش‌بینی شده محاسبه گردید. نتایج حاصل از این پیش‌بینی به همراه خطای نسبی آنها در جدول ۴-۸ آورده شده است. نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی نیز در شکل ۴-۷ نشان داده شده است.

۴-۴-۳-۲- ارزیابی مدل با استفاده از سری تست

پس از انتخاب بهترین شبکه با استفاده از نتایج حاصل از سری ارزیابی، قدرت پیش‌بینی شبکه بوسیله داده‌های سری تست که در مرحله آموزش از آنها استفاده نشده است، مورد بررسی قرار گرفت. نتایج بدست آمده از این بررسی در کنار مقادیر خطای نسبی آنها در جدول ۴-۹ و نمودار مربوط به این نتایج برحسب مقادیر تجربی در شکل ۴-۸ نشان داده شده‌اند.

جدول ۴-۸-مقادیر مشاهده شده و پیش‌بینی شده pIC_{50} برای ترکیبات سری ارزیابی با استفاده از شبکه عصبی بهینه

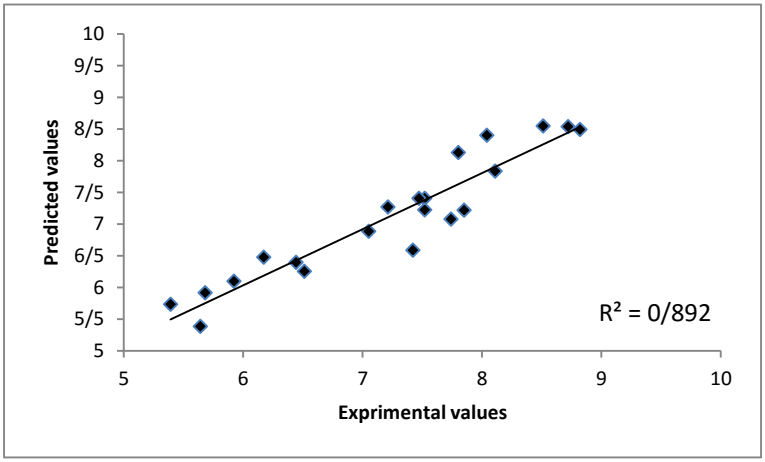
شماره	مقدار تجربی	مقدار پیش‌بینی	خطا (٪)
۳	۶/۸۲	۶/۵۷	-۳/۷۳۴۶
۸	۶/۳۵	۶/۰۱	-۵/۳۰۳۹
۱۳	۷/۸۹	۷/۶۷	-۲/۸۳۶۵
۱۸	۷/۸۲	۷/۳۸	-۵/۶۰۳۶
۲۳	۷/۸۲	۷/۶۹	-۱/۶۴۸۳
۲۸	۸/۹۲	۸/۸۴	-۰/۸۸۴۵
۳۳	۷/۴۱	۷/۱۹	-۳/۰۳۱۰
۳۸	۸/۹۲	۸/۹۴	۰/۱۹۰۶
۴۳	۹/۳۰	۹/۴۰	۱/۰۴۰۹
۴۸	۸/۲۳	۸/۳۷	۱/۷۰۲۳
۵۳	۸/۶۶	۸/۴۴	-۲/۵۳۴۶
۵۸	۸/۹۲	۸/۶۸	-۲/۶۷۸۳
۶۳	۶/۷۴	۶/۳۳	-۶/۱۵۱۳
۶۸	۵/۹۲	۵/۹۴	۰/۳۱۰۸
۷۳	۶/۸۹	۶/۸۶	-۰/۵۰۰۷
۷۸	۷/۴۸	۷/۲۳	-۳/۳۳۲۹
۸۳	۷/۰۹	۶/۵۲	-۸/۰۶۳۵
۸۸	۷/۳۶	۷/۲۵	-۱/۵۳۴۰
۹۳	۶/۸۰	۷/۲۰	۵/۸۲۲۱
۹۸	۷/۲۴	۶/۹۲	-۴/۳۷۷۱
۱۰۳	۶/۵۱	۶/۶۸	۲/۶۲۰۶



شکل ۴-۷- نمودار مقادیر پیش‌بینی شده pIC_{50} با استفاده از شبکه عصبی مصنوعی برحسب مقادیر واقعی در سری ارزیابی

جدول ۴-۹- مقادیر مشاهده شده و پیش‌بینی شده pIC_{50} برای ترکیبات سری تست با استفاده از شبکه عصبی بهینه

شماره	مقدار تجربی	مقدار پیش‌بینی	خطا (٪)
۲	۷/۷۴	۷/۰۸	-۸/۵۵۳۰
۷	۵/۶۴	۵/۳۹	-۴/۴۸۰۵
۱۲	۸/۱۱	۷/۸۴	-۳/۳۴۴۰
۱۷	۷/۸۵	۷/۲۲	-۷/۹۷۸۳
۲۲	۷/۵۲	۷/۴۱	-۱/۴۷۳۴
۲۷	۷/۲۱	۷/۲۷	۰/۸۹۳۲
۳۲	۷/۸۰	۸/۱۳	۴/۲۳۵۹
۳۷	۸/۷۲	۸/۵۴	-۲/۰۴۵۹
۴۲	۷/۵۲	۷/۲۳	-۳/۹۱۷۶
۴۷	۸/۰۴	۸/۴۱	۴/۵۴۳۵
۵۲	۸/۵۱	۸/۵۵	۰/۴۷۴۷
۵۷	۸/۸۲	۸/۴۹	-۳/۷۱۳۲
۶۲	۶/۱۷	۶/۴۸	۴/۹۸۵۴
۶۷	۵/۶۸	۵/۹۲	۴/۱۹۳۷
۷۲	۶/۵۱	۶/۲۶	-۳/۹۱۰۹
۷۷	۷/۴۲	۶/۵۹	-۱۱/۱۵۰۹
۸۲	۶/۴۴	۶/۴۰	-۰/۶۱۱۸
۸۷	۷/۴۷	۷/۴۱	-۰/۸۴۲۰
۹۲	۷/۰۵	۶/۸۹	-۲/۳۱۴۹
۹۷	۵/۹۲	۶/۱۰	۳/۰۷۰۹
۱۰۲	۵/۳۹	۵/۷۴	۶/۴۲۳۰



شکل ۴-۸- نمودار مقادیر پیش‌بینی شده pIC_{50} با استفاده از شبکه عصبی مصنوعی برحسب مقادیر واقعی در سری تست

۴-۴-۳-۳- ارزیابی مدل به روش رد مرحله‌ای تک‌تک

نتایج حاصل از ارزیابی شبکه به روش رد مرحله‌ای تک‌تک در جدول ۴-۱۰ گنجانده شده است. شکل ۴-۹ نیز نموداری از مقادیر پیش‌بینی شده به این روش را در برابر مقادیر تجربی نمایش می‌دهد. با توجه به نتایج بدست آمده می‌توان گفت مدل ارائه شده داری خطای پیش‌بینی کم و قدرت تعمیم بالایی است و می‌توان به کمک آن داده‌های IC_{50} را برای هر ترکیب جدید از دسته ترکیبات مورد بررسی با خطای کم و قابل قبول پیش‌بینی کرد.

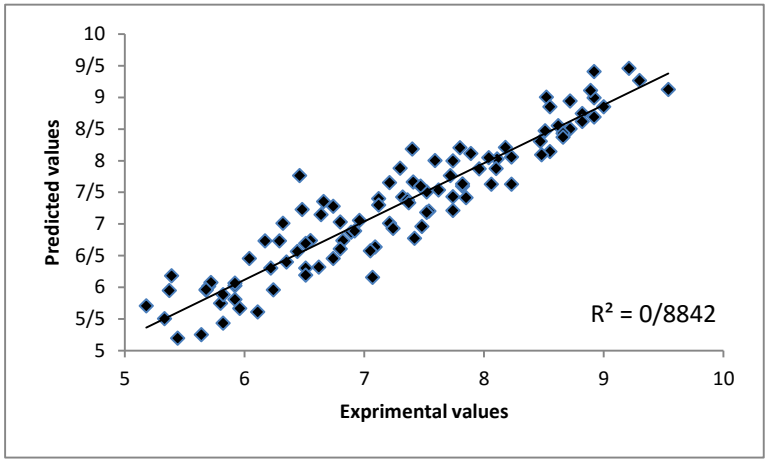
جدول ۴-۱۰- نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از روش رد مرحله‌ای تک‌تک

