

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده شیمی

رساله دکتری شیمی تجزیه

استفاده از توصیف‌گرهای مولکولی جدید در ارائه مدل‌های روابط کمی ساختار-فعالیت

(QSAR) برخی از ترکیبات شبه دارویی

نگارنده:

مژگان بگلری امنیه

استاد راهنما:

دکتر ناصر گودرزی

استاد مشاور:

دکتر داود شاهسونی

## تقدیم:

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

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

به پاس قلب های بزرگشان که فریاد رس است و سرگردانی و ترس در پناهندگان به شجاعت می گراید

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

در پایان این مجموعه راه پدر و مادر عزیزم، همسر مهربان و دلسوزم و امیدهای زندگیم کیمیا و پرللم نازنینم تقدیم می کنم،

که با حضور و همراهی شان همیشه راه را به من نشان داده و مراد این راه استوار و ثابت قدم نموده اند.

## قدردانی و تشکر:

اکنون که به یاری پروردگار و یاری و راهنمایی اساتید بزرگ، موفق به پایان این رساله شده‌ام و وظیفه خود دانسته که نهایت سپاسگزاری را از تمامی عزیزانی که در این راه به من کمک کرده‌اند را به عمل آورم:

در آغاز از استاد کرامت‌آورد "جناب آقای دکتر کوردزی" که راهنمایی این رساله را به عهده داشته‌اند کمال تشکر را دارم.

از استاد گرامی "جناب آقای دکتر داود شاهسونی" که مشاور این رساله بوده‌اند نیز قدردانی می‌نمایم.

از داوران گرامی جناب آقای دکتر محمد حسین فاطمی، جناب آقای دکتر قدح علی باقریان، دهنقی و سرکار خانم

دکتر فاطمه مصدر الامور که زحمت داوری این رساله را به عهده داشتند کمال سپاس را دارم.

خالصانه از همسر عزیزم جناب "آقای دکتر منصور عرب‌چم‌جنگلی" که وجودشان همیشه قوتی برای انجام کارهایم بوده

است و بدون شک انجام این رساله بدون کمک و راهنمایی‌های ارزنده ایشان امکان پذیر نبوده است، کمال تشکر را

دارم.

از کلیه هم‌دانشگاهیان و همراهان عزیز، به خصوص دوستان خوبم خانم مریم مظفری و خانم دکتر راحله دوستی

نهایت سپاس را دارم.

## تعهد نامه

اینجانب **مژگان بگلری امنیه** دانشجوی دوره دکتری شیمی تجزیه دانشکده شیمی دانشگاه صنعتی شاهرود، نویسنده رساله استفاده از توصیف‌گرهای مولکولی جدید در ارائه مدل‌های روابط کمی ساختار-فعالیت (QSAR) برخی از ترکیبات شبه دارویی تحت راهنمایی آقای دکتر ناصر گودرزی و مشاوره دکتر داود شاهسونی متعهد می‌شوم :

- تحقیقات در این پایان نامه 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 پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد 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 پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده ( یا بافتهای آنها ) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

امضای دانشجو

تاریخ

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است ) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

## چکیده

به منظور ساخت مدل‌های QSAR برای پیش‌بینی فعالیت دارویی سه دسته از ترکیبات دارویی، از مطالعات داکینگ مولکولی استفاده شد و توصیف‌کننده‌های مولکولی حاصل از برهم‌کنش‌های آنزیم-لیگاند دارویی با استفاده از نرم‌افزار Autodock4.2 استخراج شدند. با توجه به اینکه توصیف‌کننده‌های داکینگ مولکولی به تنهایی بیانگر ویژگی‌های ساختاری لیگاندهای دارویی نمی‌باشند و همچنین ارتباط قوی بین توصیف‌کننده‌های داکینگ مولکولی-فعالیت دارویی وجود ندارد، استفاده از توصیف‌کننده‌های ساختاری (محاسبه شده به وسیله نرم‌افزار DRAGON) علاوه بر توصیف‌کننده‌های داکینگ مولکولی ضروری می‌باشد. در بخش اول، مطالعه ساختار-فعالیت ۴۳ مشتق ۴-آنیلینو کوئینازولین به عنوان بازدارنده‌های سرطان مورد بررسی قرار گرفت. ۱۰ توصیف‌کننده‌ی داکینگ مولکولی برای همه‌ی ترکیبات استخراج شد. سری داده‌ها به دو دسته آموزش (۳۵ ترکیب) و آزمون (۸ ترکیب) تقسیم شد. رگرسیون گام به گام به عنوان روش انتخاب متغیر روی داده‌های سری آموزش اعمال شد. مدل شبکه عصبی مصنوعی با دو توصیف‌کننده منتخب توسعه یافت. ضریب تعیین ( $R^2$ ) پایین و میانگین مربعات خطای (MSE) بالای سری آزمون (به ترتیب برابر با ۰/۵۲ و ۰/۶۶) به دست آمده دلالت بر قدرت پیش‌بینی نامناسب مدل شبکه عصبی مصنوعی دارد. به منظور بهبود نتایج مدل سازی شبکه عصبی مصنوعی، توصیف‌کننده‌های ساختاری تابع توزیع شعاعی (RDF) به توصیف‌کننده‌های داکینگ مولکولی اضافه شدند. توصیف‌کننده‌های RDF بیانگر اختلاف در توزیع اتم‌ها در مولکول و اثر این توزیع بر روی فعالیت بازدارندگی ترکیبات می‌باشند. روش رگرسیون گام به گام روی ۱۶۰ توصیف‌کننده (۱۵۰ توصیف‌کننده تابع توزیع شعاعی و ۱۰ توصیف‌کننده داکینگ مولکولی) اجرا شد و ۸ توصیف‌کننده موثر انتخاب شدند و به عنوان ورودی‌های مدل شبکه عصبی مورد استفاده قرار گرفتند.  $R^2$  بالا و MSE پایین سری آزمون (به ترتیب برابر با ۰/۹۰ و ۰/۱۵)، قدرت پیش‌بینی و تعمیم‌پذیری بالای مدل شبکه عصبی توسعه یافته با مخلوط توصیف‌کننده‌های داکینگ مولکولی-ساختاری را اثبات می‌کنند. در بخش دوم، مطالعه ساختار-فعالیت ۷۳ بازدارنده ایدز با مشابهت ساختاری با دی‌آریل پیریمیدین انجام شد. پس از استخراج توصیف‌کننده‌های داکینگ مولکولی، سری داده‌ها به سه دسته آموزش (۵۳ ترکیب)، ارزیابی (۱۰ ترکیب) و آزمون (۱۰ ترکیب) تقسیم شد. دو توصیف‌کننده منتخب روش رگرسیون گام به گام وارد مدل شبکه عصبی شد و همه پارامترهای شبکه عصبی بهینه شد. پارامترهای  $R^2$  و MSE سری آزمون به ترتیب برابر با ۰/۵۰ و ۰/۲۴ به دست آمد. به منظور بهبود مدل شبکه عصبی، توصیف‌کننده‌های ساختاری شمارش گروه عاملی

برای بازدارنده‌های HIV محاسبه شدند. روش رگرسیون گام به گام روی ۱۶۴ توصیف کننده (۱۵۴ توصیف کننده شمارش گروه عاملی و ۱۰ توصیف کننده داکینگ مولکولی) اعمال شد و مدل شبکه عصبی با ۷ توصیف کننده منتخب توسعه یافت. فعالیت دارویی داده‌های سری آزمون در مدل بهینه شبکه عصبی پیش بینی شد.  $R^2$  و MSE سری آزمون به ترتیب برابر با ۰/۸۹ و ۰/۱۶ به دست آمد که نشان دهنده بهبود نتایج مدل شبکه عصبی با مخلوط توصیف کننده های داکینگ مولکولی- ساختاری می‌باشد. در بخش سوم، ارتباط کمی ساختار- فعالیت ۷۳ مشتق آذینی به عنوان بازدارنده های HIV، بررسی شد. به این منظور، ۱۰ توصیف کننده‌ی داکینگ مولکولی با استفاده از داکینگ ترکیبات مورد مطالعه در جایگاه فعال پروتئین به دست آمدند. سری داده‌ها به سه دسته آموزش (۵۳ ترکیب)، ارزیابی (۱۰ ترکیب) و آزمون (۱۰ ترکیب) تقسیم شد. از بین ۱۰ توصیف کننده داکینگ مولکولی ۲ توصیف کننده به عنوان موثرترین توصیف کننده‌ها با استفاده از روش رگرسیون گام به گام بر روی سری آموزش انتخاب شد و مدل شبکه عصبی مصنوعی با دو توصیف کننده‌ی داکینگ مولکولی بهینه شد. مقادیر فعالیت دارویی ترکیبات سری آزمون با استفاده از مدل بهینه شبکه عصبی پیش بینی شد. پارامترهای  $R^2$  و MSE سری آزمون به ترتیب برابر ۰/۷۹ و ۰/۲۲ به دست آمد. سپس به منظور بهبود مدل شبکه عصبی مصنوعی، ۱۶ شاخص شبه دارویی برای ترکیبات مورد مطالعه محاسبه شد و به ۱۰ توصیف کننده داکینگ مولکولی اضافه گردید. رگرسیون گام به گام برای انتخاب موثرترین توصیف کننده‌ها بر روی ترکیبات سری آموزش اجرا شد. ۴ توصیف کننده‌ی مهم از مجموع توصیف کننده داکینگ مولکولی و شاخص‌های شبه دارویی انتخاب و به عنوان ورودی شبکه عصبی مصنوعی استفاده گردید. فعالیت دارویی داده‌های سری آزمون با استفاده از مدل شبکه عصبی بهینه پیش بینی شد. پارامترهای  $R^2$  و MSE سری آزمون به ترتیب برابر با ۰/۸۶ و ۰/۱۱ به دست آمد. بنابراین استفاده همزمان از توصیف کننده‌های داکینگ مولکولی و شاخص‌های شبه دارویی، کارایی مدل شبکه عصبی مصنوعی را بهبود می‌بخشد.

**کلمات کلیدی:** QSAR، ضد سرطان، ضد ایدز، داکینگ مولکولی، رگرسیون گام به گام، شبکه عصبی

مصنوعی

نتایج حاصل از این رساله در مجلات خارجی و کنفرانس های داخلی، تحت عناوین زیر به چاپ رسیده است:

1. Beglari, M., Goudarzi, N., Shahsavani, D., Chamjangali, M. A., & Mozafari, Z. (2020). Combination of radial distribution functions as structural descriptors with ligand-receptor interaction information in the QSAR study of some 4-anilinoquinazoline derivatives as potent EGFR inhibitors. *Structural Chemistry*, 31, 1481–1491.
2. Beglari, M., Goudarzi, N., Shahsavani, D., Chamjangali, M. A., & Dousti, R. (2020). QSAR modeling of anti-HIV activity for DAPY-like derivatives using the mixture of ligand-receptor binding information and functional group features as a new class of descriptors. *Network modeling analysis in health informatics and bioinformatics*, 9(59), 1-15.
3. Beglari, M., Goudarzi, N., Shahsavani, D., Chamjangali, M. A., & Mozafari, Z. (2020). LM-ANN based QSAR model for the prediction of pEC<sub>50</sub> for a set of potent NNRTI using the mixture of ligand-receptor interaction information and drug-like indexes. *Network modeling analysis in health informatics and bioinformatics*, 9(53), 1-14.
4. Beglari, M., Goudarzi, N., Shahsavani, D., Chamjangali, M. A., & Mozafari, Z. Application of random forest regression in the modelling of a newly synthesized compound as potent HIV-1 reverse transcriptase inhibitors, 6th Iranian Biennial Chemometrics Seminar, University of Mazandaran, Babolsar, Iran, 26-27 Oct 2017
5. Beglari, M., Goudarzi, N., Shahsavani, D., Chamjangali, M. A., & Dousti, R. Quantitative structure activity relationship study of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors using scaled conjugate gradient artificial neural network, 6th Iranian Biennial Chemometrics Seminar, University of Mazandaran, Babolsar, Iran, 26-27 Oct 2017
6. Beglari, M., Goudarzi, N., Shahsavani, D., & Mozafari, Z. Quantitative structure activity relationship study of azine derivatives as NNRTIs using artificial neural network, 7th Iranian Biennial Chemometrics Seminar Shahrood University of Technology, Shahrood, Semnan, Iran 30-31 Oct 2019
7. Beglari, M., Goudarzi, N., Shahsavani, D., & Mozafari, Z., Multiple implementation of MARS as a new descriptor selection method in the QSAR study of a new NNRTIs using artificial neural network, 7th Iranian Biennial Chemometrics Seminar Shahrood University of Technology, Shahrood, Semnan, Iran 30-31 Oct 2019
8. Beglari, M., Goudarzi, N., Shahsavani, D., & Dousti, R., Quantitative structure activity relationship study of DAPY-like derivatives as NNRTIs using artificial neural network, 7th Iranian Biennial Chemometrics Seminar Shahrood University of Technology, Shahrood, Semnan, Iran 30-31 Oct 2019



## فهرست مطالب :

۱	۱- مقدمه
۳	۱-۱ پروتئین
۴	۲-۱ دارو
۵	۳-۱ مکانیسم اتصال دارو-گیرنده
۷	۴-۱ بیماری سرطان و اهمیت پیش‌بینی داروهای ضد سرطان
۸	۱-۴-۱ گیرنده فاکتور رشد اپیدرمی
۹	۲-۴-۱ مکانیسم عمل گیرنده فاکتور رشد اپیدرمی
۱۰	۳-۴-۱ گیرنده فاکتور رشد اپیدرمی و رابطه‌ی آن با سرطان
۱۱	۴-۴-۱ مشتقات کوئینازولین
۱۱	۵-۱ بیماری ایدز و اهمیت پیش‌بینی داروهای ضد ایدز
۱۱	۱-۵-۱ ایدز
۱۲	۲-۵-۱ ساختمان ویروس HIV
۱۳	۳-۵-۱ چرخه تکثیر ویروس HIV
۱۴	۴-۵-۱ اهداف استفاده از داروهای ضد HIV
۱۴	۵-۵-۱ درمان ضد رترو ویروسی HIV
۱۴	۶-۵-۱ انواع داروهای ضد رترو ویروسی
۱۵	۱-۶-۵-۱ بازدارنده‌های ورودی
۱۵	۲-۶-۵-۱ بازدارنده‌های نوکلئوزیدی آنزیم نسخه‌بردار معکوس
۱۷	۳-۶-۵-۱ بازدارنده‌های غیرنوکلئوزیدی نسخه‌بردار معکوس
۱۷	۴-۶-۵-۱ بازدارنده‌های پروتئاز
۱۸	۶-۱ مشتقات آزینی
۱۹	۷-۱ مشتقات دی آریل پیریمیدین‌ها و مشابه‌های آن‌ها
۲۰	۸-۱ مروری بر کارهای پیشین
۲۵	۲- فصل دوم مبانی ساختار - فعالیت
۲۶	۱-۲ مقدمه
۲۸	۲-۲ ارتباط کمی ساختار-فعالیت (QSAR)
۲۸	۱-۲-۲ تعریف
۲۸	۲-۲-۲ طبقه بندی بر اساس نوع روش‌های کمومتریکس
۲۹	۳-۲-۲ طبقه بندی بر اساس نوع اطلاعات مورد استفاده در مدل‌سازی
۲۹	۱-۳-۲-۲ QSAR مبتنی بر لیگاند
۲۹	۲-۳-۲-۲ QSAR مبتنی بر ساختار پروتئین
۳۰	۳-۲ برهمکنش پروتئین و لیگاند
۳۱	۴-۲ شبیه سازی داکینگ مولکولی
۳۳	۵-۲ مراحل انجام داکینگ مولکولی
۳۳	۱-۵-۲ فرایند اعتبارسنجی
۳۴	۲-۵-۲ آماده سازی پروتئین
۳۵	۳-۵-۲ ساخت لیگاند
۳۶	۴-۵-۲ تنظیم کردن جعبه شبکه‌ای

۳۷	۵-۵-۲	گزینه‌های داکینگ
۳۷	۶-۵-۲	انجام محاسبه‌ی داکینگ
۳۷	۷-۵-۲	آنالیز و تحلیل نتایج
۴۰	۸-۵-۲	انواع داکینگ مولکولی
۴۰	۹-۵-۲	کاربردهای داکینگ مولکولی
۴۱	۶-۲	اصول مدل سازی QSAR
۴۲	۱-۶-۲	جمع آوری و انتخاب سری داده‌ها
۴۲	۲-۶-۲	رسم و بهینه سازی ساختار ترکیبات
۴۳	۳-۶-۲	محاسبه و غربالگری توصیف کننده‌ها
۴۵	۴-۶-۲	انتخاب موثرترین توصیف کننده‌ها
۴۶	۱-۴-۶-۲	انتخاب متغیر به روش رگرسیون خطی
۴۷	۵-۶-۲	ایجاد مدل‌های آماری
۴۸	۶-۶-۲	مقدمه‌ای بر شبکه عصبی مصنوعی
۴۸	۱-۶-۶-۲	ساختمان و عملکرد نرون مصنوعی
۴۹	۲-۶-۶-۲	تابع انتقال
۴۹	۷-۶-۶-۲	آموزش شبکه‌های جلو سو با تکنیک پس انتشار
۵۰	۷-۲	ارزیابی مدل
۵۰	۱-۷-۲	استفاده از پارامترهای آماری
۵۵	۲-۷-۲	استفاده از نمودار خطای باقی مانده‌ها
۵۵	۳-۷-۲	استفاده از سری آزمون
۵۵	۴-۷-۲	تکنیک رد مرحله‌ای تک تک
۵۶	۵-۷-۲	آزمون پراکندگی
۵۶	۶-۷-۲	قلمرو کاربرد
۵۷	۷-۷-۲	بررسی میزان مشارکت توصیف کننده‌های منتخب در شبکه عصبی
۵۹	۳-	فصل سوم روش
۶۰	۱-۳	نرم افزارهای مورد استفاده
۶۰	۱-۱-۳	بسته نرم افزاری Hyperchem
۶۰	۲-۱-۳	نرم افزار DRAGON
۶۰	۳-۱-۳	نرم افزار AutoDock4.2
۶۱	۴-۱-۳	نرم افزار ViewerLite5.0
۶۱	۵-۱-۳	نرم افزار Notepad++
۶۲	۶-۱-۳	نرم افزار SPSS
۶۲	۷-۱-۳	نرم افزار MATLAB
۶۳	۲-۳	رسم و بهینه سازی ساختار ترکیبات مورد مطالعه
۶۳	۳-۳	استخراج توصیف کننده‌های حاصل از داکینگ مولکولی
۶۳	۱-۳-۳	استخراج و آماده سازی لیگاند موجود در پروتئین
۶۴	۲-۳-۳	آماده سازی پروتئین
۶۴	۳-۳-۳	فرایند اعتبارسنجی (داک-ریداک)

۶۵	۱-۳-۳-۳ فراخوانی لیگاند و پروتئین
۶۵	۲-۳-۳-۳ تنظیم فاصله بین نقاط شبکه، ابعاد شبکه و مختصات مرکز ثقل
۶۵	۳-۳-۳-۳ انتخاب الگوریتم جستجو در داکینگ مولکولی
۶۶	۴-۳-۳-۳ تنظیم پارامترهای الگوریتم ژنتیک لامارکین
۶۶	۵-۳-۳-۳ تنظیم گزینه‌های اجرای داکینگ
۶۷	۴-۳-۳-۳ داکینگ لیگاند های مورد مطالعه در پروتئین های مربوطه
۶۷	۱-۴-۳-۳ آماده سازی لیگاندهای مورد مطالعه
۶۷	۲-۴-۳-۳ اجرای عملیات داکینگ مولکولی لیگاندها و پروتئین
۶۸	۴-۳ انتخاب موثرترین توصیف کننده‌ها
۶۹	۵-۳ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه
۷۰	۶-۳ ارزیابی مدل
۷۱	۴- فصل چهارم نتایج تجربی
۷۲	۱-۴ مدل سازی فعالیت دارویی مشتقات ۴- آنیلینو کوئینازولین به عنوان ترکیبات فعال دارویی ضد سرطان
۷۳	۱-۱-۴ مجموعه داده‌ها
۷۵	۲-۱-۴ رسم و بهینه سازی ساختار ترکیبات مورد مطالعه
۷۵	۳-۱-۴ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه با استفاده از توصیف کننده‌های حاصل از داکینگ مولکولی
۷۵	۱-۳-۱-۴ استخراج توصیف کننده‌های حاصل از داکینگ مولکولی
۷۹	۲-۳-۱-۴ انتخاب موثرترین توصیف کننده‌های داکینگ مولکولی
۸۰	۳-۳-۱-۴ مدل شبکه عصبی مصنوعی
۸۱	۴-۳-۱-۴ ارزیابی مدل شبکه عصبی مصنوعی
۸۲	۴-۱-۴ مدل سازی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۸۳	۱-۴-۱-۴ توصیف کننده‌های تابع توزیع شعاعی
۸۳	۲-۴-۱-۴ انتخاب موثرترین توصیف کننده‌ها از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۸۵	۳-۴-۱-۴ مدل سازی شبکه عصبی مصنوعی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۸۶	۴-۴-۱-۴ ارزیابی مدل شبکه عصبی مصنوعی
۹۱	۵-۴-۱-۴ نتیجه‌گیری مدل سازی با مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۹۲	۶-۴-۱-۴ ارزیابی مدل با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با استفاده از پارامترهای آماری
۹۲	
۹۳	۷-۴-۱-۴ آزمون Y-تصادفی
۹۴	۲-۴ مدل سازی فعالیت دارویی مشتقات با مشابهت مولکولی دی آرل پیریمیدین‌ها (DAPY-like) به عنوان بازدارنده‌های غیرنوکلئوزیدی ویروس HIV
۹۴	۱-۲-۴ مجموعه داده‌ها
۹۷	۲-۲-۴ رسم و بهینه سازی ساختار ترکیبات مورد مطالعه
۹۷	۳-۲-۴ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه با استفاده از توصیف کننده‌های حاصل از داکینگ مولکولی
۹۷	۱-۳-۲-۴ استخراج توصیف کننده‌های حاصل از داکینگ مولکولی
۱۰۳	۲-۳-۲-۴ انتخاب موثرترین توصیف کننده‌های داکینگ مولکولی
۱۰۴	۳-۳-۲-۴ مدل شبکه عصبی مصنوعی
۱۰۵	۴-۳-۲-۴ ارزیابی مدل شبکه عصبی مصنوعی
۱۰۷	۴-۲-۴ مدل سازی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

۱۰۸.....	۱-۴-۲-۴ توصیف کننده‌های شمارش گروه‌های عاملی
۱۰۸.....	۲-۴-۲-۴ انتخاب موثرترین توصیف کننده‌ها از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۱۱۰.....	۳-۴-۲-۴ مدل سازی شبکه عصبی مصنوعی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۱۱۱.....	۴-۴-۲-۴ ارزیابی مدل شبکه عصبی مصنوعی
۱۱۷.....	۵-۴-۲-۴ نتیجه‌گیری مدل سازی با مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۱۱۸.....	۶-۴-۲-۴ ارزیابی مدل با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با استفاده از پارامترهای آماری
۱۱۸.....	۷-۴-۲-۴ آزمون Y - تصادفی
۱۱۹.....	۳-۴ مدل سازی فعالیت دارویی مشتقات آزینی به‌عنوان بازدارنده‌های غیرنوکلئوزیدی ویروس HIV
۱۲۰.....	۱-۳-۴ مجموعه داده‌ها
۱۲۳.....	۲-۳-۴ رسم و بهینه سازی ساختار ترکیبات مورد مطالعه
۱۲۳.....	۳-۳-۴ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه با استفاده از توصیف کننده‌های حاصل از داکینگ مولکولی
۱۲۳.....	۱-۳-۳-۴ استخراج توصیف کننده‌های حاصل از داکینگ مولکولی
۱۲۴.....	۲-۳-۳-۴ انتخاب موثرترین توصیف کننده‌های داکینگ مولکولی
۱۲۵.....	۳-۳-۳-۴ مدل شبکه عصبی مصنوعی
۱۲۶.....	۴-۳-۳-۴ ارزیابی مدل شبکه عصبی مصنوعی
۱۳۱.....	۴-۳-۳-۴ مدل سازی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۱۳۱.....	۱-۴-۳-۴ شاخص‌های شبه دارویی
۱۳۲.....	۲-۴-۳-۴ انتخاب موثرترین توصیف کننده‌ها از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۱۳۳.....	۳-۴-۳-۴ مدل سازی شبکه عصبی مصنوعی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۱۳۴.....	۴-۴-۳-۴ ارزیابی مدل شبکه عصبی مصنوعی
۱۴۰.....	۵-۴-۳-۴ نتیجه‌گیری مدل سازی با مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری
۱۴۱.....	۶-۴-۳-۴ ارزیابی مدل با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با استفاده از پارامترهای آماری
۱۴۱.....	۷-۴-۳-۴ آزمون Y-تصادفی
۱۴۳.....	۵- فصل پنجم نتیجه گیری و آینده نگری
۱۴۴.....	۱-۵ نتیجه‌گیری نهایی
۱۴۵.....	۱-۱-۵ بررسی ارتباط توصیف کننده‌های وارد شده در مدل با فعالیت بازدارندگی مشتقات ۴-آنیلینوکوئینازولین به عنوان بازدارنده‌های سرطان
۱۴۵.....	۲-۱-۵ بررسی ارتباط توصیف کننده‌های وارد شده در مدل با فعالیت بازدارندگی مشتقات DAPY-Like به عنوان بازدارنده‌های HIV
۱۴۷.....	۳-۱-۵ بررسی ارتباط توصیف کننده‌های وارد شده در مدل با فعالیت بازدارندگی مشتقات آزینی به عنوان بازدارنده‌های HIV
۱۵۱.....	
۱۵۴.....	۲-۵ آینده نگری
۱۵۵.....	۳-۵ منابع

## فهرست شکل‌ها:

- شکل ( ۱-۱ ) انواع گیرنده‌های سرطان و لیگاندهای آن [۱۱]..... ۸
- شکل ( ۲-۱ ) ساختار EGFR [۱۲]..... ۹
- شکل ( ۳-۱ ) فعالسازی تیروزین کیناز [۱۳]..... ۱۰
- شکل ( ۴-۱ ) ساختمان ویروس HIV [۱۶]..... ۱۳
- شکل ( ۵-۱ ) نحوه ورود ویروس HIV به دورن سلول میزبان [۱۷]..... ۱۵
- شکل ( ۶-۱ ) نحوه عملکرد بازدارنده‌های نوکلئوزیدی آنزیم نسخه بردار معکوس [۲۰]..... ۱۶
- شکل ( ۷-۱ ) نحوه عملکرد بازدارنده‌های غیر نوکلئوزیدی آنزیم نسخه بردار معکوس [۲۰]..... ۱۷
- شکل ( ۱-۲ ) اتصال لیگاند به ماکرو مولکول در طی فرایند داکینگ مولکولی [۵۷]..... ۳۱
- شکل ( ۲-۲ ) طرح شماتیک از مدل قفل و کلید [۶۰]..... ۳۲
- شکل ( ۳-۲ ) ساختار کریستالوگرافی پروتئین الف) کد کریستالوگرافی 1M17 ب) کد کریستالوگرافی 3MEC ج) کد کریستالوگرافی 3M8Q [۶۲]..... ۳۵
- شکل ( ۴-۲ ) ساختار کمپلکس لیگاند و پروتئین..... ۳۶
- شکل ( ۵-۲ ) جعبه شبکه‌ای [۶۴]..... ۳۷
- شکل ( ۱-۴ ) اسکلت اصلی ترکیبات مورد مطالعه..... ۷۳
- شکل ( ۲-۴ ) ساختار لیگاند موجود در پروتئین 1m17..... ۷۶
- شکل ( ۳-۴ ) مختصات جایگاه فعال به دست آمده از جایگاه لیگاند در پروتئین 1M17..... ۷۶
- شکل ( ۴-۴ ) تعیین مختصات شبکه در بخش اتوگرید نرم افزار..... ۷۸
- شکل ( ۵-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون..... ۸۲
- شکل ( ۶-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون..... ۸۷
- شکل ( ۷-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده بر حسب فعالیت تجربی به روش رد مرحله‌ای تک تک..... ۸۹
- شکل ( ۸-۴ ) نمودار باقی مانده بر حسب فعالیت تجربی برای کل داده‌ها..... ۸۹
- شکل ( ۹-۴ ) آزمون پراکندگی برای تقسیم بندی داده‌های ۴- آنیلینو کوئینازولین..... ۹۰
- شکل ( ۱۰-۴ ) نمودار قلمرو کاربرد برای مدل شبکه عصبی ترکیبات ۴- آنیلینو کوئینازولین..... ۹۱
- شکل ( ۱۱-۴ ) اسکلت‌های اصلی ترکیبات مورد مطالعه..... ۹۵
- شکل ( ۱۲-۴ ) ساختار لیگاندهای موجود در پروتئین الف) 3MEC ب) 3M8Q..... ۹۸
- شکل ( ۱۳-۴ ) مختصات جایگاه فعال به دست آمده از جایگاه لیگاند در پروتئین 3M8Q..... ۹۹
- شکل ( ۱۴-۴ ) مختصات جایگاه فعال به دست آمده از جایگاه لیگاند در پروتئین 3MEC..... ۹۹
- شکل ( ۱۵-۴ ) تعیین مختصات شبکه در بخش اتوگرید نرم افزار..... ۱۰۱
- شکل ( ۱۶-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی..... ۱۰۶
- شکل ( ۱۷-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون..... ۱۰۷
- شکل ( ۱۸-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی..... ۱۱۲
- شکل ( ۱۹-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون..... ۱۱۳
- شکل ( ۲۰-۴ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی به روش رد مرحله‌ای تک تک برای کل داده‌ها..... ۱۱۵
- شکل ( ۲۱-۴ ) نمودار باقی مانده بر حسب فعالیت تجربی برای کل داده‌ها..... ۱۱۵
- شکل ( ۲۲-۴ ) نمودار پراکندگی برای تقسیم بندی ترکیبات DAPY-like..... ۱۱۶
- شکل ( ۲۳-۴ ) نمودار قلمرو کاربرد برای مدل شبکه عصبی با ترکیبات DAPY-like..... ۱۱۷
- شکل ( ۲۴-۴ ) اسکلت‌های اصلی ترکیبات مورد مطالعه..... ۱۲۰

- شکل (۲۵-۴) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی ..... ۱۲۷
- شکل (۲۶-۴) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی ..... ۱۲۸
- شکل (۲۷-۴) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی به روش رد مرحله‌ای تک تک برای کل داده ها ..... ۱۳۰
- شکل (۲۸-۴) نمودار باقی‌مانده بر حسب فعالیت تجربی برای کل داده‌ها ..... ۱۳۰
- شکل (۲۹-۴) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی ..... ۱۳۵
- شکل (۳۰-۴) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی ..... ۱۳۶
- شکل (۳۱-۴) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی به روش رد مرحله‌ای تک تک برای کل داده‌ها ..... ۱۳۸
- شکل (۳۲-۴) نمودار باقی‌مانده بر حسب فعالیت تجربی برای کل داده‌ها ..... ۱۳۸
- شکل (۳۳-۴) نمودار پراکندگی برای تقسیم بندی ترکیبات آذینی ..... ۱۳۹
- شکل (۳۴-۴) نمودار قلمرو کاربرد برای مدل شبکه عصبی با ترکیبات آذینی ..... ۱۴۰
- شکل (۱-۵) نمودار مشارکت توصیف کننده‌های منتخب در مدل شبکه عصبی توسعه یافته با ترکیبات ۴-آنیلینوکوئینازولین ..... ۱۴۷
- شکل (۲-۵) نمودار مشارکت توصیف کننده‌های منتخب در مدل شبکه عصبی توسعه یافته با ترکیبات DAPY-like ..... ۱۵۱
- شکل (۳-۵) نمودار مشارکت توصیف کننده‌های منتخب در مدل شبکه عصبی توسعه یافته با ترکیبات آذینی ..... ۱۵۳