شماره	مقدار واقعی	بینی‌مقدار پیش	خطا (٪)	شماره	مقدار واقعی	بینی‌مقدار پیش	خطا (٪)
۱	۶/۳۲	۷/۰۱	۱۰/۹۶۵۲	۲۷	۷/۲۱	۷/۶۵	۶/۱۴۵۶
۲	۷/۷۴	۷/۲۱	-۶/۷۸۵۵	۲۸	۸/۹۲	۸/۹۹	۰/۷۹۲۶
۳	۶/۸۲	۶/۷۴	-۱/۱۶۲۸	۲۹	۹/۰۰	۸/۸۵	-۱/۶۵۸۹
۴	۷/۶۲	۷/۵۴	-۱/۰۹۱۹	۳۰	۷/۳۲	۷/۴۳	۱/۴۷۶۸
۵	۷/۲۱	۷/۰۱	-۲/۷۶۵۶	۳۱	۵/۳۷	۵/۹۵	۱۰/۸۵۶۶
۶	۵/۱۸	۵/۷۱	۱۰/۲۱۸۱	۳۲	۷/۸۰	۸/۲۱	۵/۲۱۷۹
۷	۵/۶۴	۵/۲۶	-۶/۸۱۰۳	۳۳	۷/۴۱	۷/۶۶	۳/۴۱۷۰
۸	۶/۳۵	۶/۴۰	۰/۸۲۵۲	۳۴	۷/۴۰	۸/۱۹	۱۰/۶۲۳۰
۹	۶/۰۴	۶/۴۵	۶/۸۶۹۲	۳۵	۸/۵۲	۹/۰۰	۵/۶۷۰۲
۱۰	۵/۷۰	۶/۰۲	۵/۵۸۴۲	۳۶	۸/۵۵	۸/۱۴	-۴/۷۴۱۵
۱۱	۸/۲۳	۷/۶۳	-۷/۳۱۴۷	۳۷	۸/۷۲	۸/۵۰	-۲/۴۸۹۷
۱۲	۸/۱۱	۸/۰۳	-۰/۹۹۱۴	۳۸	۸/۹۲	۹/۴۱	۵/۴۷۵۳
۱۳	۷/۸۹	۸/۱۱	۲/۸۳۰۲	۳۹	۹/۲۱	۹/۴۶	۲/۶۷۷۵
۱۴	۷/۷۴	۷/۴۳	-۴/۰۰۵۲	۴۰	۹/۵۴	۹/۱۲	-۴/۳۵۹۵
۱۵	۸/۱۸	۸/۲۱	۰/۳۶۵۵	۴۱	۷/۷۴	۸/۰۰	۳/۳۵۷۹
۱۶	۷/۱۲	۷/۴۰	۳/۹۲۲۸	۴۲	۷/۵۲	۷/۱۹	-۴/۴۵۰۸
۱۷	۷/۸۵	۷/۴۲	-۵/۵۱۴۶	۴۳	۹/۳۰	۹/۲۷	-۰/۳۶۵۶
۱۸	۷/۸۲	۷/۶۰	-۲/۸۴۴۰	۴۴	۸/۴۸	۸/۰۹	-۴/۵۸۰۲
۱۹	۶/۵۵	۶/۷۴	۲/۹۱۱۵	۴۵	۸/۰۶	۷/۶۳	-۵/۳۳۱۳
۲۰	۶/۴۶	۷/۷۶	۲۰/۱۸۱۱	۴۶	۸/۶۲	۸/۵۵	-۰/۷۶۹۱
۲۱	۷/۵۴	۷/۲۰	-۴/۴۷۷۵	۴۷	۸/۰۴	۸/۰۵	۰/۱۰۹۵
۲۲	۷/۵۲	۷/۵۱	-۰/۱۸۶۲	۴۸	۸/۲۳	۸/۰۶	-۲/۰۹۸۴
۲۳	۷/۸۲	۷/۶۳	-۲/۳۸۱۱	۴۹	۸/۶۶	۸/۴۳	-۲/۶۵۹۴
۲۴	۷/۵۸	۸/۰۰	۵/۴۳۳۵	۵۰	۸/۵۵	۸/۸۵	۳/۵۴۵۰
۲۵	۷/۹۶	۷/۸۷	-۱/۱۴۳۲	۵۱	۸/۱۰	۷/۸۸	-۲/۷۷۲۸
۲۶	۷/۷۲	۷/۷۷	۰/۶۰۷۵	۵۲	۸/۵۱	۸/۴۷	-۰/۴۶۰۶

ادامه جدول ۴-۱۰

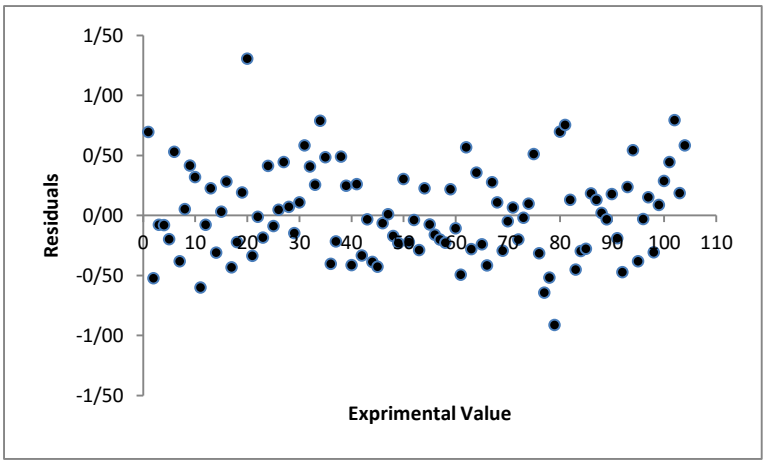
شماره	مقدار واقعی	بینی‌مقدار پیش	خطا (٪)
۵۳	۸/۷۲	۸/۹۴	۲/۵۶۶۵
۵۴	۸/۸۲	۸/۷۴	-۰/۸۷۰۷
۵۵	۸/۴۷	۸/۳۱	-۱/۹۲۲۱
۵۶	۸/۸۲	۸/۶۲	-۲/۳۰۵۰
۵۷	۸/۹۲	۸/۶۹	-۲/۵۶۹۵
۵۸	۸/۸۹	۹/۱۱	۲/۴۴۹۹
۵۹	۵/۹۲	۵/۸۱	-۱/۸۲۰۹
۶۰	۶/۱۱	۵/۶۱	-۸/۱۲۴۴
۶۱	۶/۱۷	۶/۷۴	۹//۱۷۹۹
۶۲	۶/۷۴	۶/۴۶	-۴/۲۰۰۳
۶۳	۵/۷۲	۶/۰۸	۶/۲۱۸۵
۶۴	۵/۴۴	۵/۲۰	-۴/۴۴۸۵
۶۵	۵/۲۸	۴/۸۶	-۷/۹۳۷۵
۶۶	۵/۶۸	۵/۹۶	۴/۸۶۴۴
۶۷	۵/۹۲	۶/۰۳	۱/۷۹۰۵
۶۸	۵/۹۶	۵/۶۷	-۴/۹۳۲۹
۶۹	۵/۸۰	۵/۷۵	-۰/۸۵۳۴
۷۰	۵/۸۲	۵/۸۸	۱/۰۹۴۵
۷۱	۶/۵۱	۶/۳۱	-۳/۱۱۶۷
۷۲	۶/۸۹	۶/۸۷	-۰/۲۸۱۶
۷۳	۶/۹۶	۷/۰۶	۱/۳۸۰۷
۷۴	۶/۶۴	۷/۱۵	۷/۶۸۸۳
۷۵	۶/۵۱	۶/۱۹	-۴/۸۶۴۸
۷۶	۷/۴۲	۶/۷۸	-۸/۶۶۹۸
۷۷	۷/۴۸	۶/۹۶	-۶/۹۳۱۸
۷۸	۸/۷۲	۸/۹۴	۲/۵۶۶۵

شماره	مقدار واقعی	بینی‌مقدار پیش	خطا (٪)
۷۹	۷/۰۶	۶/۱۶	-۱۲/۹۱۸۰
۸۰	۶/۶۶	۷/۳۶	۱۰/۴۷۳۰
۸۱	۵/۷۲	۷/۲۳	۱۱/۶۰۰۳
۸۲	۵/۴۴	۶/۵۷	۱/۹۸۷۶
۸۳	۵/۲۸	۶/۶۴	-۶/۳۸۳۶
۸۴	۵/۶۸	۶/۳۲	-۴/۴۹۲۴
۸۵	۵/۹۲	۵/۹۶	-۴/۴۸۵۶
۸۶	۵/۹۶	۷/۳۰	۲/۵۴۲۱
۸۷	۵/۸۰	۷/۶۰	۱/۷۰۵۵
۸۸	۵/۸۲	۷/۳۸	۰/۲۵۶۸
۸۹	۶/۵۱	۷/۳۳	-۰/۴۷۹۰
۹۰	۶/۸۹	۵/۵۱	۳/۳۲۶۵
۹۱	۶/۹۶	۶/۶۱	-۲/۸۰۱۵
۹۲	۷/۰۵	۶/۵۸	-۶/۷۲۶۲
۹۳	۶/۸۰	۷/۰۳	۳/۴۴۱۲
۹۴	۶/۷۴	۷/۲۸	۸/۰۲۹۷
۹۵	۵/۸۲	۵/۴۴	-۶/۵۸۴۲
۹۶	۶/۹۲	۶/۸۹	-۰/۴۳۹۳
۹۷	۵/۹۲	۶/۰۷	۲/۵۲۲۰
۹۸	۷/۲۴	۶/۹۳	-۴/۲۵۸۳
۹۹	۶/۲۲	۶/۳۱	۱/۳۸۲۶
۱۰۰	۵/۶۸	۵/۹۷	۵/۰۴۰۵
۱۰۱	۶/۲۹	۶/۷۳	۷/۰۳۶۶
۱۰۲	۵/۳۹	۶/۱۸	۱۴/۶۹۹۴
۱۰۳	۶/۵۱	۶/۷۰	۲/۸۴۳۳
۱۰۴	۷/۳۰	۷/۸۸	۷/۹۸۴۹