### فهرست جدول‌ها:

- جدول (۱-۱) برخی از داروهای مهارکننده آنزیم نسخه‌بردار معکوس نوکلئوزید ..... ۱۶
- جدول (۲-۱) برخی از داروهای مهارکننده آنزیم نسخه‌بردار معکوس غیر نوکلئوزید ..... ۱۷
- جدول (۳-۱) برخی از داروهای مهارکننده آنزیم پروتئاز ..... ۱۸
- جدول (۴-۱) نام و ساختار و مقادیر فعالیت بازدارندگی ترکیبات مشابه با دی آریل پیریمیدین و داروی مرجع Nevirapine ..... ۲۰
- جدول (۱-۲) پارامترهای آماری ..... ۵۵
- جدول (۱-۴) ساختار و مقادیر pIC<sub>50</sub> ترکیبات مورد مطالعه ..... ۷۴
- جدول (۲-۴) نتایج اجراهای الگوریتم ژنتیک متفاوت برای پروتئین با شناسه کریستالوگرافی 1M17 ..... ۷۸
- جدول (۳-۴) توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR ..... ۸۰
- جدول (۴-۴) ماتریس همبستگی توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR ..... ۸۰
- جدول (۵-۴) مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR ..... ۸۰
- جدول (۶-۴) توابع و پارامترهای شبکه‌های بهینه شده با توصیف کننده‌های انتخاب شده داکینگ مولکولی ..... ۸۱
- جدول (۷-۴) نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از داده‌های سری آزمون ..... ۸۲
- جدول (۸-۴) توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی به روش SR ..... ۸۴
- جدول (۹-۴) ماتریس همبستگی توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR ..... ۸۵
- جدول (۱۰-۴) مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با روش SR ..... ۸۵
- جدول (۱۱-۴) توابع و پارامترهای شبکه‌های بهینه با توصیف کننده‌های منتخب از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR ..... ۸۶

جدول ( ۴-۱۲ )	نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از داده‌های سری آزمون.....	۸۷
جدول ( ۴-۱۳ )	نتایج حاصل از ارزیابی مدل شبکه عصبی به روش رد مرحله‌ای تک تک برای کل داده‌ها.....	۸۸
جدول ( ۴-۱۴ )	مقایسه مدل های شبکه عصبی توسعه یافته با توصیف کننده‌های متفاوت در ارزیابی با استفاده از سری آزمون.....	۹۱
جدول ( ۴-۱۵ )	پارامترهای آماری برای مدل ANN با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری.....	۹۲
جدول ( ۴-۱۶ )	مقادیر $R^2$ برای سری آزمون با استفاده از آزمون Y-تصادفی.....	۹۳
جدول ( ۴-۱۷ )	ساختار و مقادیر $pEC_{50}$ ترکیبات مورد مطالعه.....	۹۶
جدول ( ۴-۱۸ )	نتایج اجراهای الگوریتم ژنتیک متفاوت برای پروتئین با شناسه کریستالوگرافی 3MQ8.....	۱۰۱
جدول ( ۴-۱۹ )	نتایج اجراهای الگوریتم ژنتیک متفاوت برای پروتئین با شناسه کریستالوگرافی 3MEC.....	۱۰۱
جدول ( ۴-۲۰ )	توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR.....	۱۰۳
جدول ( ۴-۲۱ )	ماتریس همبستگی توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR.....	۱۰۳
جدول ( ۴-۲۲ )	مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR.....	۱۰۴
جدول ( ۴-۲۳ )	توابع و پارامترهای شبکه‌های بهینه شده با توصیف کننده‌های انتخاب شده داکینگ مولکولی.....	۱۰۵
جدول ( ۴-۲۴ )	نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری ارزیابی.....	۱۰۵
جدول ( ۴-۲۵ )	نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون.....	۱۰۶
جدول ( ۴-۲۶ )	توصیف کننده‌های انتخاب شده از مجموع توصیف کننده های ساختاری و داکینگ مولکولی با روش SR.....	۱۰۹
جدول ( ۴-۲۷ )	ماتریس همبستگی توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR.....	۱۰۹
جدول ( ۴-۲۸ )	مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های داکینگ مولکولی و توصیف کننده‌های ساختاری با روش SR.....	۱۰۹
جدول ( ۴-۲۹ )	توابع و پارامترهای شبکه‌های بهینه توصیف کننده‌های منتخب از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR.....	۱۱۰
جدول ( ۴-۳۰ )	نتایج حاصل از ارزیابی مدل شبکه عصبی بهینه با استفاده از داده‌های سری ارزیابی.....	۱۱۱
جدول ( ۴-۳۱ )	نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون.....	۱۱۲
جدول ( ۴-۳۲ )	نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از روش رد مرحله‌ای تک تک برای کل داده‌ها.....	۱۱۴
جدول ( ۴-۳۳ )	مقایسه مدل های شبکه عصبی توسعه یافته با توصیف کننده‌های متفاوت در ارزیابی مدل با استفاده از سری آزمون.....	۱۱۷
جدول ( ۴-۳۴ )	پارامترهای آماری برای مدل ANN با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری.....	۱۱۸
جدول ( ۴-۳۵ )	مقادیر $R^2$ برای سری آزمون با استفاده از آزمون Y-تصادفی.....	۱۱۹
جدول ( ۴-۳۶ )	ساختار و مقادیر $pEC_{50}$ ترکیبات مورد مطالعه.....	۱۲۱
جدول ( ۴-۳۷ )	توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR.....	۱۲۵
جدول ( ۴-۳۸ )	ماتریس همبستگی توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR.....	۱۲۵
جدول ( ۴-۳۹ )	مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR.....	۱۲۵
جدول ( ۴-۴۰ )	توابع و پارامترهای شبکه‌های بهینه با توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR.....	۱۲۶
جدول ( ۴-۴۱ )	نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری ارزیابی.....	۱۲۷

جدول ( ۴-۴۲ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون.....	۱۲۸
جدول ( ۴-۴۳ ) نتایج حاصل از ارزیابی مدل با استفاده از روش رد مرحله‌ای تک تک.....	۱۲۹
جدول ( ۴-۴۴ ) توصیف کننده‌های انتخاب شده از مجموع توصیف کننده های ساختاری و داکینگ مولکولی با روش SR.....	۱۳۳
جدول ( ۴-۴۵ ) ماتریس همبستگی توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR.....	۱۳۳
جدول ( ۴-۴۶ ) مقادیر تورم واریانس توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR.....	۱۳۳
جدول ( ۴-۴۷ ) توابع و پارامترهای شبکه‌های بهینه توصیف کننده‌های منتخب از مجموع توصیف کننده های ساختاری و داکینگ مولکولی با روش SR.....	۱۳۴
جدول ( ۴-۴۸ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری ارزیابی.....	۱۳۵
جدول ( ۴-۴۹ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون.....	۱۳۶
جدول ( ۴-۵۰ ) نتایج حاصل از ارزیابی مدل با استفاده از روش رد مرحله‌ای تک تک.....	۱۳۷
جدول ( ۴-۵۱ ) مقایسه مدل های شبکه عصبی توسعه یافته با توصیف کننده‌های متفاوت در ارزیابی مدل با استفاده از سری آزمون.....	۱۴۰
جدول ( ۴-۵۲ ) پارامترهای آماری برای مدل ANN با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری.....	۱۴۱
جدول ( ۴-۵۳ ) مقادیر $R^2$ برای سری آزمون با استفاده از آزمون Y-تصادفی.....	۱۴۲
جدول ( ۵-۱ ) محدوده قابل قبول ویژگی‌های مولکولی برای بروز خاصیت آنتی نئوپلاستیک در DRAGON و محدوده این خواص برای ترکیبات مورد مطالعه.....	۱۵۲

### فهرست رابطه‌ها:

رابطه ( ۱-۱ ).....	۶
رابطه ( ۲-۱ ).....	۶
رابطه ( ۳-۱ ).....	۶
رابطه ( ۴-۱ ).....	۷
رابطه ( ۱-۲ ).....	۵۰
رابطه ( ۲-۲ ).....	۵۱
رابطه ( ۳-۲ ).....	۵۱
رابطه ( ۴-۲ ).....	۵۱
رابطه ( ۵-۲ ).....	۵۱
رابطه ( ۶-۲ ).....	۵۱
رابطه ( ۷-۲ ).....	۵۲
رابطه ( ۸-۲ ).....	۵۲
رابطه ( ۹-۲ ).....	۵۲
رابطه ( ۱۰-۲ ).....	۵۲
رابطه ( ۱۱-۲ ).....	۵۲
رابطه ( ۱۲-۲ ).....	۵۳



۵۳.....	رابطه (۱۳-۲)
۵۳.....	رابطه (۱۴-۲)
۵۳.....	رابطه (۱۵-۲)
۵۴.....	رابطه (۱۶-۲)
۵۶.....	رابطه (۱۷-۲)
۵۷.....	رابطه (۱۸-۲)
۵۸.....	رابطه (۱۹-۲)
۸۳.....	رابطه (۱-۴)
۱۳۲.....	رابطه (۲-۴)
۱۴۵.....	رابطه (۱-۵)
۱۴۸.....	رابطه (۲-۵)
۱۴۹.....	رابطه (۳-۵)
۱۴۹.....	رابطه (۴-۵)
۱۵۱.....	رابطه (۵-۵)



مقدمه

## مقدمه

به‌طور کلی هر بیماری حاصل به هم خوردن نظم طبیعی یکی از چرخه‌های فعالیت بدن است، بنابراین با دانستن این که کدام چرخه بدن از نظم طبیعی خود خارج شده است، می‌توان دارویی را به‌کار برد که دوباره آن را به حالت اولیه باز گرداند. بروز انواع بیماری‌ها از قبیل سرطان، آلزایمر، ایدز، بیماری‌های مشترک انسان و دام و مقاوم شدن ویروس‌ها در برابر آنتی بیوتیک‌ها و ... همه از جمله مواردی می‌باشند که ذهن دانشمندان را در دستیابی و طراحی داروهای نسل جدید با عوارض جانبی کمتر و خواص بهبود یافته در زمان محدود معطوف به خود نموده‌اند. عمدتاً مولکول دارو (لیگاند) با یک مولکول خاص (هدف) در سیستم زیستی که نقش مشخصی بازی می‌کند برهمکنش دارد. امروزه طراحی داروهای نوترکیب بیشتر بر اساس سنتز ترکیب و سپس انجام آزمایش‌های لازم جهت بررسی فعالیت دارویی آن ترکیب می‌باشد. این روش‌ها معمولاً بر اساس سعی و خطا می‌باشد تا ترکیبی با فعالیت دارویی چشم‌گیر و مناسب شناخته شود. با توجه به زمان‌بر و هزینه‌بر بودن سنتز آزمایشگاهی ترکیبات و بررسی‌های فارماکولوژی، به کارگیری روش‌های QSAR به جهت ارایه مدلی برای پیش‌بینی فعالیت ترکیبات هم‌مرده (حتی ترکیباتی که هنوز سنتز نشده‌اند) توانسته است علاوه بر کاهش هزینه‌ها، کمک موثری در طراحی هدفمند داروها و مهیا نمودن اطلاعاتی در مورد چگونگی تعیین خواص شیمیایی سازگار برای فعالیت دارویی ترکیبات داشته باشد.

امروزه بیماری‌هایی همچون ایدز و سرطان ذهن دانشمندان و متخصصان بی‌شماری را به خود جلب کرده‌اند. این دو بیماری در حال حاضر گریبان‌گیر تمامی جوامع بشری بوده و روز به روز به آمار مبتلایان و قربانیان آن‌ها افزوده می‌گردد و علی‌رغم تلاش‌های صورت گرفته جهت درمان و پیش‌گیری از ایدز و سرطان، این دو پدیده همچنان از چالش‌های مهم سلامتی در سراسر جهان می‌باشند. لذا سنتز ترکیبات دارویی جدید برای آن‌ها از اهمیت ویژه‌ای برخوردار است. از آنجایی که انجام آزمایش‌های تجربی وقت‌گیر و پرهزینه می‌باشند، لذا توسعه روش‌های تئوری جهت ارزیابی و پیش‌بینی فعالیت

دارویی ترکیبات جدید از اهمیت قابل توجهی برخوردار است.

در این پروژه با توجه به ضرورت مدل سازی QSAR برای پیش بینی فعالیت دارویی مهارکننده های این دو گروه از بیماری ها، مشتقات وسیعی از ترکیبات هتروسیکل با خواص دارویی ضد سرطان و ضد HIV مورد بررسی قرار گرفته اند.

## ۱-۱ پروتئین

پروتئین ها یکی از انواع درشت مولکول های زیستی هستند که واحد تشکیل دهنده آن اسید آمینه می باشد. پروتئین ها مانند یک کلاف سه بعدی از بسپارها<sup>۱</sup> می باشند که از ترکیب اسید آمینه ها ساخته می شوند. پروتئین ها در اکثر فرایندهای درون سلولی نقش دارند و معمولاً به وسیله برهمکنش با دیگر پروتئین ها نقش خود را اعمال کرده و شبکه برهمکنشی پروتئین ها را تشکیل می دهند. به همین دلیل تشخیص چگونگی برهمکنش آن ها گامی کلیدی برای شناخت عملیات درون سلول است [۱]. برخی از پروتئین ها (مانند هورمون های پروتئینی) به عنوان سیگنال بین سلول ها، وظیفه پیام رسانی را بر عهده دارند. برخی از آن ها قسمتی از سیستم ایمنی هستند (مانند آنتی بادهای)، برخی در ساختمان سلول دخالت دارند، برخی دیگر نیز به عنوان کانال هایی جهت عبور و مرور به داخل و خارج سلول عمل می کنند. بیش از ۳۰۰۰۰ نوع پروتئین در بدن انسان وجود دارد که بعضی از آن ها با پروتئین های دیگر جانداران یکسان و بعضی نیز مشابه هستند [۲].

مطالعه پروتئین ها و عملکرد آن ها در واقع موجب فهم بهتر خود سلول ها و موجودات می شود. در سطح مولکولی، کلیه مکانیسم های زیستی سلول ها، توسط پروتئین ها انجام می شود. پروتئین ها در ارتباط با یکدیگر به طور دقیق و بسیار کنترل شده وظایف خود را انجام می دهند. اساس مولکولی اغلب بیماری ها، بروز نقص یا تداخل در کارکرد عادی پروتئین هایی است که از طریق آن مکانیسم های زیستی سلولی را انجام می دهند. اگر فعالیت پروتئین ها از حالت طبیعی خارج شود می تواند باعث بیماری های

---

<sup>۱</sup> Polymer

مختلف شوند [۳].

## ۲-۱ دارو

در بیوشیمی، گیرنده<sup>۱</sup> به یک مولکول پروتئینی گفته می‌شود که بر سطح و یا در عرض غشای پلاسمایی سلول یا درون سیتوپلاسم جاسازی شده است تا مولکولی به آن بچسبد، این مولکول که به گیرنده می‌چسبد لیگاند نام دارد و ممکن است یک پپتید، هورمون، یک دارو یا یک سم باشد و وقتی که اتصال انجام شد، گیرنده تغییر شکل داده و باعث شروع یک زنجیره از پاسخ‌ها توسط سلول می‌شود. گیرنده‌ها به صورت اختصاصی عمل کرده و هر لیگاند، گیرنده مخصوص به خود را دارد. هر گیرنده پیام‌های خارج سلولی مخصوص به خود را دریافت و اثر این پیام‌ها یا خود پیام را به درون سلول منتقل می‌کند [۴].

اتصال دارو به گیرنده از طریق پیوندهای شیمیایی صورت می‌گیرد که شامل پیوندهای کوالانسی، یونی، هیدروفوبی و همچنین هیدروژنی می‌باشد. نیروهایی که دارو را به گیرنده‌اش متصل می‌کنند باید به اندازه‌ای قوی و پایدار باشند که بتواند منجر به پاسخ فیزیولوژیک گردد. پس از مصرف دارو و عبور مولکول دارو از غشاها و قرار گرفتن در نزدیک سطح گیرنده، جاذبه الکترواستاتیک اولین نیرویی است که داروها را به طرف سطح گیرنده با بار مخالف می‌کشد. این نوع پیوند از فاصله‌ی دورتر می‌تواند ایجاد جاذبه کند که پیوند هیدروژنی از این فاصله قادر به ایجاد جاذبه نمی‌باشد. عموماً پیوند یونی (الکترواستاتیک) باید قبل از آن که گیرنده فعال شود از طریق ایجاد پیوند هیدروژنی یا واندروالس یا هر دو تقویت شود در غیر این صورت به آسانی و با سرعت توسط انرژی گرمایی شکسته می‌شود. حتی ممکن است با وجود تشکیل پیوندهای ثانویه متعدد، کمپلکس دارو-گیرنده باز هم تفکیک شود مگر این که پیوند کوالانسی ایجاد شده باشد.

---

<sup>۱</sup> Receptor

در بحث تعامل دارو با گیرنده، داروی‌های آگونیست<sup>۱</sup> پس از اتصال به گیرنده با ایجاد تغییراتی در سطح گیرنده آن را فعال می‌کنند و سبب ایجاد پاسخ می‌شوند، در بعضی گیرنده‌ها مولکول اثر کننده درون خود گیرنده واقع شده و با اتصال دارو مستقیماً فعال می‌شود. مانند باز شدن یک کانال یونی و در برخی دیگر مولکول اثر کننده یک مولکول واسطه مجزا می‌باشد که گیرنده پس از اتصال دارو با آن ارتباط برقرار می‌کند. داروهای آنتاگونیست<sup>۲</sup> وقتی روی گیرنده می‌نشینند باعث بلوک یا مهار آن می‌شوند و در واقع از اتصال مولکول‌های دیگر به گیرنده جلوگیری می‌کند [۴].