شکل ۴-۹- نمودار مقادیر پیش‌بینی شده pIC_{50} با استفاده از مدل شبکه عصبی مصنوعی بهینه شده برحسب مقادیر واقعی به روش رد مرحله‌ای تک‌تک

باقیمانده مقادیر محاسبه شده از مقادیر تجربی برحسب مقادیر تجربی pIC_{50} برای مولکول‌های مورد بررسی در شکل ۴-۱۰ رسم شده است. توزیع نسبتا یکنواخت و اتفاقی باقیمانده‌ها در دو طرف خط صفر نشان دهنده عدم وجود خطای معین در مدل حاصل از شبکه عصبی مصنوعی است.



شکل ۴-۱۰- نمودار مقادیر خطای مطلق در ارزیابی مدل شبکه عصبی مصنوعی به روش LOO در مقابل مقادیر تجربی pIC_{50}

۴-۴-۳-۴- آزمون Y- تصادفی

به منظور اجتناب از همبستگی‌های تصادفی و تضمین قدرت شبکه در پیش‌بینی مقادیر مورد نظر، آزمون Y- تصادفی انجام شد. در این مرحله، با استفاده از کدی که در نرم افزار MATLAB نوشته و

اجرا شد، مقادیری به عنوان پاسخ به صورت تصادفی و در محدوده مقادیر pIC50 برای ترکیبات مورد بررسی ایجاد گردید. سپس مدل شبکه عصبی بهینه با استفاده از ماتریس متغیرهای مستقل و مقادیر تصادفی پاسخ توسعه یافت. در نهایت با توجه به پاسخ شبکه، همبستگی متغیرهای مستقل با متغیرهای پاسخ با استفاده از شاخص R^2 مورد بررسی قرار گرفت. این عملیات ده بار تکرار شد. در جدول ۳-۱۱ مقادیر R^2 برای سری‌های ارزیابی و تست حاصل از این فرایند آورده شده است. مقادیر پایین R^2 نشان‌دهنده عدم وجود همبستگی شانسی در مدل به دست آمده از شبکه هستند.

جدول ۴-۱۱- مقادیر R^2 برای سری‌های آموزش،ارزیابی و تست پس از چندین آزمون Y- تصادفی

تعداد دفعات تکرار آزمون	R^2 سری ارزیابی (n=۲۱)	R^2 سری تست (n=۲۱)
۱	۰/۱۰۶۳	۰/۱۹۸۴
۲	۰/۰۳۰۲	۰/۰۹۰۵
۳	۰/۰۰۹۴	۰/۰۰۷۹
۴	۰/۰۰۷۸	۰/۰۴۷۱
۵	۰/۰۳۹۴	۰/۰۹۹۶
۶	۰/۱۷۳۳	۰/۰۷۰۰
۷	۰/۰۱۹۱	۰/۰۰۴۱
۸	۰/۰۹۲۷	۰/۰۰۱۷
۹	۰/۰۲۸۴	۰/۰۰۱۶
۱۰	۰/۰۰۵۸	۰/۰۸۴۳

۴-۴-۳-۵- ارزیابی مدل شبکه عصبی مصنوعی با استفاده از پارامترهای آماری

برای بررسی میزان کارایی مدل شبکه عصبی ارائه شده چهار پارامتر آماری MSE، MAE، REP و R^2 محاسبه شد. نتایج حاصل در جدول ۴-۱۲ آورده شده است.

نتایج حاصل نشان می‌دهند که مدل شبکه عصبی عصبی ارائه شده می‌تواند با قدرت بالا و خطای کم و قابل قبول قدرت بازدارندگی ترکیبات ۵-اکسوپیرولیدین-۳-کربوکسامید را پیش‌بینی نماید.

جدول ۴-۱۲-پارامترهای آماری محاسبه شده برای مدل شبکه عصبی مصنوعی

پارامتر	سری ارزیابی (n=۲۱)	سری تست (n=۲۱)	کل داده‌ها (n=۱۰۴)
MSE	۰/۰۷۰۹	۰/۱۲۱۰	۰/۱۴۰۵
MAE	۰/۲۲۳۰	۰/۲۸۳۱	۰/۳۰۱۰
REP	۳/۵۱۵۷	۴/۸۲۰۱	۵/۱۸۰۱
R^2	۰/۹۴۷۶	۰/۸۹۲۰	۰/۸۸۴۲

۴-۴-۴- مدل رگرسیون خطی چندگانه

با توجه به زیاد بودن تعداد توصیف‌کننده‌ها مدل رگرسیون خطی متناظر با شبکه عصبی مصنوعی، دارای ارزش علمی نمی‌باشد. در اینجا صرفا جهت ارائه ضرایب رگرسیون، رابطه رگرسیون خطی چندگانه با ۱۹ توصیف‌کننده آورده شده است.

$$pIC_{50} = 27.306 - 4.585\ GATS5p + 0.554\ C001 - 32.377\ R8v - \quad (۱-۴)$$

$$12.477\ RBF - 1.701\ Mor16v - 0.537\ Hy - 13.428\ R3u - 1.460\ GATS2e +$$
$$0.564\ Mor18m - 120.458\ JGI5 + 0.884\ Mr16m + 3.729\ BEHe6 - 5.679\ BEHp8 +$$
$$0.261\ X3sol + 1.672\ GATS8p - 11.660\ ATS8p - 165.898\ GATS1m -$$
$$4.009\ MATS3e + 117.120\ GATS6m$$

فصل پنجم

بحث و نتیجه‌گیری

۵-۱- بررسی توصیف کننده‌های انتخاب شده

مهمتر از قدرت پیش‌گویی میزان فعالیت ترکیبات جدید، رسالت دیگر مدل‌های QSAR ایجاد ارتباط بین فعالیت مورد نظر و پارامترهای ساختاری یا توصیف‌کننده‌هاست که بتوان به کمک این ارتباط به بررسی تئوری و در نهایت عملی ترکیباتی جدید با خواص بهتر پرداخت. در این بخش به بررسی توصیف‌کننده‌های انتخاب شده در فصول ۳ و ۴ می‌پردازیم و سعی می‌کنیم تا حد امکان توضیحی برای ارتباط بین این توصیف‌کننده‌ها و خاصیت مورد نظر بیابیم.

۵-۱-۱- توصیف‌کننده‌های گروه *2D Autocorrelation* [۴۴]

این توصیف‌کننده‌ها بر اساس تابع خود ارتباطی AC_l تعریف می‌شوند:

$$AC_l = \int_b^a f(x).f(x+l)dx \qquad (۱-۵)$$

که در آن $f(x)$ تابعی از متغیر x و l فاصله از x می‌باشد؛ a و b فاصله مطالعه شده را بیان می‌کنند. $f(x)$ معمولا یک تابع وابسته به زمان مانند یک سیگنال الکتریکی یا وابسته به مکان مانند دانسیته جمعیت در فضا می‌باشد. این گروه از توصیف‌کننده‌ها به بررسی اطلاعات مختلفی در مورد ترکیبات می‌پردازند که از دایره بحث این نوشتار خارج است و ما فقط به بررسی سه نوع از آنها که در مدل منتخب در فصل چهار به آنها اشاره شده است می‌پردازیم.

نوعی از این توصیف‌کننده‌ها که از گروه توصیف‌کننده‌های گروه Geary Autocorrelation است به شکل زیر محاسبه می‌شود:

$$C(d) = \frac{\frac{1}{2\Delta}\sum_{i=1}^{A}\sum_{j=1}^{A}\delta_{ij}.(w_i - w_j)^2}{\frac{1}{A-1}.\sum_{i=1}^{A}(w_i - \bar{w})^2} \qquad (۲-۵)$$

که w یک ویژگی اتم، $\overline{w}$ مقدار میانگین آن روی مولکول، A تعداد اتم‌ها و d فاصله توپولوژیکی است. δ_{ij} نیز به تابع کرونکر معروف است (در حالتی که $d_{ij} = d$ باشد δ برابر یک است، در غیر اینصورت صفر است) و Δ هم مجموعه δ هاست. مقدار این توصیف‌کننده‌ها که از جنس فاصله است از صفر تا بینهایت تغییر می‌کند.