### ۳-۱ مکانیسم اتصال دارو-گیرنده

نقاط هدف دارو، اغلب پروتئین‌هایی همچون آنزیم و یا گیرنده‌هایی در سطح سلول می‌باشند. برای انتخاب داروی مناسب دانستن مکانیسم مهار کنندگی آن الزامی است. مکانیسم مهار کنندگی شامل موارد زیر می‌باشد:

۱. مهار کنندگی رقابتی

۲. مهار کنندگی غیر رقابتی

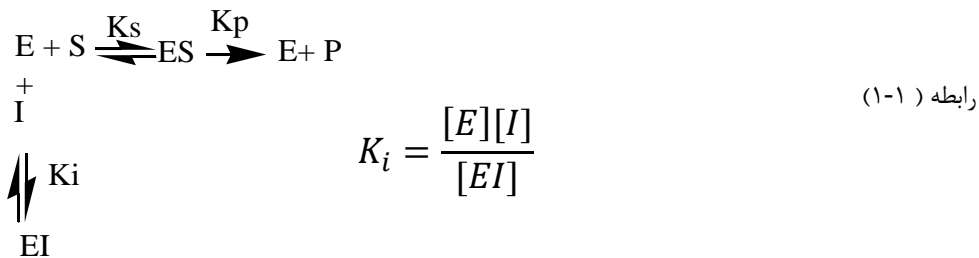
اغلب داروها از راه مکانیسم رقابتی عمل می‌کنند. به این معنا که به‌طور برگشت‌پذیری به جایگاه فعال مولکول هدف متصل می‌شوند. هنگامی که دارو در جایگاه فعال است، سوبسترای اصلی نمی‌تواند متصل شود و این مسئله کارایی پروتئین را کاهش می‌دهد، بدون این که آن را کاملاً از بدن حذف نماید. مولکول‌هایی وجود دارند که ساختار شیمیایی آن‌ها شبیه سوبسترای طبیعی یک آنزیم است. این امر ممکن است آن‌ها را قادر سازد تا با جایگاه فعال موجود روی آنزیم پیوند برقرار کنند و به همین دلیل از طریق کاهش تعداد جایگاه‌های فعال در دسترس سوبسترای اصلی، در کار آن‌ها اختلال ایجاد کنند. این مواد بر سر اشغال جایگاه فعال با سوبسترا رقابت می‌کنند؛ بنابراین به‌عنوان مهار کننده‌های رقابتی شناخته می‌شوند. واکنش آنزیمی با مهار کنندگی رقابتی از طریق ایجاد کمپلکس را می‌توان به صورت

---

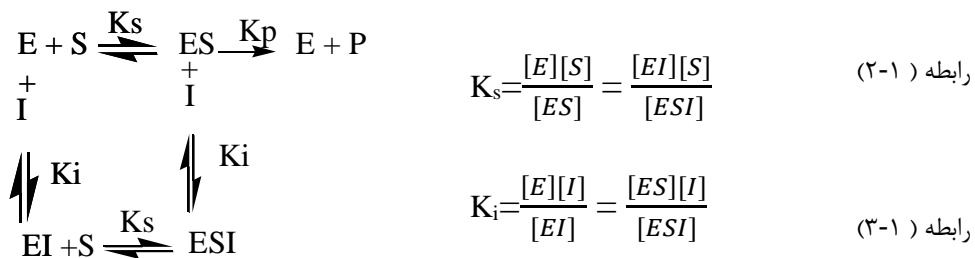
<sup>۱</sup> Agonist

<sup>۲</sup> Antagonist

رابطه ( ۱-۱ ) نمایش داد:



E, S, I, ES, EI و P به ترتیب مربوط به آنزیم، سوبسترا، بازدارنده، کمپلکس آنزیم-سوبسترا، کمپلکس آنزیم-بازدارنده و محصول می‌باشد.  $K_s$  و  $K_p$  به ترتیب مربوط به ثابت‌های تفکیک آنزیم-سوبسترا و آنزیم-محصول است و  $K_i$  ثابت بازداری نام دارد. مهارکننده‌های غیر رقابتی به جایگاه‌هایی به جز جایگاه‌های آنزیم متصل می‌شوند که در این صورت جایگاه فعال آنزیم برای سوبسترا آزاد می‌باشد، اما این مهارکننده‌ها شکل پروتئین را تغییر می‌دهند و بنابراین فعالیت کاتالیتیک جایگاه فعال را کاهش می‌دهند [۵].



غلظت مهارکنندگی در برهم‌کنش لیگاند-پروتئین حائز اهمیت بوده و آگاهی از آن در بررسی فعالیت آنزیم ضروری است.  $IC_{50}$  یا غلظت مهارکنندگی عبارت است از کمترین غلظتی از دارو که دارای ۵۰٪ اثر بازداری باشد. به عبارت دیگر به غلظتی از دارو که از فعالیت دارویی و یا بیوشیمیایی آنزیم تا حدود ۵۰٪ جلوگیری کند،  $IC_{50}$  گفته می‌شود. مقدار  $IC_{50}$  با استفاده از رابطه ( ۴-۱ ) به ثابت بازداری  $K_i$  تبدیل می‌شود.



$$K_i = \frac{IC_{50}}{1 + \frac{S}{K_m}}$$

رابطه ( ۴-۱ )

S غلظت بازدارندگی (دارو) بوده و  $K_m$  نیز ثابت متناسب با هر آنزیم می‌باشد و غلظت بازدارندگی (دارو) در نصف ماکسیمم فعالیت آنزیم می‌باشد [۶].

## ۴-۱ بیماری سرطان و اهمیت پیش‌بینی داروهای ضد سرطان

سرطان یک بیماری کشنده ژنتیکی است که در این بیماری ناهنجاری در ژن‌های کنترل‌کننده تکثیر سلولی موجب رشد نامحدود سلول‌ها می‌شود. طبق بررسی‌های اپیدمیولوژیک انجام شده در ایالت متحده آمریکا، از هر ۴ مورد مرگ و میر، یک نفر با ابتلاء به سرطان می‌میرد. در حال حاضر سرطان دومین علت اصلی مرگ و میر در کشور های توسعه یافته به لحاظ اقتصادی و سومین علت مرگ و میر در کشورهای در حال توسعه می‌باشد [۷، ۸]. تکثیر سلول‌ها در حالت نرمال کاملاً کنترل شده و بر اساس نیازهای بدن انجام می‌شود و در هر زمان به‌طور کاملاً دقیق، به‌وسیله فاکتورهای رشد، پیام‌های محیطی و برخی پروتئین‌ها کنترل می‌شوند. جهش‌هایی که منجر به تغییر هر یک از فاکتورهای موثر در سرنوشت سلول می‌شوند، باعث بهم خوردن نظم دقیقی که در تنظیم رشد، تکثیر و تمایز سلول‌ها وجود دارد، شده و می‌تواند منجر به بروز سرطان شود. سلول‌های سرطانی کنترل خود را بر چرخه سلولی از دست داده و به‌طور مدارم و بدون توجه به پیام‌های سلولی و فاکتورهای رشد، به تکثیر ادامه می‌دهند. در سلول‌های سرطانی بیان پروتئین‌های سلولی تغییر کرده و برخی از آن‌ها که نباید بیان شوند، بیان شده و بیان برخی دیگر از ژن‌ها متوقف می‌شوند. تغییرات گوناگون در ژنوم، ماشین رونویسی<sup>۱</sup>، ماشین پروتئین‌سازی<sup>۲</sup> سلول‌های طبیعی منجر به تبدیل ژن‌ها به انکوژن<sup>۳</sup> و افزایش تولید محصولات ژنی و یا غیر فعال شدن سرکوب‌کننده‌های تومور می‌شود [۹]. گیرنده

<sup>۱</sup> Transcriptome

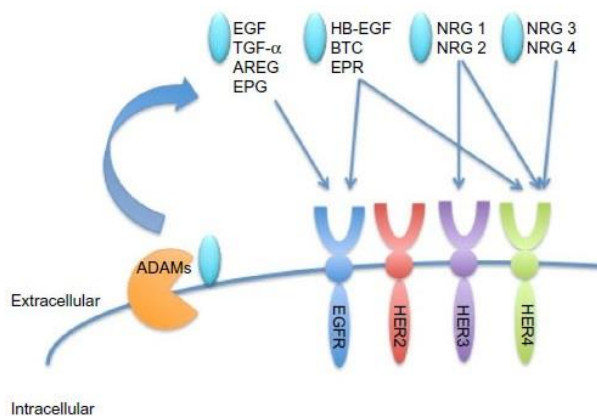
<sup>۲</sup> Proteome

<sup>۳</sup> Oncogene

فاکتور رشد اپیدرمی یکی از اعضای مهم خانواده گیرنده تیروزین کیناز<sup>۱</sup> (RTK) محسوب می‌شود.

### ۱-۴-۱ گیرنده فاکتور رشد اپیدرمی

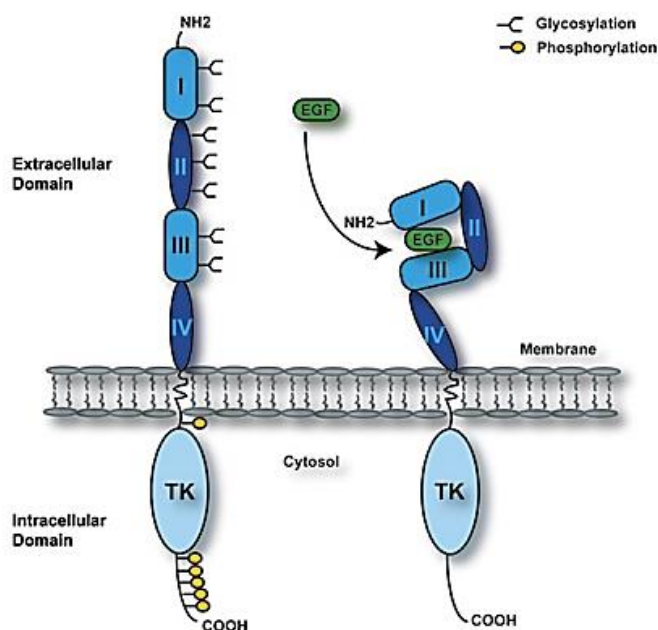
این گیرنده نقش مهمی در فرآیندهای مهم و پایه‌ای سلول از جمله تکثیر، بقای سلولی و مهاجرت سلولی دارا می‌باشد. این خانواده شامل ۴ عضو مهم می‌باشد که عبارتند از HER1 (EGFR)، HER2، HER3 و HER4. همانگونه که در شکل (۱-۱) نیز نشان داده شده است، بعضی از لیگاندها اختصاصی یک نوع گیرنده بوده ولی بعضی از آنها نیز به‌طور مشترک بین چند گیرنده عمل می‌نمایند [۱۰].



شکل ( ۱-۱) انواع گیرنده‌های سرطان و لیگاندهای آن [۱۱]

گیرنده فاکتور رشد اپیدرمی که مهم‌ترین عضو این خانواده می‌باشد پروتئینی با وزن ۱۷۰ کیلو دالتون بوده و از سه قسمت که در شکل ( ۲-۱) مشاهده می‌شود، تشکیل شده است که عبارتند از قسمت خارج سلولی، بین غشایی (هیدروفوبیک) و داخل سلولی (تیروزین کیناز) [۱۲].

<sup>۱</sup> Receptor tyrosine kinase

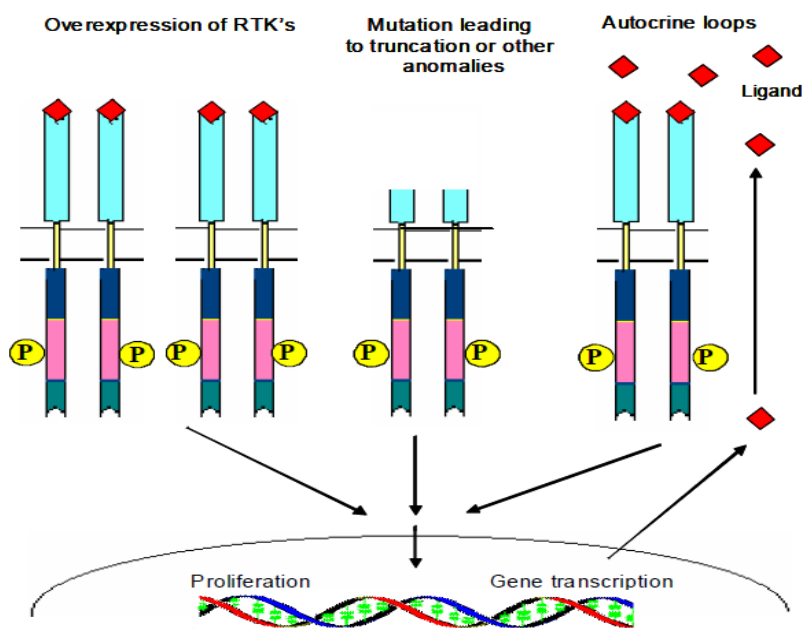


شکل (۲-۱) ساختار EGFR [۱۲]

#### ۲-۴-۱ مکانیسم عمل گیرنده فاکتور رشد اپیدرمی

فعال‌سازی رسپتور EGFR پس از اتصال لیگاند از طریق دیمریز شدن و فسفریلاسیون جایگاه‌های ویژه‌ای در قسمت داخل سلولی انجام می‌گیرد. شکل (۳-۱) فعال شدن تیروزین کیناز در فسفریلاسیون بافت سرطانی در اثر فرایندهایی در طی انتقال پیام<sup>۱</sup> (فرایندهایی که به‌وسیله‌ی آن اطلاعات بیولوژیکی منتقل می‌شود) که یک مرحله‌ی مهم در ایجاد تومورهای سرطانی می‌باشد، اتفاق می‌افتد. بنابراین اثر کلیدی تیروزین کیناز در ایجاد تومور سرطانی بسیار قابل تأمل است و اخیراً تیروزین کینازها به‌عنوان هدف برای داروهای ضد سرطان به‌کار می‌روند [۱۳].

<sup>۱</sup> - Signal transduction



شکل ( ۳-۱ ) فعالسازی تیروزین کیناز [۱۳]

### ۳-۴-۱ گیرنده‌ی فاکتور رشد اپیدرمی و رابطه‌ی آن با سرطان

وقتی گیرنده‌ی فاکتور رشد اپیدرمی فعال شود، بسیاری از فرایندهای سلولی از قبیل تکثیر سلولی، تمایز، مهاجرت، چسبندگی، محافظت در برابر آپوپتوسیس، تحریک رگزایی، آغاز می‌گردد. افزایش بیان گیرنده‌ی فاکتور رشد اپیدرمی و لیگاند آن باعث ایجاد بسیاری از سرطان‌ها از جمله سرطان ریه با یاخته‌های غیرکوچک، سرطان مثانه، سرطان سینه و سرطان مولتی‌فرم گلیوبلاستوما<sup>۱</sup> می‌شود. بیان فراوان گیرنده‌ی فاکتور رشد اپیدرمی در سرطان ریه با یاخته‌های غیرکوچک حدود ۸۰-۴۰ درصد و در سرطان ریه با یاخته‌های کوچک حدود ۵۰ درصد افزایش می‌یابد. فعالیت تومورزایی تیروزین کیناز در سلول‌های سرطانی می‌تواند به‌وسیله انتخاب مهارکننده‌های تیروزین کیناز متوقف شود. روش‌های مختلفی برای مهار تیروزین کیناز وجود دارد، که یکی از آن روش‌ها بازداری به وسیله بازدارنده‌های فسفریلاسیون تیروزین با وزن مولکولی پایین انجام می‌شود. در سال ۱۹۸۰ بازدارنده‌های

<sup>۱</sup> - Multiform Glioblastoma

با وزن مولکولی پایین پیشنهاد شد که می‌توانستند مانع تکثیر سلولی وابسته به لیگاند EGF شوند. بیشتر بازدارنده‌های تیروزین کیناز تقلیدی از ATP هستند و حداقل دو حلقه آروماتیک دارند. اختلاف کوچکی در ATP منجر به تغییراتی در پیوند هیدروژنی و برهمکنش‌های هیدروفوبیکی شده و سبب تغییراتی در تمایل آن می‌شود. بازدارنده در جایگاه پیوند ATP واقع در شکاف تیروزین کیناز اتصالات پایدار با باقی‌مانده‌ی تیروزین کیناز برقرار کرده و از فسفوریلاسیون در لوپ فعال جلوگیری می‌کند [۱۰، ۱۳]. در این تحقیق رابطه‌ی کمی ساختار-فعالیت بازدارنده‌های تیروزین کیناز گیرنده‌ی فاکتور رشد اپیدرمی مورد بررسی قرار گرفت.