نوع دوم از این دسته توصیف‌کننده‌ها که در اینجا بررسی می‌شود توصیف‌کننده‌های Moran autocorrelation است که به صورت زیر محاسبه می‌شوند:

$$I(d) = \frac{\frac{1}{\Delta}\sum_{i=1}^{A}\sum_{j=1}^{A}\delta_{ij}(w_i - \overline{w}).(w_j - \overline{w})}{\frac{1}{A}\sum_{i=1}^{A}(w_i - \overline{w})} \qquad (۳-۵)$$

پارامترها در این رابطه مشابه رابطه (۵-۲) هستند. این توصیف‌کننده‌ها مقادیری بین ۱+ و ۱- دارند.

نوع دیگری از این توصیف‌کننده‌ها، توصیف‌کننده‌های Moreau-Broto autocorrelation هستند که به وسیله رابطه (۵-۴) محاسبه می‌شوند:

$$ATS_d = \sum_{i=1}^{A}\sum_{j=1}^{A}\delta_{ij}.(w_i.w_j)_d = \mathbf{w^T}.{}^{\mathbf{m}}\mathbf{B}.\mathbf{w} \qquad (۴-۵)$$

که در اینجا پارامترهای w، δ و A نیز مانند دو رابطه قبل تعریف می‌شوند. $^{\mathbf{m}}\mathbf{B}$ یک ماتریس دو دویی خلوت[۱] از مرتبه mام است و $\mathbf{w}$ یک بردار A بعدی از ویژگی اتمی مورد نظر می‌باشد. ویژگی‌های اتمی نوعی عبارتند از: جرم اتمی، قطبش‌پذیری، بار و الکترونگاتیویته.

از این گروه توصیف کننده‌ها، GATS1m، GATS5p، GATS8p، GATS2e، MATS3e و ATS8p در مرحله انتخاب توصیف‌کننده‌های موثر در مدل ارائه شده در فصل چهارم برگزیده شده‌اند.

[۱] Binary sparse matrix

سه توصیف‌کننده GATS5p، GATS8p و ATS8p چگونگی توزیع بار و قطبش‌پذیری در سطح اتمی را نشان می‌دهند، که حضور این سه توصیف‌کننده نشان دهنده اهمیت این فاکتورها در نحوه عملکرد ترکیبات مورد بررسی است. توصیف‌کننده‌های GATS2e و MATS3e نقش خاصیت فیزیکولوژی الکترونگاتیویته ساندرسون را در ترکیبات بیان می‌کنند و توصیف‌کننده GATS1m نیز براساس وزن اتمی محاسبه شده است. با توجه به منفی بودن ضریب رگرسیون برای اغلب این توصیف‌کننده‌ها، افزایش مقدار آنها سبب کاهش pIC_{50} برای ترکیبات مورد بررسی می‌شود.

۵-۱-۲- توصیف‌کننده‌های *3D-MoRSE* [۴۵ و ۴]

ایده اصلی توصیف کننده‌های 3D-MoRSE، به دست آوردن اطلاعاتی از مختصات اتمی سه بعدی با استفاده از تبدیلی است که در مطالعات پراش الکترونی، برای فراهم آوردن منحنی‌های نظری تفرق کاربرد دارد.

$$I(s) = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} w_i . w_j . \frac{\sin(s.r_{ij})}{s.r_{ij}} \qquad (۵-۵)$$

در این رابطه $I(s)$ شدت الکترون پراکنده شده، w یک ویژگی اتمی، s زاویه پراکندگی، r_{ij} فاصله بین اتم‌های i و j و A تعداد کل اتم‌ها را نشان‌می‌دهد. این توصیف‌کننده‌ها چینش سه بعدی اتم‌ها را نشان می‌دهند و به اندازه مولکول وابسته نیستند، بنابراین برای تعداد زیادی از مولکول‌ها با اندازه‌های متفاوت قابل استفاده هستند.

از این گروه توصیف‌کننده‌ها Mor16v، Mor16m و Mor18m در مدل ارائه شده در فصل چهارم ظاهر شده‌اند.

۵-۱-۳- توصیف‌کننده‌های $BCUT$[۱] [۴۶]

این توصیف‌کننده‌های مولکولی ویژه‌مقادیری[۲] از یک ماتریس ارتباطی تغییر شکل یافته به نام ماتریس بوردن[۳] هستند. ماتریس بوردن یک گراف مولکولی تهی از هیدروژن است که در آن مولفه‌های قطری اعداد اتمی، الکترونگاتیویته و ... و عناصر غیر قطری در ارتباط با مرتبه پیوند دو اتم پیوند شده، هستند.

اگر عناصر قطری این ماتریس شامل اعداد اتمی اتم‌های سازنده مولکول باشند، توالی nتایی مرتبی از کوچکترین ویژه‌مقادیر ماتریس بوردن به عنوان توصیف‌کننده‌های مولکولی با قدرت تفکیک بالا در نظر گرفته می‌شود و برای بازیابی و مرتب‌سازی ساختارهای مولکولی از آنها استفاده می‌شود. فرض اساسی در اینجا این است که کوچکترین ویژه‌مقادیر، محتوی سهم‌هایی برای تمام اتم‌ها هستند و در نتیجه انعکاسی از توپولوژی کل مولکول می‌باشند. توصیف‌کننده BELm7 نیز، که با توجه به وزن اتمی محاسبه گردیده است، از این گروه توصیف‌کننده‌هاست و در مدل ارائه شده در فصل سوم ظاهر شده است. منفی بودن ضریب رگرسیون برای آن نشان می‌دهد که با افزایش مقادیر این توصیف‌کننده توان بازدارندگی افزایش می‌یابد.

حال اگر ماتریس بوردن یک گراف مولکولی شامل اتم‌های هیدروژن باشد عناصر قطری این ماتریس، تقریبی از الکترونگاتیویته اتم‌ها خواهند بود. ماتریسی که شامل اطلاعاتی در مورد محیط الکترونی اطراف اتم‌هاست، باید بیانگر ماتریس ویژه‌مقادیری باشد که توزیع الکترونی را در سراسر مولکول توصیف می‌کند. این ماتریس توالی مرتبی از nتا از بزرگترین ویژه‌مقادیر مطلق از ماتریس بوردن است و براساس الکترونگاتیویته اتم‌هاست. در اینجا مقیاس الکترونگاتیویته بر مبنای کربن است و الکترونگاتیویته کربن برابر با صفر در نظر گرفته می‌شود.

---

۱ Burden-CAS-University of texas eigenvalues

۲ Eigenvalue

۳ Burden matrix

از این گروه توصیف‌کننده‌ها BEHe6 و BEHp8 که محاسبات آنها به ترتیب بر اساس الکترونگاتیویته سندرسون و قطبش‌پذیری اتمی می‌باشد، در مدل رگرسیون خطی چندگانه فصل چهارم به کار گرفته شده‌اند. ضریب رگرسیون BEHe6 مثبت می‌باشد که نشان می‌دهد افزایش آن سبب افزایش pIC_{50} خواهد شد، اما ضریب رگرسیون برای BEHp8 منفی است و با افزایش مقدار آن pIC_{50} کاهش می‌یابد.

۵-۱-۴- توصیف‌کننده‌های *Constitutional* [۴۷]

این گروه از توصیف‌کننده‌ها، ساده‌ترین و معمول‌ترین توصیف‌کننده‌های استفاده شده می‌باشند که خصوصیات مولکول را بدون داشتن اطلاعات زیادی در مورد آن منعکس می‌کنند. تعداد اتم‌ها، تعداد پیوندها، تعداد مطلق و نسبی یک نوع اتم خاص در مولکول، تعداد پیوندهای ساده، دوگانه، سه‌گانه و آروماتیک، تعداد حلقه‌ها، تعداد حلقه‌ها بر اساس تعداد اتم‌های آنها و وزن مولکولی متداول‌ترین توصیف‌کننده‌های ساختاری می‌باشند.

یکی از این توصیف‌کننده‌ها C001 است که در مدل بدست آمده در فصل چهارم ظاهر شده است و بیانگر تعداد شاخه‌های متیل انتهایی در مولکول می‌باشد. با توجه به توصیف‌کننده‌های X3sol و Hy (در ادامه به تشریح آنها خواهیم پرداخت) که در مدل منتخب به کار گرفته شده‌اند می‌توان گفت حضور بیشتر گروه‌های متیل که به معنای شاخه‌دارتر بودن ساختار می‌باشد موجب می‌شود که ترکیب آبگریزتر شده و بهتر بتواند با ساختار پروتئینی وارد برهمکنش شود.