#### ۴-۴-۱ مشتقات کوئینازولین

کوئینازولین‌ها یک دسته مهم از ترکیبات هتروسیکلی هستند که از دو حلقه‌ی جوش خورده بنزنی و پیریدین تشکیل شده‌اند. کوئینازولین‌ها دارای طیف گسترده‌ای از خواص دارویی و فعالیت‌های بیولوژیکی مانند ضد سرطان، ضد ویروس بودن، ضد سل، ضد دیابت، چاقی، ضد التهاب، ضد میکروب و ضد تشنج هستند. این ترکیبات با مکانیسم رقابتی به‌عنوان بازدارنده‌های ضدسرطان مورد توجه قرار گرفته‌اند [۱۴].

#### ۵-۱ بیماری ایدز و اهمیت پیش‌بینی داروهای ضد ایدز

##### ۵-۱-۱ ایدز

سندرم نقص ایمنی اکتسابی<sup>۱</sup> (AIDS) یک بیماری عفونی کشنده است که اولین بار در سال ۱۹۸۳ شناسایی شد. ویروس ایدز سلول‌های ایمنی موجود در خون (لنفوسیت‌ها) و سلول‌های ایمنی موجود در بافت‌ها را درگیر می‌سازد. این سلول‌ها در تولید پادتن برای مقابله با بیماری‌ها و سرطان‌ها نقش دارند. وقتی دستگاه ایمنی آسیب ببیند، نه تنها در برابر ویروس HIV بلکه نسبت به بقیه عفونت‌ها و ویروس‌هایی که قبلاً ایجاد مشکل نمی‌کردند نیز آسیب‌پذیر می‌شود. لذا با گذشت زمان افراد آلوده به

<sup>۱</sup> Acquired Immunodeficiency Syndrom

HIV بیشتر و بیشتر بیمار می‌شوند و معمولاً سال‌ها پس از آلودگی با این ویروس، به یکی از بیماری‌های خاص (بیماری‌های متأثر از نقص دستگاه ایمنی) مبتلا شده و در این زمان گفته می‌شود که آن‌ها به ایدز مبتلا شده‌اند. بنابراین زمانی که فرد آلوده به ویروس HIV برای اولین بار به یک بیماری جدی مبتلا شود و یا وقتی که تعداد سلول‌های ایمنی باقی‌مانده در بدن او از حد معینی کم‌تر شود (زیر ۲۰۰ عدد در هر میکرولیتر)، مبتلا به بیماری ایدز در نظر گرفته می‌شود [۱۵].

ویروس HIV از خانواده رتروویروس‌ها<sup>۱</sup> می‌باشد. رتروویروس‌ها، گروهی از ویروس‌ها هستند که ژنوم آن‌ها از RNA تشکیل شده است. بنابراین برای تکثیر خود به آنزیمی به نام آنزیم نسخه بردار معکوس<sup>۲</sup> وابسته‌اند که ژنوم RNA آن‌ها را به DNA نسخه‌برداری کند تا بعد بتواند آن را به کمک آنزیم اینتگرز<sup>۳</sup> وارد ژنوم میزبان کند و به این ترتیب امکان تکثیر ویروس به وجود آید. کلمه رترو (معکوس) در نام این ویروس‌ها به‌همین خاطر است چرا که معمولاً نسخه‌برداری از DNA به RNA انجام می‌شود اما در این دسته از ویروس‌ها عکس این عمل صورت می‌گیرد. رتروویروس‌ها باعث کاهش توانایی سیستم ایمنی بدن میزبان می‌شوند. علائمی که در بیماری ایدز مشاهده می‌شود مربوط به بیماری‌هایی است که در اثر نقص دستگاه ایمنی بدن تولید می‌شوند [۱۵].

## ۲-۵-۱ ساختمان ویروس HIV

HIV یک ذره کاملاً کروی با قطر ۱۲۰-۱۰۰ نانومتر است که با غشای دو لایه لیپیدی پوشیده شده است. در سطح غشای لیپیدی دو نوع گلیکوپروتئین (GP41, GP120) وجود دارد که ویروس از طریق این گلیکوپروتئین‌ها قادر خواهد بود به گیرنده‌های CD4<sup>+</sup> موجود در سطح لنفوسیت T کمک کننده<sup>۴</sup> بچسبد. سطح داخلی غشای لیپیدی توسط ماتریکس احاطه شده و اطلاعات ژنتیکی مورد نیاز برای

---

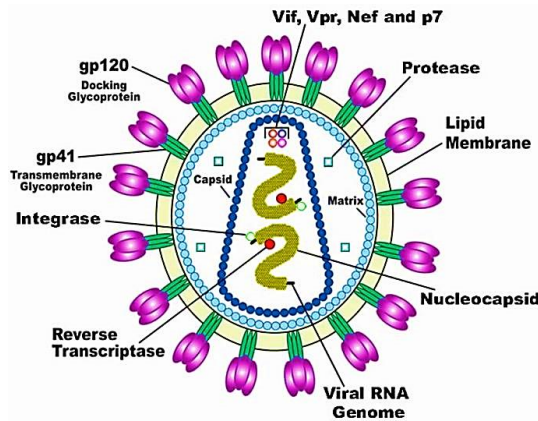
<sup>۱</sup> Retrovirus

<sup>۲</sup> Reverse transcriptase

<sup>۳</sup> Integrase

<sup>۴</sup> T helper lymphocyte

تکثیر ویروس را در خود جای داده است. این اطلاعات شامل دو کپی یکسان از ژنوم RNA ویروسی و سه آنزیم اینتگراز، پروتئاز<sup>۱</sup> و نسخه بردار معکوس می باشد [۴]. شکل (۴-۱) ساختمان ویروس HIV را نشان می دهد.



شکل (۴-۱) ساختمان ویروس HIV [۱۶]

### ۳-۵-۱ چرخه تکثیر ویروس HIV

چرخه تولید و تکثیر ویروس شامل مراحل زیر است:

۱- چسبیدن ویروس به سلول های خونی و آلوده کردن سلول ها

۲- تولید DNA پروویروسی و اتصال به ژنوم میزبان

۳- رونویسی<sup>۲</sup> ژن های ویروسی

۴- تولید قطعات ویروسی و رهاسازی ویروس

سلول هایی که مورد تهاجم این ویروس قرار می گیرند، نوعی از سلول های خونی انسان به نام لنفوسیت های نوع T کمک کننده می باشند. این سلول ها دارای گیرنده ای به نام  $CD_4^+$  هستند که ویروس پس از اتصال به این گیرنده، توسط پروتئین GP41 (موجود در سطح ویروس HIV) به داخل غشای سلول نفوذ می کند و RNA ویروس و آنزیم های مختلف (از جمله نسخه بردار معکوس، اینتگراز و پروتئاز) به داخل سیتوپلاسم سلول تزریق می شود. مرحله بعد تولید DNA از روی RNA ویروس با کمک آنزیم

<sup>۱</sup> Protease

<sup>۲</sup> Transcription

نسخه‌بردار معکوس است. در صورت موفقیت این عمل، DNA ساخته شده که DNA پروویروسی نام دارد، با استفاده از آنزیم اینتگراز به DNA سلول میزبان ملحق می‌گردد. بعد از رونویسی در هسته سلول، RNA ناقل<sup>۱</sup> به سیتوپلاسم سلول میزبان منتقل می‌شود و از اجزای سلولی میزبان برای ساختن پروتئین‌های ویروسی استفاده می‌کند. آنزیم پروتئاز فراهم شده به وسیله HIV وظیفه شکستن پلی‌پپتید ساخته شده به رشته‌های پروتئینی سازنده HIV را بر عهده دارد. به این ترتیب ویروس HIV می‌تواند یک ذره کامل بسازد که به غشای میزبان مهاجرت کرده و غشای لیبیدی مورد نیازش را از غشای سلول میزبان می‌گیرد و آماده‌ی آلوده کردن سلول‌های دیگر است [۱۵].

#### ۱-۵-۴ اهداف استفاده از داروهای ضد HIV

۱. سرکوب ویروس و کاهش روند تکثیر آن‌ها
۲. تحکیم و بهبود عملکرد سیستم ایمنی که با شمارش سلول‌های  $CD4^+$  ارزیابی می‌شود.
۳. بهبود کیفیت زندگی بیماران با کاهش عوارض بیماری
۴. بهبود امید به زندگی و افزایش طول عمر بیماران
۵. کاهش احتمال سرایت

#### ۱-۵-۵ درمان ضد رترو ویروسی HIV

درمان ضد رتروویروسی نوع اصلی درمان ایدز می‌باشد که البته شفا بخش نیست اما زندگی مبتلایان به این بیماری را طولانی‌تر می‌کند. درمان شامل داروهایی است که بایستی روزانه و تا آخر عمر مصرف شوند [۱۵].

#### ۱-۵-۶ انواع داروهای ضد رتروویروسی

از زمان آغاز اپیدمی HIV مجموعه‌ای از داروها ساخته شده‌اند که از طریق مهار توانایی تکثیر این ویروس، عمر افراد مبتلا به این ویروس را به‌طور قابل ملاحظه‌ای افزایش می‌دهند. این داروها از سرعت

---

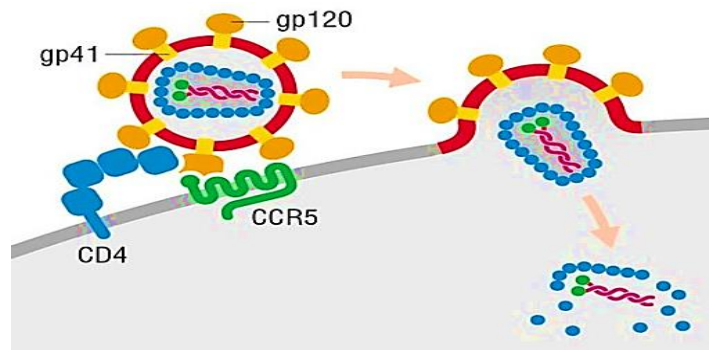
<sup>۱</sup> Messenger RNA



نابودی  $CD4^+$  بیماران می‌کاهد و می‌توانند بروز بیماری ایدز را به تأخیر بیندازد، اما آن را معالجه نمی‌کند. در مجموع چهار دسته دارو وجود دارد که در مراحل مختلف آلودگی و رشد ویروس HIV عمل می‌کنند و با توجه به اینکه این ویروس از سه آنزیم نسخه‌بردار معکوس، پروتئاز و اینتگراز برای تکثیر خود و آلوده کردن سلول‌ها استفاده می‌کند، بنابراین داروهای سنتز شده بیش‌تر این نوع آنزیم‌ها را هدف قرار می‌دهند.

### ۱-۵-۱ بازدارنده‌های ورودی<sup>۱</sup>

با توجه به شکل ( ۱-۵ ) در سطح بیرونی ویروس HIV گلیکوپروتئین‌های GP41, GP120 وجود دارد که به‌وسیله آن‌ها ویروس به گیرنده‌های  $CD4^+$  موجود در سطح لنفوسیت‌های T سلول میزبان متصل شده و وارد سلول می‌گردد. بازدارنده‌های ورودی به این پروتئین‌ها می‌چسبند و از پیوستن ویروس به سلول‌های  $CD4^+$  جلوگیری می‌کند. به این ترتیب با بلوکه کردن یکی از این پروتئین‌ها، تکثیر آهسته می‌شود.



شکل ( ۱-۵ ) نحوه ورود ویروس HIV به دورن سلول میزبان [۱۷]

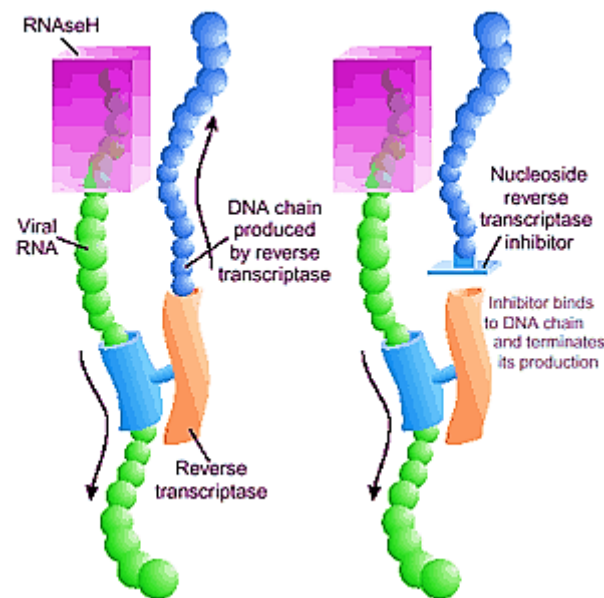
### ۱-۵-۲ بازدارنده‌های نوکلئوزیدی آنزیم نسخه‌بردار معکوس<sup>۲</sup>

این دسته دارویی از سال ۱۹۸۷ در درمان HIV موجود است و بیشتر به‌عنوان مشابه نوکلئوزید (Nucleoside analogues) شناخته می‌شود. این نوع از مهارکننده‌ها بازدارنده‌های رقابتی هستند.

<sup>۱</sup> Fusion or Entry inhibitors

<sup>۲</sup> Nucleoside Reverse Transcriptase Inhibitors (NRTIs)

مهارکننده‌های نوکلئوزیدی به زنجیره نوکلئوزیدی DNA که توسط آنزیم نسخه‌بردار معکوس در حال تولید است، می‌چسبند و مانع از اتصال نوکلئوزیدهای بیشتر به DNA می‌شوند به این ترتیب یک DNA ناقص تولید می‌شود. شکل ( ۶-۱ ) عملکرد این مهارکننده‌ها را به خوبی نشان می‌دهد [۱۸, ۱۹].



شکل ( ۶-۱ ) نحوه عملکرد بازدارنده‌های نوکلئوزیدی آنزیم نسخه بردار معکوس [۲۰]

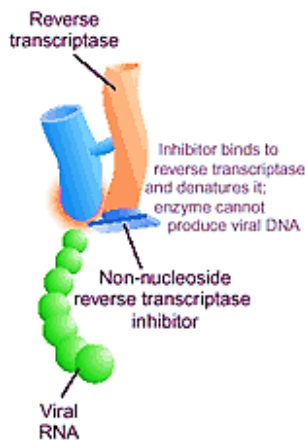
از این دسته به داروهای موجود در جدول ( ۱-۱ ) می‌توان اشاره کرد [۲۱].

جدول ( ۱-۱ ) برخی از داروهای مهارکننده آنزیم نسخه‌بردار معکوس نوکلئوزید

نام دارو در یک رده شیمیایی خاص و نام تجاری	نام شیمیایی (نام تحقیقاتی)
Epivir, lamivudine	3TC
Ziagen, ABC	Abacavir
Retrovir, zidovudine	AZT
(AZT/3TC combined)	Combivir
(AZT/3TC/abacavir combined)	Trizivir†
Zerit, Stavudine	d4T
Hivid, zalcitabine	ddC
Videx (tablet), Videx FC (Capsule), Zalcitabine	ddl
Emtriva, emtricitabine	FTC

## ۱-۵-۳ بازدارنده‌های غیرنوکلئوزیدی نسخه‌بردار معکوس<sup>۱</sup>

این دسته دارویی از سال ۱۹۹۷ در درمان HIV مورد تایید قرار گرفته‌اند. این دسته از مهارکننده‌ها، بازدارنده‌های غیر رقابتی هستند و به بخش هیدروفوبی جایگاه فعال متصل می‌شوند. نویراپین<sup>۲</sup> نخستین گروه از این مهارکننده‌ها بود که در سال ۱۹۹۶ توسط وزارت بهداشت کانادا مورد تایید قرار گرفت. این نوع مهارکننده‌ها نیز فرایند رونویسی را مختل می‌کنند. با توجه به شکل (۷-۱) آن‌ها با چسباندن خود به آنزیم نسخه‌بردار معکوس و محدود کردن آنزیم، مانع از نسخه‌برداری می‌شوند. از این دسته داروها می‌توان به موارد موجود در جدول (۲-۱) اشاره کرد [۲۱].



شکل (۷-۱) نحوه عملکرد بازدارنده‌های غیر نوکلئوزیدی آنزیم نسخه‌بردار معکوس [۲۰]

جدول (۲-۱) برخی از داروهای مهارکننده آنزیم نسخه‌بردار معکوس غیر نوکلئوزید

نام دارو در یک رده شیمیایی خاص و نام تجاری	نام شیمیایی (نام تحقیقاتی)
Sustiva	Efavirenz
Viramune	Nevirapine

## ۱-۵-۴ بازدارنده‌های پروتئاز<sup>۳</sup>

این داروها اولین بار در سال ۱۹۹۵ تایید شدند. پروتئاز یک آنزیم گوارشی است که پروتئین را تجزیه می‌کند و تقریباً در هر سلول زنده‌ای وجود دارد. آنزیم پروتئاز در HIV به زنجیره طویل آنزیم‌ها

<sup>۱</sup> Non Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

<sup>۲</sup> Nevirapin

<sup>۳</sup> Protease inhibitor

و پروتئین‌ها، در سلول حمله کرده و آن‌ها را به قطعات کوچک‌تری تقسیم می‌کند تا از این طریق پروتئین‌های لازم برای تولید یک ویروس بالغ تولید شود. ولی مهارکننده پروتئاز به مکان‌هایی از آنزیم که بریدن<sup>۱</sup> زنجیره پلی‌پپتیدی از آن مکان رخ می‌دهد، می‌چسبد و مانع از تولید پروتئین‌های لازم برای تولید ویروس جدید می‌شود. به عبارتی، مهارکننده پروتئاز قادر است فرایند تبدیل ویروس نابالغ غیرعفونی را به ویروس بالغ عفونی، آهسته کند [۱۸]. در این روش آنزیم مهارکننده پروتئاز، تکثیر ویروس و در نتیجه آلوده کردن سلول‌های جدید را آهسته می‌کند. NRTIs و NNRTIs فقط بر روی سلول‌هایی که جدیداً آلوده شده‌اند، موثر هستند. آنزیم مهارکننده پروتئاز در سلول‌هایی که مدت طولانی است آلوده شده‌اند، اثر خود را از طریق آهسته کردن تکثیر ویروس اعمال می‌کند. دسته ای از این داروها در جدول ( ۳-۱ ) خلاصه شده است [۲۱].

جدول ( ۳-۱ ) برخی از داروهای مهارکننده آنزیم پروتئاز

نام دارو در یک رده شیمیایی خاص و نام تجاری	نام شیمیایی (نام تحقیقاتی)
Agenerase	Amprenavir
Reyataz	Atazanavir
Crixivan	Indinavir
Kaletra	Lopinavir/Ritonavir
Kaletra	Lopinavir/Ritonavir
Viracept	Nelfinavir
Norvir	Ritonavir
Fortovase (soft gell), Invirase (hard gell)	Saquinavir

## ۱-۶ مشتقات آزینی

آزین‌ها، مولکول‌هایی هستند که در آن‌ها ۱ تا ۶ اتم نیتروژن با کربن موجود در حلقه بنزن جایگزین می‌شود. نیتروژن‌ها می‌توانند به صورت پیوسته (۲،۱-۳،۲،۱-۴،۱-۳،۱-۴،۱-۳،۱-۴،۱) در حلقه توزیع شوند. پیریدین ترکیبی با یک نیتروژن، پیریدازین با دو پیوند نیتروژن پیوسته، پیریمیدین و پیرازین با دو پیوند نیتروژن گسسته و تری‌آزین‌ها با سه نیتروژن به‌عنوان ترکیبات پایدار و آروماتیک‌های شناخته شده هستند. سیستم‌هایی با چهار تا شش نیتروژن موجود در حلقه نیز مورد

<sup>۱</sup> Cut

توجه قرار گرفته و نوع تترا آزین کاملاً شناخته شده است اما پنتا آزین و هگزا آزین آن چنان شناخته شده نیستند [۲۲]

همان طور که گفته شد مهارکننده‌های غیر نوکلئوزیدی با هدف قرار دادن آنزیم نسخه‌بردار معکوس، در درمان عفونت HIV مؤثر هستند. مشتقات آزین‌ها (پیریدین‌ها، پیریدازین‌ها، پیریمیدین‌ها، پیرازین‌ها و تری آزین‌ها و ...) به‌عنوان دسته جدیدی از مهارکننده‌های غیر نوکلئوزیدی شناخته شده‌اند.

### ۷-۱ مشتقات دی آریل پیریمیدین‌ها و مشابه‌های<sup>۱</sup> آن‌ها

مشتقات مشابه با دی آریل پیریمیدین<sup>۲</sup> (DAPY-like) به‌عنوان گروه مهمی از بازدارنده‌های NNRTIs در مقابله با HIV-1 به تایید سازمان غذا و دارو ایالت متحده آمریکا<sup>۳</sup> (US FDA) رسیده است [۲۳]. این ترکیبات با اصلاح ساختاری مشتقات دی آریل پیریمیدین به وسیله‌ی جایگزینی حلقه پیریمیدین با ساختار دو حلقه‌ای شامل نیتروژن به‌دست آمده و مقاومت مناسبی را در برابر HIV-1 از خود نشان داده‌اند و همچنین ویژگی‌های فارموکینتیک این مشتقات نسبت به مشتقات DAPY بهبود یافته و دارای فعالیت بازدارندگی قابل توجهی می‌باشند. دسته‌ای از ترکیبات DAPY-like در جدول (۱-۴) نمایش داده شده است که فعالیت بازدارندگی این ترکیبات از داروی مرجع (Nevirapine) نیز بیشتر بوده است [۲۴-۲۶].

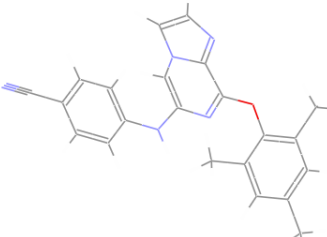
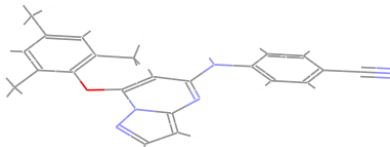
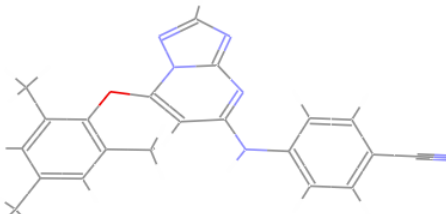
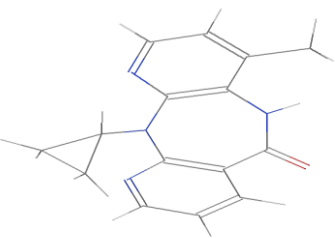
---

<sup>۱</sup> Analogues

<sup>۲</sup> Diaryl pyrimidine like

<sup>۳</sup> US Food and Drug Administration

جدول ( ۴-۱ ) نام و ساختار و مقادیر فعالیت بازدارندگی ترکیبات مشابه با دی آریل پیریمیدین و داروی مرجع Nevirapine

Formula	structure	EC <sub>50</sub>
4-((8- mesityloxy)imidazo [1,2-a]pyrazin-6-yl)amino)benzotrile		0.26 μM[24]
4-(7-Mesityloxy)pyrazolo[1,5-a]pyrimidin-5-ylamino)benzotrile		0.07 μM[25]
4-((7-Mesityloxy)-[1,2,4]triazolo[1,5-a]pyrimidin-5-yl)amino)benzotrile		0.05 μM[26]
Nevirapine		0.31 μM[24]

## ۸-۱ مروری بر کارهای پیشین

در سال های اخیر ساختارهای کریستالوگرافی گیرنده ها (پروتئین ها) بیشتر در دسترس محققین قرار گرفته است و همچنین استفاده از داکینگ مولکولی، به عنوان یک روش مطالعه ساختار کمپلکس لیگاند-گیرنده گسترش پیدا کرده است. از این رو استفاده از توصیف کننده های داکینگ مولکولی که حاوی اطلاعاتی از ساختار هدف و نحوه برهمکنش لیگاند با جایگاه فعال گیرنده می باشند در مدل سازی ارتباط کمی ساختار- فعالیت<sup>۱</sup> (QSAR) بسیار مورد توجه قرار گرفته اند. اغلب توصیف کننده های

<sup>۱</sup> Quantitative structure- activity relationship

داکینگ مولکولی به دو روش قابل محاسبه می‌باشند. در روش اول به منظور استخراج توصیف کننده‌های ساختاری، بهترین پیکربندی لیگاند در پایدارترین حالت کمپلکس لیگاند-گیرنده به نرم افزارهایی از جمله [29] CODESSA<sup>۲</sup>, [28] BINANA<sup>۱</sup>, [27] DRAGON منتقل می‌شود و در نهایت از این توصیف کننده‌ها برای ساخت مدل‌های QSAR استفاده می‌شود [۳۸-۳۰]. در روش دوم از اطلاعات خروجی داکینگ به‌عنوان توصیف کننده در ساخت مدل QSAR استفاده می‌شود، این توصیف کننده‌ها اغلب از نوع انرژی و یا برهمکنش اسیدآمینها با گیرنده می‌باشد. مدل‌های QSAR توسعه داده شده با توصیف کننده‌های داکینگ مولکولی در طی سال‌های ۲۰۰۹-۲۰۱۹ بررسی شده‌اند که در ادامه به اختصار شرح داده می‌شوند.

گرگ<sup>۳</sup> و همکارانش در سال ۲۰۱۰ مدل QSAR را برای ۲۳ ترکیب آنتی باکتریایی توسعه دادند. ۱۱ توصیف کننده با استفاده از داکینگ مولکولی محاسبه شد. توصیف کننده‌های با ضریب همبستگی<sup>۴</sup> بالای ۰/۵ حذف شد و مدل رگرسیون خطی چندگانه<sup>۵</sup> (MLR) و ماشین بردار پشتیبان<sup>۶</sup> (SVM) با ۳ توصیف کننده باقی‌مانده توسعه داده شد. ضریب تعیین<sup>۷</sup> سری آموزش برای این دو مدل به ترتیب برابر با ۰/۶۷ و ۰/۸۶ به دست آمد [۳۹].

سینگلا<sup>۸</sup> و همکارانش در سال ۲۰۱۱ مدل QSAR را برای پیش‌بینی غلظت بازدارندگی ۸۴ ترکیب ضد باکتریایی توسعه دادند. ۳۴۳۰ توصیف کننده با استفاده از نرم افزارهای V-life، Web-DRAGON، Cdk و داکینگ مولکولی استخراج شدند. ضریب تعیین مدل به روش ارزیابی تقاطعی<sup>۹</sup> (CV) CV-SVM-QSAR با استفاده از ۱۱ توصیف کننده‌ی موثر انتخاب شده از مجموع توصیف

---

<sup>۱</sup> BINDing ANALyzer

<sup>۲</sup> Comprehensive Descriptors for Structural and Statistical Analysis

<sup>۳</sup> Garg

<sup>۴</sup> Correlation coefficient

<sup>۵</sup> Multiple linear regression

<sup>۶</sup> Support vector machine

<sup>۷</sup> The coefficient of determination

<sup>۸</sup> Singla

<sup>۹</sup> Cross validation

کننده‌های داکینگ مولکولی، V-life و Web-Cdk، ۰/۶۳ به‌دست آمد. [۴۰]

ابراهیمی و همکارانش در سال ۲۰۱۲، از ۱۳۴ توصیف کننده داکینگ مولکولی و ساختاری برای مدل سازی ۶۱ بازدارنده کینازی با مشابهت اکتیوین (ALK5)<sup>۱</sup> استفاده کردند. ۸ توصیف کننده موثر با استفاده از روش رگرسیون خطی چندگانه مرحله‌ای<sup>۲</sup> (SR) انتخاب شدند. مدل SVM به روش ارزیابی تقاطعی (CV) و با داده‌های سری آموزش توسعه داده شد و ضریب تعیین سری آزمون ۰/۸۴ به‌دست آمد [۴۱].

بایونوچی<sup>۳</sup> و همکارانش در سال ۲۰۱۳، ۸۴۲ توصیف کننده داکینگ مولکولی و ساختاری را برای ۵۹ داروی قلبی-عروقی با استفاده از نرم افزارهای PaDEL، AutoDock4.2 و Intersurf tool محاسبه کردند. موثرترین توصیف کننده‌ها با استفاده از روش MLR انتخاب و مدل شدند. ارزیابی مدل با استفاده از سری آزمون و روش رد مرحله‌ای تک تک انجام شد و ضریب تعیین به ترتیب برابر با ۰/۶۹ و ۰/۶۸ به‌دست آمد [۴۲].

قرقانی و همکارانش در سال ۲۰۱۳ مدل SVM را برای ۶۸ بازدارنده آلزایمر توسعه دادند. ۱۲۹ توصیف کننده با استفاده از نرم افزارهای Autodock4.2، BINANA و Hyperchem محاسبه شد. انتخاب موثرترین توصیف کننده‌ها با استفاده از SR انجام شد. مدل رگرسیون بردار پشتیبان با استفاده از ۴ توصیف کننده داکینگ مولکولی (BINANA، Autodock4.2) و یک توصیف کننده ساختاری (Hyperchem) توسعه داده شد و ضریب تعیین سری آزمون برابر با ۰/۸۶ به‌دست آمد [۴۳].

چارکراپورتی<sup>۴</sup> و همکارانش در سال ۲۰۱۴ از ۵۰ بازدارنده‌ی بتا سکریتاز<sup>۵</sup> برای پیش‌بینی فعالیت دارویی استفاده کردند. سه توصیف کننده‌ی داکینگ مولکولی شامل انرژی داکینگ، اندروالسی و

---

<sup>۱</sup> Activin-like kinase-5

<sup>۲</sup> Stepwise MLR

<sup>۳</sup> Bianucci

<sup>۴</sup> Chakraborty

<sup>۵</sup> B-secretase



هیدروژنی با استفاده از نرم افزار iGEMDOCK و ۱۲ توصیف کننده‌ی ساختاری با استفاده از نرم افزار HyperChem محاسبه شدند. مدل MLR-QSAR با سه توصیف کننده‌ی داکینگ مولکولی ساخته شد و ضریب تعیین مدل برابر با ۰/۵ به دست آمد. سپس مدل MLR-QSAR با ۸ توصیف کننده‌ی داکینگ مولکولی و ساختاری ساخته شد و ضریب تعیین مدل به ۰/۹۵ بهبود یافت [۳۳].

ابراهیمی و همکارش در سال ۲۰۱۴ مدل SVM را برای ۲۹ مشتق بربرین<sup>۱</sup> به عنوان داروهای ضد سرطانی توسعه دادند. ۱۰۴ توصیف کننده با به کارگیری نرم افزارهای BINANA, Autodock4.2, و Hyperchem محاسبه شد. ۵ توصیف کننده مهم با استفاده از stepwise MLR انتخاب شد و مدل SVM با ۴ توصیف کننده داکینگ مولکولی (BINANA, Autodock4.2) و یک توصیف کننده ساختاری (Hyperchem) توسعه داده شد. ضریب تعیین سری آزمون برابر با ۰/۸۲ به دست آمد و سایر پارامترهای آماری مدل نیز دارای مقادیر قابل قبول بودند [۴۴].

ژنگ<sup>۲</sup> و همکارانش در سال ۲۰۱۴ از ۹۳ مشتق بوتیریل کولین استراز<sup>۳</sup> به عنوان بازدارنده‌های آلزایمر برای ساخت مدل QSAR استفاده کردند. ۶ توصیف کننده با استفاده از داکینگ مولکولی محاسبه شد. مدل MLR با دو توصیف کننده انرژی آزاد اتصال و انرژی چرخشی توسعه داده شد. ضریب تعیین مدل برابر با ۰/۶۷ به دست آمد [۴۵].

امینی و همکارانش در سال ۲۰۱۶، با استفاده از مطالعه QSAR فعالیت ضد دیابت ۳۳ ترکیب دی پپتیدیل پپتیداز-۴ (DPP-IV) را پیش‌بینی کردند. در این مطالعه به منظور استخراج توصیف کننده‌ها از نرم افزارهای BINANA, AutoDock 4.2 و DRAGON استفاده شد. مدل MLR و ANN با استفاده از توصیف کننده‌های منتخب روش رگرسیون گام به گام توسعه یافت. ضریب تعیین و خطای استاندارد (SE)<sup>۴</sup> سری آزمون به ترتیب برابر با ۰/۸۵ و ۰/۲۱۳ برای مدل MLR و ۰/۹۵۴ و ۰/۱۶۸

---

<sup>۱</sup> Berberine

<sup>۲</sup> Zheng

<sup>۳</sup> Butyrylcholinesterase

<sup>۴</sup> Sum of Squares Error

برای مدل ANN به دست آمد. نتایج به دست آمده توانمندی مدل ANN برای پیش بینی فعالیت ۳۳ بازدارنده DPP-IV را نشان داد [۳۵].