۵-۱-۵- توصیف‌کننده‌های گروه *GETAWAY* [۴۸]

این توصیف‌کننده‌ها که با توجه به مختصات فضایی اتم‌ها در یک مولکول به راحتی قابل محاسبه‌اند با ماتریس قدرت نفوذ[۱] تعریف می‌شوند و بیان‌کننده ویژگی‌های هندسی و توپولوژیکی مولکول‌ها هستند. ماتریس قدرت نفوذ یا ماتریس تاثیر مولکول[۲] (MIM) با رابطه زیر بیان می‌شود:

$$H = M.(M^T.M)^{-1}.M^T \qquad (۶-۵)$$

که M ماتریس مختصات اتمی، T به معنای ماتریس ترنس پوز و H ماتریس قدرت نفوذ است. H یک ماتریس متقارن A×A است که A تعداد اتم‌هاست. عناصر قطری ماتریس H لوریج‌ها نام دارند که هر یک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. بالطبع اتم‌های سطحی اعداد لوریج بزرگتری نسبت به اتم‌های مرکزی دارند. همچنین مولکول‌های کروی اعداد لوریج پایین‌تری نسبت به مولکول‌های کشیده دارند.

توصیف‌کننده‌های GETAWAY به دو زیر مجموعه H و R تقسیم می‌شوند. توصیف‌کننده‌های H-GETAWAY به اندازه و شکل مولکول وابسته بوده بگونه‌ای که با افزایش اندازه اتم و فاصله‌ی اتم از مرکز مولکول، مقدار آنها افزایش می‌یابد.

گروه دیگر توصیف‌کننده‌های R-GETAWAY هستند که بر اساس ماتریس تاثیر– فاصله به صورت زیر تعریف می‌شوند.

$$[R]_{ij} = \left[\frac{\sqrt{h_{ii}.h_{jj}}}{r_{ij}}\right] \qquad i \neq j \qquad (۷-۵)$$

که h_{ii} و h_{jj} لوریج‌های دو اتم i و j و r_{ij} فاصله آن دو است. عناصر قطری این ماتریس صفر است.

این توصیف‌کننده‌ها به ساختار سه بعدی مولکول‌ها حساس بوده و در برگیرنده‌ی اطلاعاتی نظیر شکل، سایز، تقارن مولکول و نحوه توزیع اتم‌ها در مولکول می‌باشند.

---

[۱] Leverage matrix

[۲] Molecular Influence Matrix

از این گروه، توصیف‌کننده‌های R6m در مدل ارائه شده برای ترکیبات سولفونانیلید در فصل سوم و R8v و R3u در مدل ارائه شده برای پیش‌بینی توان بازدارندگی ترکیبات ۵-اکسوپیرولیدین-۳-کربوکسامید در فصل چهارم انتخاب شده‌اند. اندیس‌های m و v نشان می‌دهند که این توصیف‌کننده‌ها به ترتیب بر اساس جرم اتمی و حجم واندروالس اتمی محاسبه شده‌اند و اندیس u بیان می‌دارد که توصیف‌کننده مربوطه با ویژگی خاصی از اتم‌ها محاسبه نشده‌اند.

۵-۱-۶- توصیف‌کننده‌های *RDF*[۱] [۴۹]

می‌توان گفت RDF یا تابع توزیع شعاعی مربوط به یک دسته از اتم‌ها، معادل توزیع احتمال یافتن یک اتم در یک فضای کروی به شعاع R است. رابطه (۵-۸) نحوه محاسبه تابع RDF را نشان می‌دهد:

$$g(R) = f \cdot \sum_{i}^{N-1} \sum_{j>i}^{N} A_i \cdot A_j \cdot e^{-\beta \cdot (R - r_{ij})^2} \qquad (۵-۸)$$

که f یک فاکتور مقیاس، N تعداد اتم‌های، r_{ij} فاصله بین دو اتم i و j و A یک ویژگی اتمی (جرم اتمی) است. β یک فاکتور تسهیل کننده است که توزیع احتمال فاصله بین اتمی را مشخص می‌کند و می‌توان از آن به فاکتور دما برای تعریف تحرک اتمی تعبیر کرد. $g(R)$ در نقاط گسسته‌ای با فواصل معین حساب می‌شود. با مجموعه‌ای از کدهای RDF با ویژگی‌های مختلف اتمی می‌توان ساختار سه بعدی یک مولکول را به طور واضح توصیف کرد. این توصیف‌کننده‌ها اهمیت توزیع اتم‌ها در مولکول روی خواص دارویی ترکیبات مورد نظر را نشان می‌دهند.

در مدل رگرسیون خطی چندگانه در فصل سوم (رابطه (۳-۵)) توصیف‌کننده RDF035u با ضریب منفی وجود دارد که نشان دهنده رابطه معکوس این توصیف‌کننده با مقادیر pIC_{50} است.

[۱] Radial Distribution Function

۵-۱-۷- توصیف‌کننده‌های گروه *Topological*

این توصیف‌کننده‌ها بر اساس نمایش گراف مولکول می‌باشند. در این گراف‌ها هر نقطه نشان‌دهنده یک اتم بوده و خطوط بین نقاط نیز نشان‌دهنده پیوند شیمیایی بین اتم‌ها می‌باشد. معمولا در گراف‌های مولکولی اتم هیدروژن را نشان نمی‌دهند. این توصیف‌کننده‌ها اطلاعاتی راجع به ساختمان، اندازه، شکل، تقارن، شاخه‌دار شدن، نحوه اتصال اتم‌ها و نوع اتم‌های موجود در یک مولکول را در اختیار ما قرار می‌دهند. محاسبه این توصیف‌کننده‌ها به سادگی از روی ساختمان دوبعدی مولکول‌ها امکان‌پذیر می‌باشد.

از این گروه توصیف‌کننده‌ها در مدل ارائه شده در فصل سوم توصیف‌کننده‌ی GGI3 و در مدل ارائه شده در فصل چهارم توصیف‌کننده‌ی JGI5 انتخاب شد که اندیس‌های توپولوژیکی بار نامیده می‌شوند. این توصیف‌کننده‌ها میزان انتقال بار بین هر جفت از اتم‌ها و در نتیجه انتقال بار در کل مولکول را مورد ارزیابی قرار می‌دهند[۵۰].

توصیف‌کننده دیگری از این دسته که در مدل شبکه عصبی در فصل چهارم به کار گرفته شده است X3sol می‌باشد که اندیس‌های ارتباطی حلالپوشی نام دارند. به طور کلی این توصیف‌کننده‌های مولکولی آنتروپی حلالپوشی و برهمکنش‌های پراکندگی در محلول را شرح می‌دهند که به صورت زیر تعریف می‌شود:

$$ {}^{m}\chi_q^s = \frac{1}{2^{m+1}} \cdot \sum_{k=1}^{K} \frac{(\prod_{a=1}^{n} L_a)_k}{(\prod_{a=1}^{n} \delta_a)_k^{1/2}} \qquad (۹-۵) $$

در رابطه (۵-۹) L_a عدد کوانتمی اصلی (۲ برای اتم C، N و O و ۳ برای Si، S و Cl و...) مربوط به aامین اتم در kامین زیرگراف، δ_a درجه‌ی رأس، K تعداد کل زیرگراف‌های mام و n تعداد یال‌ها در زیرگراف می‌باشد. حضور این توصیف‌کننده در مدل نشان دهنده اهمیت حلالپوشی ترکیب در توان

بازدارندگی آن می‌باشد. ضریب رگرسیون برای این توصیف‌کننده مثبت است که نشان می‌دهد با افزایش آن مقدار pIC_{50} افزایش یافته و در نتیجه اثربخشی دارو زیاد می‌شود که با توجه به برهمکنش این ترکیبات با ساختار ماکرومولکولی همچون پروتئین CCR5 این موضوع دور از انتظار نیست.

۵-۱-۸- توصیف‌کننده‌های گروه *WHIM* [۵۱]

این توصیف‌کننده‌ها که به وسیله گروه تحقیقی تودسچینی[۱] و همکارانش توسعه یافته‌اند حاوی اطلاعات سه بعدی هستند و بر اساس تجزیه اجزاء اصلی ماتریس کوواریانس وزن‌دار شده حاصل از کوردینه‌های کارتزین مولکول می‌باشند. اهمیت این توصیف‌کننده‌ها بخاطر تاثیر کنفورماسیون ترکیب در طول برهمکنش با هدف می‌باشد. این توصیف‌کننده‌ها اطلاعات سه بعدی در مورد اندازه، شکل، تقارن و توزیع اتمی در مولکول‌ها را در اختیار ما قرار می‌دهند. رابطه (۵-۱۰) روش کلی محاسبه این توصیف‌کننده‌ها را بیان می‌کند:

$$S_{jk} = \frac{\sum_{i=1}^{A} w_i (q_{ij} - \bar{q}_i)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^{A} w_i} \qquad (۱۰-۵)$$

که S_{jk} کوواریانس وزن‌دار بین کوئوردینه‌های اتمی jام و kام، A تعداد اتم‌ها، w_i وزن iامین اتم، q_{ij} و q_{ik} به ترتیب jامین و kامین کئوردینه‌های iامین اتم و $\bar{q}$ مقدار میانگین مربوطه را نشان می‌دهند.

توصیف‌کننده‌های WHIM به دو دسته مهم تقسیم می‌شوند: توصیف‌کننده‌های جهت‌دار[۲] و توصیف‌کننده‌های کروی[۳]. گروه توصیف‌کننده‌های جهت‌دار توصیف‌کننده‌هایی در مورد اندازه، شکل و

[۱] Todeschini
[۲] Directional WHIM descriptors
[۳] Global WHIM descriptors

تقارن مولکولی در جهت محورهای اصلی از مولکول ارائه می‌دهند ولی اطلاعات توصیف‌کننده‌های کروی با دید کروی از مولکول محاسبه می‌شود.

مدل رگرسیون خطی جهت پیش‌بینی میزان فعالیت دارویی ترکیبات سولفونانیلید در فصل سوم توصیف‌کننده L2m که از دسته توصیف‌کننده‌های WHIMجهت‌دار است و بر اساس وزن اتمی وزن‌دار شده است را انتخاب کرد. این دسته از توصیف‌کننده‌ها اندازه‌های مولکولی را در راستاهای اصلی محاسبه می‌کنند. با توجه به اینکه ضریب رگرسیون برای این توصیف‌کننده منفی است، افزایش ابعاد مولکول موجب کاهش فعالیت دارویی ترکیب می‌گردد.

۵-۱-۹- توصیف‌کننده‌ی *Hy*

این توصیف‌کننده بیانگر فاکتور آبدوستی ترکیب می‌باشد. ضریب منفی برای این فاکتور در مدل حاصل از رگرسیون خطی در فصل چهارم نشان می‌دهد که حضور گروه‌های آبدوست در ترکیبات مورد بررسی موجب کاهش فعالیت بازدارندگی آنها می‌گردد.

۵-۲- بررسی میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی

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

۱. توصیف‌کننده‌ی مورد نظر از سری توصیف‌کننده‌های ارائه شده به شبکه بهینه حذف شد.

۲. با استفاده از توصیف‌کننده‌های باقیمانده مقادیر پاسخ برای داده‌های سری ارزیابی محاسبه گردید.

۳. با استفاده از مقادیر پاسخ محاسبه شده و مقادیر پاسخ تجربی، میانگین خطای مطلق در غیاب توصیف‌کننده‌ی مورد نظر محاسبه شد.

۴. مراحل ۱ تا ۳ برای هریک از توصیف‌کننده‌ها تکرار شد.

۵. در پایان درصد مشارکت هر یک از توصیف‌کننده‌ها به کمک رابطه (۵-۱۱) برآورد شد.

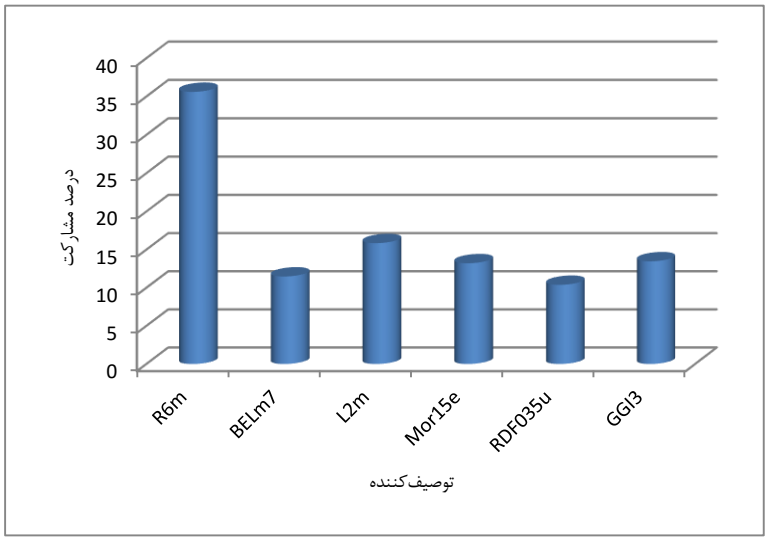
$$c_i = 100 \frac{mae_i}{\sum_{i=1}^{N} mae_i} \qquad (۱۱-۵)$$

در رابطه بالا c_i درصد مشارکت توصیف‌کننده iام، N تعداد توصیف‌کننده‌های به کارگرفته شده در مدل شبکه عصبی و mae_i میانگین خطای مطلق حاصل از سری ارزیابی در غیاب توصیف‌کننده iام می‌باشند[۵۲].

۵-۲-۱- بررسی میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی مشتقات سولفونانیلید

با توجه به اینکه در بهینه‌سازی مدل شبکه عصبی برای مشتقات سولفونانیلید از روش ارزیابی تقاطعی استفاده شده است، میانگین خطای مطلق در رابطه (۵-۱۱)، برای این ترکیبات، از میانگین خطای مطلق حاصل از ارزیابی تقاطعی سری آموزش به جای میانگین خطای مطلق سری ارزیابی استفاده شده است. نمودار شکل ۵-۱ نتایج این محاسبات را نمایش می‌دهد.

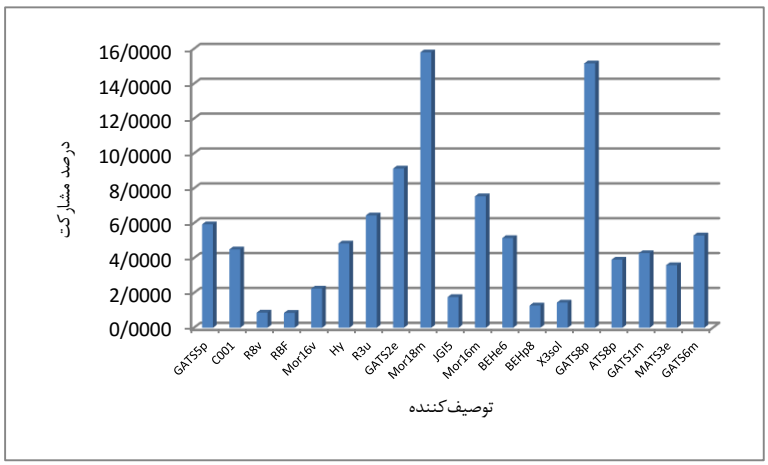
براساس نمودار شکل ۵-۱، توصیف‌کننده R6m بیشترین اثر مشارکت را در مدل ارائه شده نشان می‌دهد. این توصیف‌کننده از دسته توصیف‌کننده‌های GETAWAY می‌باشد که در بر دارنده اطلاعات سه بعدی هندسی و توپولوژیکی مولکول می‌باشند. توصیف‌کننده‌ی بعدی که دارای بیشترین اثر مشارکت می‌باشد L2m است که از دسته توصیف‌کننده‌های WHIMجهت‌دار است و بر اساس وزن اتمی وزن‌دار شده است. این توصیف‌کننده اندازه‌های مولکولی را در راستاهای اصلی محاسبه می‌کند. با توجه به این توصیف‌کننده‌ها می‌توان گفت که توپولوژی و هندسه مولکول در نحوه فعالیت این ترکیبات نقش بسیار مهمی ایفا می‌کند.



شکل ۵-۱- نمودار درصد مشارکت توصیف‌کننده‌های به کار گرفته شده در مدل شبکه عصبی برای ترکیبات سولفونانیلید

۵-۲-۲- بررسی میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی مشتقات ۵-اکسوپیرولیدین-۳-کربوکسامید

جهت بررسی و ارزیابی اثر مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی همانگونه که ذکر شد مقادیر اثر مشارکت برای هر یک از توصیف‌کننده‌ها محاسبه شد. نتایج محاسبات به صورت نمودار در شکل ۵-۲ آورده شده است.



شکل ۵-۲- نمودار درصد مشارکت توصیف‌کننده‌های به کار گرفته شده در مدل شبکه عصبی برای ترکیبات ۵-اکسوپیرولیدین-۳-کربوکسامید

با توجه به نمودار شکل ۵-۲ مشاهده می‌شود که دو توصیف‌کننده‌ی GATS8p و Mor18m دارای بیشترین اثر مشارکت در مدل می‌باشند. GATS8p از دسته توصیف‌کننده‌های 2D-autocorrelation می‌باشد و چگونگی توزیع بار و قطبش‌پذیری در سطح مولکولی را نشان می‌دهد.