صفری زاده و همکارش در سال ۲۰۱۹، درصد بازدارندگی آلزایمر ۳۶ مشتق کوئینولین را پیش‌بینی کردند. توصیف‌کننده‌های داکینگ مولکولی (۱۲ توصیف‌کننده)، Hyperchm (۸ توصیف‌کننده)، DRAGON (۱۵۶۰ توصیف‌کننده) و Gaussian (۲۲ توصیف‌کننده) محاسبه شدند. مدل MLR با ۳ توصیف‌کننده ساختاری و یک توصیف‌کننده داکینگ مولکولی توسعه داده شد. مجموع مربعات خطا (SSE) و ضریب تعیین سری آزمون به ترتیب برابر با ۵/۶۷ و ۰/۸۹ به دست آمد [۴۶].

با توجه به مطالعات QSAR اخیر، توصیف‌کننده‌های داکینگ مولکولی به‌عنوان توصیف‌کننده‌های جدید مورد توجه قرار گرفته‌اند. اما این توصیف‌کننده‌ها اغلب به تنهایی منجر به ساخت مدل‌های پیشگو نمی‌شوند [۳۳، ۴۰، ۴۵، ۴۶]. لذا به‌کارگیری توصیف‌کننده‌های ساختاری به‌همراه توصیف‌کننده‌های داکینگ مولکولی و استفاده از یک روش مدل‌سازی کارآمد می‌تواند سبب بهبود قدرت پیش‌بینی مدل شود. مطالعات انجام شده، دسته‌ وسیعی از توصیف‌کننده‌ها شامل توصیف‌کننده‌های ساختاری، الکترونی و کوانتومی و ... را به همراه توصیف‌کننده‌های داکینگ مولکولی پیشنهاد می‌دهد [۳۳، ۳۵، ۴۷-۴۰] و انتخاب یک طبقه بندی خاص از توصیف‌کننده‌های ساختاری که بتواند سبب بهبود نتایج مدل شود، همچنان ناشناخته است. در این پژوهش توصیف‌کننده‌های داکینگ مولکولی به صورت مخلوط‌های دوتایی به همراه توصیف‌کننده‌های تابع توزیع شعاعی<sup>۱</sup>، توصیف‌کننده‌های خواص مولکولی و توصیف‌کننده‌های گروه‌های عاملی<sup>۲</sup> برای ساخت مدل به کار گرفته شد.

---

<sup>۱</sup> Radial distribution function

<sup>۲</sup> Functional group counts

## فصل دوم

مبانی ساختار - فعالیت

## ۱-۲ مقدمه

در طول دهه‌های اخیر کمومتریکس به یک حوزه‌ی علمی بسیار پرکاربرد و مهم در علوم مختلف تبدیل شده است. واژه‌ی کمومتریکس در سال ۱۹۷۰ ابداع شد و به تدریج با به کارگیری روش‌های مختلف چند متغیره به منظور آنالیز داده‌های شیمیایی، به یک شاخه علمی توانمند تبدیل شد [۴۸]. در حقیقت، کمومتریکس یکی از شاخه‌های علم شیمی می‌باشد که از ریاضیات و آمار برای رسیدن به اهداف ذیل استفاده می‌کند.

➤ طراحی و یا انتخاب فرآیند آزمایشگاهی بهینه

➤ فراهم آوردن حداکثر اطلاعات شیمیایی از طریق تجزیه‌ی داده‌های آزمایشگاهی

➤ به دست آوردن دانش کافی از سیستم شیمیایی مورد مطالعه

اکنون متخصصین علم کمومتریکس از روش‌های تجزیه‌ی داده جهت استخراج اطلاعات از داده‌های خام استفاده می‌کنند. یک شیمیدان می‌تواند از این اطلاعات و همچنین از هوش شیمی خود برای به دست آوردن دانش کافی از سیستم مورد مطالعه استفاده کند [۴۹].

کمومتریکس یکی از شاخه‌های با اهمیت علمی است که کاربردهای زیادی در بسیاری از شاخه‌های علوم مانند آمار چند متغیره، مدل‌سازی ریاضی، علوم کامپیوتر و شیمی تجزیه داراست. این موضوع باعث شده است تا در بسیاری از منابع علمی از کمومتریکس به عنوان فرزند ریاضیات، کامپیوتر و شیمی نام برده شود [۴۸]. بنابراین گسترش سریع فناوری به خصوص در قسمت‌های مربوط به کاربرد کامپیوتر در شیمی تجزیه موجب پیشرفت‌های سریعی در زمینه کمومتریکس در چند سال اخیر شده است.

طراحی دارو از جمله کاربردهای کمومتریکس است. امروزه پیشرفت در علوم کامپیوتر و روش‌های محاسباتی منجر به جایگزینی روش‌های منطقی و محاسباتی به جای روش‌های سنتی و تصادفی در طراحی دارو شده است، این موضوع باعث صرفه‌جویی در زمان و هزینه‌های آزمایشگاهی و پژوهشی

می‌شود. هدف از این فصل بیان تعاریف و مفاهیم مقدماتی در فرآیند طراحی محاسباتی دارو است. اصطلاح کمومتریکس برای اولین بار توسط اسوانت ولد<sup>۱</sup> دانشمند جوان سوئدی که در زمینه شیمی فیزیک آلی فعالیت داشت، مطرح گردید. همکاری ولد با بروس آر. کووالسکی<sup>۲</sup> که در دانشگاه واشنگتن بر روی الگوشناسی در شیمی تجزیه مطالعه می‌کرد، منجر به تأسیس انجمن بین‌المللی کمومتریکس (ICS)<sup>۳</sup> در سال ۱۹۷۴ گردید [۵۰]. بنا به تعریف ICS، کمومتریکس عبارت است از کاربرد روش‌های ریاضی، آماری برای برقراری ارتباط بین سنجش‌های انجام شده روی یک سیستم یا فرایند شیمیایی به منظور درک بهتر اطلاعات شیمیایی است.

برخی از کاربردهای کمومتریکس عبارتند از:

۱. بهینه‌سازی

۲. بررسی رفتار کروماتوگرافی غیر خطی

۳. پردازش سیگنال

۴. طراحی دارو

۵. کالیبراسیون غیر خطی و مدل‌سازی [۴۸].

در میان کاربردهای مختلف کمومتریکس یکی از مهم‌ترین و شاخص‌ترین این کاربردها طراحی دارو می‌باشد که به بررسی ارتباط کمی ساختار-فعالیت (QSAR) می‌پردازد، که خواص مولکول‌ها را به ویژگی‌های ساختاری آن‌ها نسبت می‌دهد [۵۱]. در ادامه به بحث در مورد QSAR، انواع روش‌های QSAR، کاربرد آن در طراحی دارو و مراحل مخت‌لف مطالعات QSAR پرداخته خواهد شد.

---

<sup>۱</sup> Svante wold

<sup>۲</sup> Bruce R. Kowaski

<sup>۳</sup> International Chemometrics Society

## ۲-۲ ارتباط کمی ساختار-فعالیت (QSAR)

### ۱-۲-۲ تعریف

از نظر شیمیدانان فعالیت‌ها و یا خواص یک ترکیب ناشی از ویژگی‌های ساختاری آن است. هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده‌ی مولکول انجام گیرد، به آن ارتباط کمی ساختار-خاصیت<sup>۱</sup> (QSPR) می‌گویند. اما وقتی خواصی از نوع بیولوژیکی (مانند فعالیت دارو) توصیف می‌شود، اشاره به ارتباط کمی ساختار-فعالیت (QSAR) خواهد داشت. QSAR سعی در پیدا کردن رابطه‌ی هماهنگ میان فعالیت‌های شیمیایی و فیزیکی با ویژگی‌های مولکولی دارد، 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 [۵۲]. مدل‌های QSAR را می‌توان بر اساس پارامترهایی که به صورت تجربی به دست آمده و یا از طریق تئوری محاسبه شده‌اند، به زبان ریاضی بیان نمود [۵۳]. روش‌های QSAR را می‌توان بر اساس معیارهای گوناگونی طبقه بندی کرد که رایج‌ترین این طبقه‌بندی‌ها در بخش بعدی مطرح شده است.

### ۲-۲-۲ طبقه بندی بر اساس نوع روش‌های کمومتریکس

بعضی مواقع روش‌های QSAR بر اساس نوع تکنیک به کار رفته برای ایجاد رابطه‌ی بین ویژگی‌های ساختاری و فعالیت دارویی (نوع مدل‌سازی) تقسیم می‌شوند.

✓ روش‌های خطی مانند تحلیل مؤلفه‌های اصلی<sup>۲</sup> (PCA)، رگرسیون کمترین توان‌های دوم

جزئی<sup>۳</sup> (PLS)، رگرسیون خطی چندگانه (MLR) و جنگل‌های تصادفی<sup>۴</sup> (RF) و...

<sup>۱</sup> Quantitative structure – property relationship

<sup>۲</sup> Principal component analysis

<sup>۳</sup> Partial least squares regression

<sup>۴</sup> Random forest

✓ روش‌های غیر خطی مانند شبکه عصبی مصنوعی<sup>۱</sup> (ANN)، الگوریتم نزدیک ترین همسایه ها<sup>۲</sup> (kNN) [۵۴] و ...

## ۲-۲-۳ طبقه بندی بر اساس نوع اطلاعات مورد استفاده در مدل سازی

روش‌های QSAR جهت مطالعه برهمکنش دارو (لیگاند)-پروتئین را نیز می‌توان به دو دسته روش‌های QSAR مبتنی بر لیگاند و روش‌های QSAR مبتنی بر ساختار پروتئین تقسیم کرد که در ادامه به شرح مختصر آن‌ها می‌پردازیم .

### ۲-۲-۳-۱ QSAR مبتنی بر لیگاند<sup>۳</sup>

اکثر مطالعات QSAR انجام گرفته با استفاده از روش مبتنی بر لیگاند بوده است. این روش به‌طور معمول از اطلاعات مجموعه‌ای از ترکیبات با فعالیت مشخص استفاده می‌کند. عمده‌ترین کاربرد این روش برای محاسبه ویژگی‌هایی است که تابعی از تداخل‌های غیر اختصاصی بین مولکول و اطراف آن می‌باشد که تغییرات اندک در ساختار مولکولی باعث تغییرات کوچکی در ویژگی می‌گردد. اغلب، این روش در مواردی که اطلاعات ساختاری پروتئین هدف وجود نداشته باشد استفاده می‌شود. مفهوم کلیدی در روش‌های مبتنی بر لیگاند این است که ترکیباتی که ساختار مشابه و یا اجزای ساختاری مشابه نسبت به ترکیبات فعال شناخته شده دارند، به احتمال زیاد فعالیتی همانند ترکیبات مشابه خود دارند [۵۵].

### ۲-۲-۳-۲ QSAR مبتنی بر ساختار پروتئین

آگاهی از ساختار هدف بیولوژیکی، به ویژه جایگاه اتصال پروتئین، از دیدگاه طراحی دارو بسیار حائز اهمیت است، زیرا اطلاعات دقیق برهم‌کنش اتصال لیگاند-پروتئین می‌تواند در طراحی داروها استفاده شود. استفاده از این روش با عنوان طراحی داروی منطقی<sup>۴</sup> یا طراحی داروی مبتنی بر ساختار

<sup>۱</sup> Artificial neural network

<sup>۲</sup> k-Nearest-neighbors

<sup>۳</sup> Structure-based drug design

<sup>۴</sup> Rational drug design

پروتئین شناخته می‌شود. این روش به میزان اطلاعاتی که راجع به دارو و پروتئین در دسترس می‌باشد وابسته است. در حال حاضر، بلورنگاری اشعه‌ی ایکس<sup>۱</sup> و طیف سنجی NMR<sup>۲</sup> عمده‌ترین روش‌های تجربی تعیین ساختار ماکرومولکول‌های زیستی در سطح اتمی به شمار می‌رود.

ممکن است هیچ نوع اطلاعات تجربی راجع به دارو و پروتئین وجود نداشته باشد که در چنین مواردی باید از روش‌های محاسباتی استفاده کرد. با داشتن این اطلاعات مسئله پیش رو عبارت است از طراحی دارویی که به‌طور مناسبی با پروتئین که جایگاه هدف می‌باشد، برهمکنش برقرار کند که به این نوع طراحی داکینگ مولکولی<sup>۳</sup> می‌گویند. داکینگ دارو و پروتئین می‌تواند از طریق برهمکنش‌های آب‌گریز، الکترواستاتیک، پیوندهای هیدروژنی و واندروالس باشد. امروزه استفاده از ابزارهای محاسباتی چون مطالعات داکینگ مولکولی می‌تواند اطلاعات ارزشمندی در جهت طراحی دارو مفید و موثر در اختیار بگذارند. مطالعات داکینگ مولکولی اطلاعاتی نظیر نحوه برهمکنش دارو و پروتئین، جهت‌گیری دارو در سایت اتصال پروتئین و انرژی آزاد اتصال دارو به پروتئین را فراهم می‌کنند [۵۵]. انرژی اتصال لیگاند و پروتئین، ثابت بازداری تئوری، انرژی الکترواستاتیک و برخی از اطلاعات خروجی داکینگ مولکولی از مجموعه اطلاعات ارزشمندی است که در این پروژه به‌عنوان توصیف‌کننده مولکولی مورد بررسی قرار گرفته است.

## ۳-۲ برهمکنش پروتئین و لیگاند

برهمکنش پروتئین‌ها با لیگاندهای مختلف نقشی اساسی در فرایندهای زیستی دارد. مهم‌ترین کاربرد مطالعه برهمکنش پروتئین-لیگاندها، طراحی داروهای جدید بر مبنای ساختار است. روش‌های آزمایشگاهی مطالعه‌ی این برهمکنش‌ها، روش‌های تجربی همچون کالریمتری، طیف‌سنجی و ساختاری است. بهبود قدرت کامپیوترها این امکان را فراهم کرده است که برهمکنش پروتئین-لیگاند با استفاده

---

<sup>۱</sup> X-ray crystallography

<sup>۲</sup> Nuclear magnetic resonance

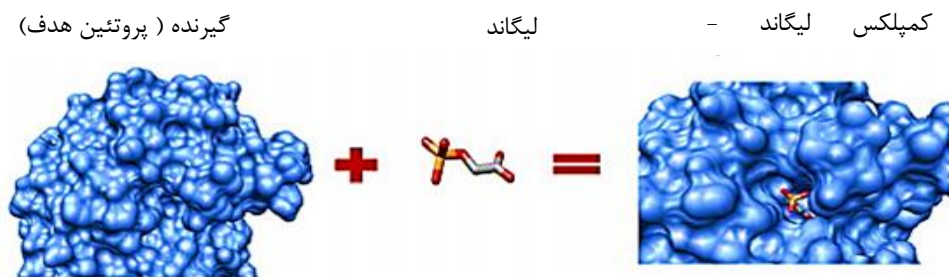
<sup>۳</sup> Molecular docking



از کامپیوتر نیز مورد مطالعه قرار گیرد. در نتیجه با استفاده از روش‌های کامپیوتری همچون داکینگ مولکولی، می‌توان پیش از انجام آزمایش تجربی، برهمکنش ترکیبات شیمیایی (داروها) را با پروتئین بررسی نمود و پس از انتخاب ترکیباتی که با احتمال بیشتری به هدف متصل می‌شوند، آن‌ها را در آزمایشگاه یا در سیستم‌های زنده مورد بررسی قرار داد.

## ۴-۲ شبیه سازی داکینگ مولکولی

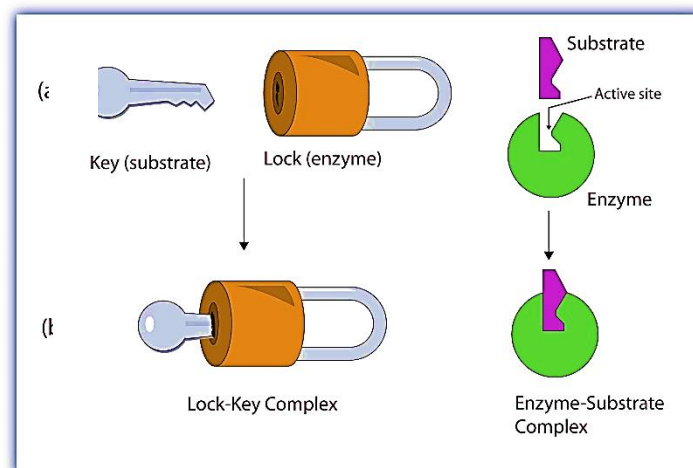
داکینگ مولکولی روشی رایج جهت طراحی و کشف داروست که از اوایل ۱۹۸۰ به‌طور گسترده مورد استفاده قرار گرفته است. مطالعات داکینگ یکی از روش‌هایی است که می‌تواند صورت‌بندی فعال زیستی و یا توصیف کننده‌های عددی را جهت مدل‌سازی‌های ارتباط کمی ساختار-فعالیت ایجاد نماید. همچنین این روش برای تعیین برهمکنش‌های کلیدی هر یک از بازدارنده‌ها با گیرنده زیستی مورد استفاده قرار می‌گیرد. در شکل (۱-۲) ایده روش داکینگ مولکولی به‌صورت اجمالی نمایش داده شده است. در این فرایند لیگاند در حفرات مولکول گیرنده (پروتئین هدف) قرار می‌گیرد. از این روش جهت توصیف رفتار مولکول‌های کوچک در محل اتصال پروتئین‌های هدف و نیز برای توضیح فرایندهای بیوشیمیایی استفاده می‌شود. یکی از اهداف مطالعات داکینگ مولکولی دستیابی به ساختار پیچیده‌ی گیرنده-لیگاند با استفاده از روش‌های محاسباتی می‌باشد [۵۶].