۵-۳- بحث و نتیجه‌گیری در مورد خواص بازدارندگی ترکیبات سولفونانیلید

روند نتایج در این تحقیق همانند پروژه‌های مشابه نشان داد که روابط ساختار- فعالیت در مولکول‌های آلی با فعالیت زیستی، روابطی غیرخطی و پیچیده است که نمی‌توان به وسیله مدل‌های خطی آنها را به خوبی توصیف نمود. در این تحقیق از مدل شبکه عصبی مصنوعی با الگوریتم آموزشی تنظیم بایزین برای مطالعات QSAR استفاده شده است که براساس اطلاعات موجود این اولین گزارش در مورد به کار گیری روش فوق برای پیش‌بینی فعالیت بازدارندگی سری مشتقات سولفونانیلید یاد شده در جدول ۳-۱ می‌باشد. نتایج بدست آمده از مدل ارائه شده نشان می‌دهند که این مدل قادر است به خوبی و به طور رضایت بخشی بین ساختار مولکولی و فعالیت بازدارندگی این ترکیبات ارتباط بر قرار نماید. نتایج ارائه شده در جدول ۳-۱۰ نشان می‌دهند که مدل شبکه عصبی در پیش‌بینی مقادیر $\log(1/IC_{50})$ برای ترکیبات سری تست که قبلا آنها را ندیده است بسیار توانمند ظاهر می‌شود و در مقایسه با مدل رگرسیون خطی چندگانه نتایج بهتری را ارائه می‌نماید..

۵-۴- بحث و نتیجه‌گیری در مورد خواص بازدارندگی ترکیبات ۵-اکسوپیرولیدین-۳-کربوکسامید

با توجه به اینکه تاکنون مطالعات QSAR ترکیبات بازدارنده‌ی پروتئین CCR5، ۵-اکسوپیرولیدین-۳-کربوکسامید گزارش نشده است و این اولین گزارش در مورد به کار گیری شبکه عصبی مصنوعی با الگوریتم آموزشی لونبرگ-مارکورت در مورد ترکیبات فوق می‌باشد، این مدل می‌تواند به محققان در طراحی ترکیبات جدید از این دسته با استخلاف‌های جدید کمک فراوانی نماید. نتایج به دست آمده برای این سری از ترکیبات و تعداد ۱۹ توصیف‌کننده‌ایی که در مدل شبکه عصبی، به منظور ایجاد مدلی کارآمد در پیش‌بینی توان بازدارندگی این ترکیبات، مورد استفاده قرار گرفته‌اند، نشان می‌دهند که رابطه بسیار پیچیده‌ای میان ساختار و فعالیت دارویی آنها وجود دارد که با مدل‌های خطی ساده نمی‌توان آنها را به خوبی توصیف نمود. از این رو در این نوشتار نتایج مربوط به مدل خطی که با این تعداد توصیف‌کننده از نظر علمی اهمیت چندانی ندارند آورده نشده‌اند.

۵-۵- آینده‌نگری

استفاده از مدل‌های QSAR می‌تواند منجر به صرفه‌جویی زیادی در وقت و هزینه در راستای طراحی و بهینه‌سازی ساختار داروها شود. به کمک نتایج حاصل از این پروژه می‌توان به طراحی و آزمایش ترکیبات جدید با ویژگی بازدارندگی قوی‌تر فعالیت آنزیمی آروماتازها و نیز مهار نمودن برهمکنش پروتئین CCR5 با ویروس HIV-1 پرداخت. با توجه به توصیف‌کننده‌های به کار گرفته شده در مدل‌های ارائه شده و با استفاده از میزان مشارکت آنها در مدل می‌توان به جنبه‌هایی که موجب تقویت خاصیت دارویی مورد نظر می‌شوند پی برد و با استفاده از گروه‌های استخلافی مناسب به مشتقاتی از این ترکیبات دست یافت که دارای توان دارویی بالاتری نسبت به سایر ترکیبات باشند.

صحت و دقت نتایج مدل‌سازی QSAR به دقت داده‌های تجربی در مورد ساختار و فعالیت مولکلول‌های مورد استفاده در سری داده‌ها وابستگی زیادی دارد. تعداد ترکیبات نیز در سری داده‌ها می‌تواند در میزان درستی نتایج تاثیر بسزایی داشته باشد. بنابراین می‌توان با افزایش تعداد ترکیبات سری داده‌ها مدل‌های ارائه شده را بسط داده و به مدل‌های دقیق‌تر با قدرت پیش‌بینی بالاتری دست پیدا کرد.

با توجه به اینکه انتخاب توصیف‌کننده‌های مهم در مدل‌های ارائه شده به کمک روش رگرسیون خطی مرحله‌ای انجام گرفته است، تنها برهمکنش‌های خطی در روابط ساختار-فعالیت ترکیبات لحاظ شده است، لذا می‌توان با استفاده از روش‌های دیگر از جمله الگوریتم ژنتیک به انتخاب توصیف‌کننده‌ها اقدام نمود و نتایج را با نتایج ارائه شده در این تحقیق مقایسه کرد.

مراجع و 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]. منصور عرب چم‌جنگلی (۱۳۸۶)، "پیش‌بینی فعالیت دارویی ضد ایدز (سیتوتوکسیتی) مشتقات ۵-فنیل -۱-فنیل آمینو۱H- ایمیدازول بوسیله شبکه عصبی مصنوعی"، دانشگاه صنعتی شاهرود، گزارش طرح پژوهشی.

[۴]. M. Arab Chamjangali, M. Beglari, G. Bagherian (2007), “Prediction of cytotoxicity data (CC_{50}) of anti-HIV 5-pheny-1-phenylamino-1H-imidazole derivatives by artificial neural network trained with Levenberg-Marquardt algorithm”, *J. Mol. Graphics Modell.*, 26,pp 360-367.

[۵].Todeschini R. and Consonni V. (2000), “Handbook of Molecular Descriptors”, John Wiley, New York, pp 98.

[۶]. J. A. Pople, D. Beveridge (1970), “Approximate Molecular Orbital Theory”, McGraw-Hill.

[۷]. Wayne W. Daniel (1999), “Biostatistics Afoundation for Analysis in the Health Science”, John WILEY.

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

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

[۱۰]. Todeschini R. and Consonni V. (2000), “Handbook of Molecular Descriptors”, John Wiley, New York, pp 307.

[۱۱]. Todeschini R. and Consonni V. (2000), “Handbook of Molecular Descriptors”, John Wiley, New York, pp 304.

[۱۲]. Pitts W., MacCullouch S. (1947) “On how we know universals: The perception of auditory and visual forms”, *Bull. Math. Biophysic*. 9, pp 127-147.

[۱۳]. Wayne W. Daniel (1999), “Biostatistics Afoundation for Analysis in the Health Science”, John WILEY.

[۱۴]. جنی ر. بینز و. "هوش مصنوعی از الف تا ی" (۱۳۷۶)، بیگدلی قمی س. محسنی م. بدیع ک.، دفتر تحقیقات یاسین، ص ۲۲۴- ۲۲۸.

[۱۵]. منهاج م."مبانی شبکه‌های عصبی هوش محاسباتی"، جلد اول (۱۳۸۷)، چاپ پنجم، انتشارات دانشگاه صنعتی امیرکبیر، ص ۲۵-۲۷.

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

[۱۷].جنی ر. بینز و. "هوش مصنوعی از الف تا ی" (۱۳۷۶)، بیگدلی قمی س. محسنی م. بدیع ک.، دفتر تحقیقات یاسین، ص ۱۰۵.

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

[۱۹]. Soma Samanta, K. Srikanth, Suchandra Banerjee, BikashDebnath, ShovanlalGayenand TarunJha (2004), “5-N-Substituted-2-(substituted benzenesulphonyl) glutamines asantitumor agents. Part II: Synthesis, biological activity and QSAR study”. *Bioorgan. Med. Chem.*, 12, pp 1413–1423.

[۲۰]. Shuzhong Zhang, Xinning Yang, Robert A. Coburn, Marilyn E. Morris (2005), “Structure activity relationships and quantitative structure activity relationships for the flavonoid-mediated inhibition of breast cancer resistance protein”, *Biochem. Pharmacol.*, 70(4), pp 627-639.

[۲۱]. Alan R. Katritzky, MinatiKuanar, Dimittar A. Dobchev, Barbara W. A. Vanhoecke, MatiKarelson, Virinder S. Parmar, Christian V. Stevense and Marc E.

Bracke (2006), “QSAR modeling of anti-invasive activity of organic compounds using structural descriptors”, *Bioorgan. Med. Chem.*, 14, pp 6933-6939.

[۲۲]. Katarina M. Nikolic (2007), “QSAR study of α-tocopherol derivatives with chemotherapeutic activity against human breast cancer cells”, *J. Mol. Struct.*, 809(1-3), pp 137-143.

[۲۳]. K.M. Nikolic (2008), “Design and QSAR study of analogs of α-tocopherol with enhanced antiproliferative activity against human breast adenocarcinoma cells”, *J. Mol. Graphics Modell.*, 26(5), pp 868-873.

[۲۴]. Katarina Nikolic, DanicaAgababa (2009), “Design and QSAR study of analogs of γ-tocotrienol with enhanced antiproliferative activity against human breast cancer cells”, *J. Mol. Graphics Modell.*, 27(7), pp 777-783.

[۲۵]. RaziehSabet, MehrdadMohammadpour, Amir Sadeghi, AfshinFassihi (2010), “QSAR study of isatin analogues as in vitro anti-cancer agents”, *Europ. J. Med. Chem.*, 45, pp 1113–1118.

[۲۶]. J. M. J. Tronchet, M. Grigorov, N. Dolatshahi, F. Moriaud, J. Weber (1997), “A QSAR study conforming the heterogeneity of the HEPT derivative series regarding their interaction with HIV reverse transcriptase”, *Europ. J. Med. Chem.*, 32, pp 279.

[۲۷]. Jalali-Heravi, M. and Parastar, F. (2000), “Use of Artificial Neural Network in a QSAR Study of Anti-HIV Activity for a Large Group of HETP Derivatives”, *J. Chem. Inf. Comput.Sci.*, 40, pp 147-154.

[۲۸]. F.A. Molfetta, C.N. Alves, A.B.F. da Silva (2002), ‘A quantum chemical and statistical study of biflavonoid compounds with anti-HIV activity”, *J. Mol. Struct. Theochem*, 577, pp 187-195.

[۲۹]. Dana Weekes, Gary B. Fogel (2003), “Evolutionary optimization, backpropagation, and data preparation issues in QSAR modeling of HIV inhibition by HETP derivatives”, *BioSystems*, 72, pp 149-158.

[۳۰]. Shovanlal Gayen, Bikash Debnath, Soma Samanta and TarunJha (2004), “QSAR study on some anti-HIV HEPT analogues using physicochemical and topological parameters”, *Bioorgan. Med. Chem.*, 12, pp 1493–1503.

[۳۱]. Sayyed Hamed Sadat Hayatshahi, Parviz Abdolmaleki, Shahrokh Safarian, Khosro Khajeh (2005), ‘Non-linear quantitative structure-activity relationship for adenine derivatives as competitive inhibitors of adenosine deaminase”, *Biochem. Bioph. Res. Co.,* 338, pp 1138-1142.

[۳۲]. Pablo R. Duchowicz, Michael Fernandez, Julio Caballero, Eduardo A. Castro and Francisco M. Fernandez (2006), “QSAR for non-nucleoside inhibitors of HIV-1 reverse transcriptase”, *Bioorgan. Med. Chem.*, 14, pp 5876-5889.

[۳۳]. David Hecht, Mars Cheung, Gary B. Fogel (2008),”QSAR using evolved neural networks for the inhibition of mutant *Pf*DHFR by pyrimethamine derivatives”, *BioSystems*, 92, pp10-15.

[۳۴]. Yong-Hong Liang, Fen-Er Chen (2009), “QSAR studies for diarylpyrimidines against HIV-1 reverse transcriptase wild-type and mutant strains”, *Europ. J. Med. Chem.,* 44, pp 625-631.

[۳۵]. M. Arab Chamjangali (2009), “Modelling of Cytotoxicity Data (CC50) of Anti-HIV 1-[5-Chlorophenyl) Sulfonyl]-1H-Pyrrole Derivatives Using Calculated Molecular Descriptors and Levenberg–Marquardt Artificial Neural Network”, *Chem. Biol. Drug Des.*, 73 (4), pp 456-465.

[۳۶]. Bin Su, Ran Tian, Michael V. Darby, and Robert W. Brueggemeier (2008), “Novel Sulfonanilide Analogs Decrease Aromatase Activity in Breast Cancer Cells: Synthesis, Biological Evaluation, and Ligand-Based Pharmacophore Identification”, *J. Med. Chem.,* 51 (5), pp 1126-1135.

[۳۷]. Todeschini R., Milano Chemometrics and QSAR Group.

http://www.disat.unimib.it

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

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

[۴۰]. MATLAB 7.7 (2008b), The Math Works, Inc., Natick, MA, USA.

[۴۱]. Shinichi Imamura, Yuji Ishihara, Taeko Hattori, Osamu Kurasawa, Yoshihiro Matsushita, Yoshihiro Sugihara, NaoyukiKanzaki, Yuji Iizawa, Masanori Baba, Shohei Hashiguchi (2004), "CCR5 Antagonists as Anti-HIV-1 Agents. 1. Synthesis and Biological Evaluation of 5-Oxopyrrolidine-3-carboxamide Derivatives", *Chem. Pharm. Bull.* 52(1), pp 63-73.

[۴۲]. Shinichi Imamura, Osamu Kurasawa, Yoshi Nara, Takashi Ichikawa, Youichi Nishikawa, Takehiro Iida, Shohei Hashiguchi, NaoyukiKanzaki, YujiIizawa, Masanori Babab and Yoshihiro Sugihara (2004), "CCR5 antagonists as anti-HIV-1 agents. Part 2: Synthesis and biological evaluation of N-[3-(4-benzylpiperidin-1-yl)propyl]-N, N0-diphenylureas", *Bioorgan. Med. Chem.*,12, pp 2295-2306.

[۴۳]. Shinichi Imamura, Youichi Nishikawa, Takashi Ichikawa, Taeko Hattori, Yoshihiro Matsushita, Shohei Hashiguchi, NaoyukiKanzaki, Yuji Iizawa, Masanori Baba and Yoshihiro Sugihara (2005) "CCR5 antagonists as anti-HIV-1 agents. Part 3: Synthesis and biological evaluation of piperidine-4-carboxamide derivatives", *Bioorgan. Med. Chem.*, 13, pp 397-416.

[۴۴]. Roberto Todeschini, Viviana Consonni, (2000), "Handbook of molecular Descriptors", WILEY-VCH, pp 17-19.

[۴۵]. Roberto Todeschini, Viviana Consonni, (2000), "Handbook of molecular Descriptors", WILEY-VCH, pp 513-514.

[۴۶]. Roberto Todeschini, Viviana Consonni, (2000), "Handbook of molecular Descriptors", WILEY-VCH, pp 132-133.

[۴۷]. Roberto Todeschini, Viviana Consonni, (2000), "Handbook of molecular Descriptors", WILEY-VCH, pp 90-91.

[۴۸]. Consonni V, Todeschini R., Pavan M. (2002), "Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors", *J. Chem. Inf. Comput. Sci*, 42, pp 682-692.

[۴۹]. Roberto Todeschini, Viviana Consonni (2000), “Handbook of molecular Descriptors”, WILEY-VCH, pp 366-367.

[۵۰]. Roberto Todeschini, Viviana Consonni (2000), “Handbook of molecular Descriptors”, WILEY-VCH, pp 445-446.

[۵۱]. Roberto Todeschini, Viviana Consonni (2000), “Handbook of molecular Descriptors”, WILEY-VCH, pp 492-497

[۵۲]. Julio Caballero, Miguel Garrigab and Michael Fernandez (2006), “2D Autocorrelation modeling of the negative inotropic activity of calcium entry blockers using Bayesian-regularized genetic neural networks”, *Bioorgan. Med. Chem.*, 14, pp 3330–3340

Abstract

Recent investigations show that estrogen plays important role in breast cancer disease. Thus some researches have been carried out on compounds which can inhibit pathways of this hormone in body at some ways. These studies are focused on the two categories of compounds. The first group of compounds directly influence on estrogen activities. The second group of compounds inhibits the estrogen production procedures. One group of these compounds that can inhibit estrogen production is Sufonanilide derivatives. In the first section of this thesis, QSAR models for 21 analogues of Sulfonanilide were constructed using MLR and Bayesian regularized ANN (BR-ANN). Among of a large number of calculated descriptors only 20 significant molecular descriptors were obtained by stepwise regression, as the most feasible descriptors, and then 6 descriptors were used as inputs for ANN. The data set was randomly divided into train (16 molecules) and test (5 molecules) sets and the models parameters and ANN architecture were optimized by cross-validation method. The prediction ability of each model was evaluated using the test data set and Leave-One-Out cross-validation (LOO) Method. The MSEs for the test data set and LOO method were 0.0227 and 0.0132 for MLR, respectively. The MSEs values obtained from applying ANN model to test set and cross-validation were 0.0098 and 0.0161, respectively. The results obtained from ANN showed the excellent prediction of the inhibitory activity data of the corresponding analogues.

In the second section, quantitative-structure activity of some 5-oxopyrrolidine-3-carboxamide derivatives was studied. These compounds suppress disease by interfering in interaction between CCR5 and HIV. Data set consist of 104 analogues of 5-oxopyrrolidine-3-carboxamide. In this study, data set was divided into 62 trains, 21 validations and 21 test sets. The ANN model for prediction of pIC_{50} was constructed using 19 descriptors. MSE values for test and LOO method were 0.1210 and 0.1405, respectively.

Key word: QSAR, ANN, MLR, Sulfonanilide..

Shahrood University of Technology
Faculty of Chemistry

Quantitative Structure-Activity study of some Sulfonanilide derivatives as a new group of anti-cancer drugs and anti-HIV activity of some new compounds

Mahdi Mohamad Rezaei

Superviser:
Dr. M. Arab Chamjangali

Advisor:
Dr. Z. Kalntar

Date: December – 2010