شکل (۱-۲) اتصال لیگاند به ماکرو مولکول در طی فرایند داکینگ مولکولی [۵۷]

فرایند داکینگ شامل دو مرحله اصلی است. مرحله اول شامل نمونه‌برداری و ایجاد صورت‌بندی‌های مختلف لیگاند و بررسی جهت‌گیری آن‌ها نسبت به جایگاه فعال گیرنده است. وقتی

انعطاف‌پذیری گیرنده هم مدنظر باشد، این مرحله در برگیرنده تغییرات صورت‌بندی جایگاه گیرنده نیز می‌شود. در مرحله دوم، تمایل اتصال لیگاند به گیرنده توسط امتیازدهی تخمین زده می‌شود. امتیاز اختصاص یافته به هر حالت بر اساس ارزیابی تابع امتیازدهی که اغلب انرژی آزاد اتصال لیگاند-گیرنده را نشان می‌دهد به دست می‌آید. در حالت ایده‌آل، بهترین امتیاز (منفی‌ترین انرژی آزاد اتصال گیرنده-لیگاند) به صورت‌بندی از لیگاند متصل به گیرنده اختصاص می‌یابد که بیشترین برهمکنش‌های جاذبه و کمترین نیروهای دافعه در آن‌ها وجود دارد [۵۸]. برخی از الگوریتم‌های امتیازدهی بر ویژگی‌های آنتالپی سیستم متمرکز شده‌اند، درحالی‌که برخی دیگر اثر آنتروپی را نیز در اتصال مولکول‌ها به یکدیگر در نظر می‌گیرند. در بیشتر موارد دارو از طریق مهار رقابتی یعنی اتصال به جایگاه فعال و جلوگیری از ورود سوبسترا به آن جایگاه عمل می‌کند، از این موضوع به‌عنوان تئوری قفل و کلید در عملکرد دارو نام برده می‌شود. وجود موقعیت فضایی خاص بر هم‌کنش دارو با گیرنده در سال ۱۸۹۴ توسط فیشر کشف شد [۵۹]. او الگوی مشهور قفل و کلید را بیان کرد و دارو را همانند یک کلید و گیرنده را همانند قفل معرفی نمود. که دارو به‌عنوان کلیدی می‌تواند فعالیت گیرنده را افزایش یا کاهش دهد [۶۰]



شکل ( ۲-۲ ) طرح شماتیک از مدل قفل و کلید [۶۰]

داکینگ مولکولی از پرکاربردترین ابزارهای طراحی محاسباتی دارو محسوب می‌شود. همچنین

بهترین روش برای ارزیابی توان یک مهارکننده نسبت به یک پروتئین ویژه می‌باشد. ساده‌ترین برنامه‌های داکینگ، انعطاف‌پذیری<sup>۱</sup> لیگاند و سایت هدف را در نظر نمی‌گیرند و برای هر یک، منحصراً از یک صورت‌بندی قفل شده<sup>۲</sup> استفاده می‌کنند. ولی برخی دیگر از برنامه‌ها، لیگاند را به صورت منعطف و سایت هدف پروتئین را صلب<sup>۳</sup> در نظر می‌گیرند. همچنین برنامه‌هایی وجود دارند که لیگاند و پروتئین را منعطف در نظر می‌گیرند اما به دلیل زمان زیادی که برای محاسبه صرف می‌کنند مرسوم نیستند [۶۱]. از آن جایی که محاسبات داکینگ برهمکنش بین ترکیب و جایگاه فعال پروتئین را شبیه سازی<sup>۴</sup> می‌کنند، نتایج آن با نتایج به دست آمده از بررسی‌های بیوشیمیایی همخوانی خوبی دارد.

## ۵-۲ مراحل انجام داکینگ مولکولی

داکینگ مولکولی به دو صورت انجام می‌گیرد:

۱. داک لیگاند کریستالوگرافی درون پروتئین (کوگنیت لیگاند)<sup>۵</sup> با پروتئین جهت اعتبار

سنجی فرایند داکینگ مولکولی

۲. داک لیگاندهای مشابه با لیگاند درون پروتئین در شرایط تعیین شده توسط فرایند اعتبار

سنجی به جهت مدل سازی

در هر دو مورد مذکور انجام مراحل زیر ضروری است.

### ۱-۵-۲ فرایند اعتبارسنجی

قبل از هر عملیات داکینگ گیرنده-لیگاند، اعتبارسنجی برای فرایند داکینگ مولکولی ضروری

است. به منظور انجام فرایند اعتبارسنجی ابتدا کوگنیت لیگاند یک‌بار از گیرنده خارج و دوباره به آن

وارد می‌شود و سپس این دو با هم داک می‌شوند که به این عملیات داک-ریداک<sup>۶</sup> یا خود-داکینگ<sup>۷</sup> هم

---

<sup>۱</sup> Flexibility

<sup>۲</sup> Locked

<sup>۳</sup> Rigid

<sup>۴</sup> Simulate

<sup>۵</sup> Cognate Ligand

<sup>۶</sup> Dock-redock

<sup>۷</sup> Selfdocking

می‌گویند و به این ترتیب بهترین حالت اتصال محاسبه شده توسط نرم افزار با آنچه که از ساختار کریستالوگرافی به دست آمده است مقایسه می‌شود. به طوری که باید میانگین ریشه توان‌های دوم انحراف<sup>۱</sup> (RMSD) حاصل از فرایند اعتبارسنجی کمتر از مقدار  $2A^\circ$  به دست آید تا فرایند داکینگ دارای اعتبار باشد، به طوری که در طی این فرایند ساختار کریستالوگرافی مجدداً با لیگاند خودش داک می‌شود و RMSD آن محاسبه می‌گردد. RMSD نشان‌دهنده‌ی تناظر نقطه به نقطه دو ترکیب با هم است و میزان اختلاف دو ساختار را در حالت سه بعدی نشان می‌دهد. RMSD اتصال لیگاند و گیرنده باید از  $2A^\circ$  کمتر باشد تا بتوان داکینگ را با شرایط تعریف شده ادامه داد و اگر نه باید شرایط را تا جایی که به زیر  $2A^\circ$  بیاید تغییر داد. با توجه به شرایط بهینه حاصل از این فرایند، داکینگ سایر لیگاندها در پروتئین با همین شرایط انجام می‌شود.

## ۲-۵-۲ آماده سازی پروتئین

صحت نتایج داکینگ به طور مستقیم به کیفیت ساختار کریستالوگرافی جایگاه فعال پروتئین وابسته است. ساختار کریستالوگرافی پروتئین‌ها با اشعه ایکس مشخص می‌شوند. ساختار پروتئین به همراه لیگاند "هولو"<sup>۲</sup> و ساختار پروتئین بدون لیگاند و کوفاکتور "آپو"<sup>۳</sup> نامیده می‌شود. ساختارهای هولو مهم‌ترین روش در شناسایی جایگاه فعال پروتئین هستند. مرکز ثقل لیگاندی که در پروتئین است همان مرکز تقریبی جایگاه فعال پروتئین می‌باشد. از آن جا که حتی بهترین ساختارهای کریستالوگرافی اغلب ارزش تفکیک یک آنگستروم یا بیشتر دارند، ابتدا لیگاند و مولکول‌های آب موجود در ساختار کریستالوگرافی را حذف کرده و با فرمت بانک اطلاعاتی پروتئین<sup>۴</sup> ذخیره می‌کنند و پس از ورود به محیط برنامه، هیدروژن‌هایی را که توسط کریستالوگرافی پروتئین دیده نمی‌شوند، اضافه می‌نمایند.

---

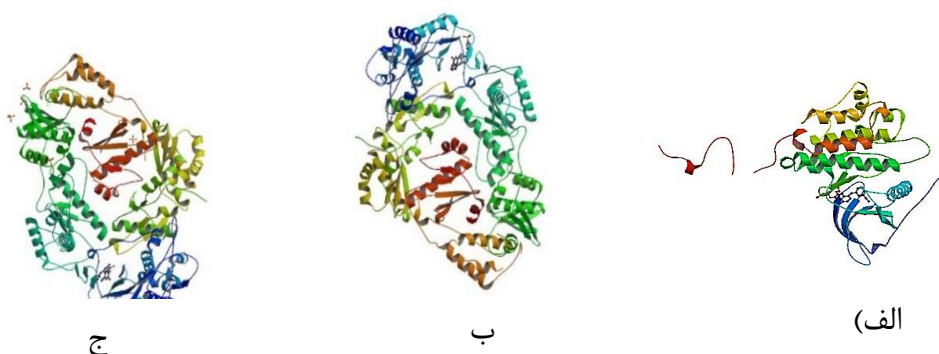
<sup>۱</sup> Root mean square deviation

<sup>۲</sup> HOLO structure

<sup>۳</sup> Apo structure

<sup>۴</sup> Protein data bank

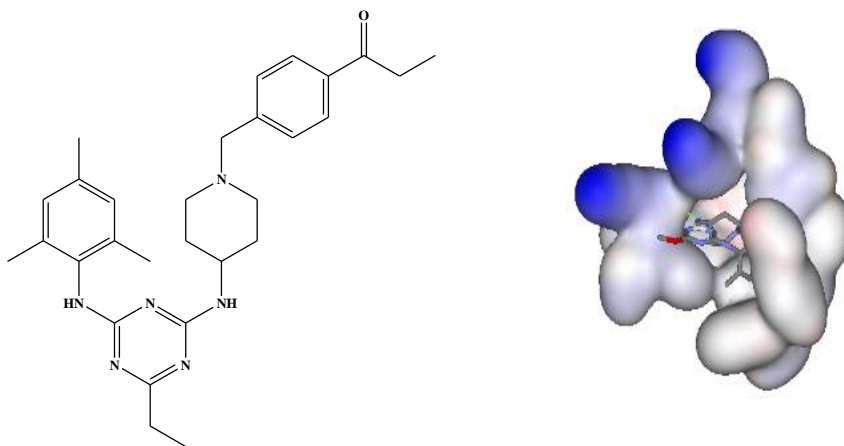
شکل ( ۳-۲ ) ساختارهای کریستالوگرافی سه پروتئین استفاده شده در این پروژه را نشان می‌دهد [۶۲]



شکل ( ۳-۲ ) ساختار کریستالوگرافی پروتئین الف) کد کریستالوگرافی 1M17 ب) کد کریستالوگرافی 3MEC ج) کد کریستالوگرافی 3M8Q [۶۲]

## ۳-۵-۲ ساخت لیگاند

برنامه‌های داکینگ برای تولید لیگاند و قرار دادن آن در جایگاه فعال چندین انتخاب دارند، در خودکارترین حالت، تعدادی از برنامه‌های داکینگ می‌توانند یک مجموعه داده از لیگاندها را گرفته و هر یک را در جایگاه فعال قرار دهند و داکینگ را انجام دهند. با استفاده از این روش کاملاً خودکار، می‌توان هزاران ترکیب را بدون مداخله دستی کاربر آنالیز نمود. در روش دیگر ابتدا همه ساختارهای ممکن تولید شده و در یک فایل ذخیره می‌شوند. سپس این ساختارها به صورت خودکار خوانده شده و میزان امتیاز داکینگ آن‌ها نمره‌دهی می‌شوند. اغلب برنامه‌های داکینگ، یک ساختار ورودی را گرفته و هزاران ساختار تولید می‌کنند، اما فقط بهترین آن‌ها را به عنوان نتایج نهایی ذخیره می‌کنند (۲۰۰-۱۰۰ ساختار به انتخاب کاربر) این امر سبب می‌شود که فضای حافظه کامپیوتر بیش از حد اشغال نشود. نقطه مقابل برنامه‌های خودکار داکینگ، برنامه‌هایی هستند که کاربر لیگاند را داخل جایگاه فعال قرار می‌دهد سپس هر بار داکینگ را انجام می‌دهد. بسیاری از کارهای طراحی دارو بر مبنای ساختار، به صورت دستی و با وارد ساختن ترکیبات در جایگاه فعال پروتئین هدف انجام می‌شود. نمایشی از کمپلکس لیگاند با جایگاه فعال در شکل ( ۴-۲ ) آورده شده است.



شکل ( ۴-۲ ) ساختار کمپلکس لیگاند و پروتئین

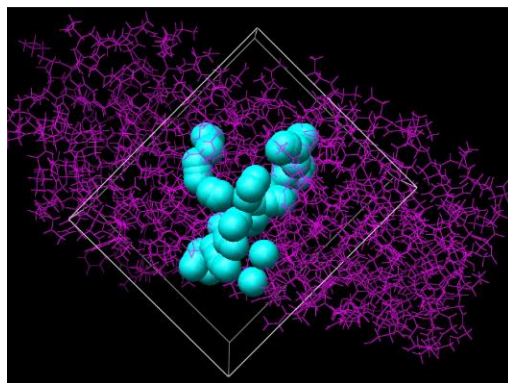
## ۴-۵-۲ تنظیم کردن جعبه شبکه‌ای<sup>۱</sup>

از آن جایی که در محاسبات داکینگ زمان زیادی صرف می‌شود، می‌توان اسیدهای آمینه‌ای را که دور از جایگاه فعال قرار دارند و عموماً اثر معنی‌داری روی نتایج نمره دهی ندارند، در محاسبات در نظر نگرفت. به این منظور یک حد آستانه تعیین شده به طوری که هیچ برهمکنشی برای اسیدهای آمینه‌ای که خارج از این حد از لیگاند قرار دارند، محاسبه نمی‌شود. برای سادگی محاسبات، این مرز معمولاً به صورت یک شبکه مکعبی که شبکه پیوندی<sup>۲</sup> نامیده می‌شود تعیین می‌گردد، و در مورد محاسبه پتانسیل‌های شبکه، تنها پتانسیل نقاطی که داخل این شبکه مکعبی هستند در نظر گرفته می‌شود.

اگر فایل مربوط به ساختار کریستالوگرافی هولو باشد به این معنی که دارای لیگاند کمپلکس شده باشد جایگاه فعال پروتئین مشخص خواهد بود و مختصات جعبه پیوندی همان مختصات جایگاه فعال و به شکل مکعبی با ابعاد یکسان تعریف می‌شود. ابعاد جعبه باید طوری انتخاب شود که اسیدهای آمینه جایگاه فعال را در بر بگیرد. شکل ( ۵-۲ ) جعبه شبکه‌ای در بر گیرنده اسیدهای آمینه را نشان می‌دهد [۶۳، ۶۴].

<sup>۱</sup> Grid box

<sup>۲</sup> Bounding box



شکل ( ۲-۵ ) جعبه شبکه‌ای [۶۴]

## ۵-۵-۲ گزینه‌های داکینگ

در هنگام تنظیم ورودی‌ها برای یک محاسبه داکینگ، گزینه‌هایی برای جایگاه فعال انعطاف‌پذیر، روش‌های نمره دهی، روش‌های جستجو، حلال پوشی، برخورد با جایگاه فعال احاطه شده و غیره موجود هستند.

## ۶-۵-۲ انجام محاسبه‌ی داکینگ

زمانی که گزینه‌های داکینگ تنظیم شدند می‌توان محاسبات داکینگ را انجام داد. این محاسبات گاهی توسط همان کامپیوتری که صفحه‌ی رابط گرافیکی از آن استفاده شده، انجام می‌شوند و گاهی می‌توانند به یک سرور دیگر فرستاده شوند.

## ۷-۵-۲ آنالیز و تحلیل نتایج

مهم‌ترین نتیجه محاسبات داکینگ، انرژی اتصال لیگاند به جایگاه فعال است. بر اساس این انرژی بهترین پیکربندی مهارکننده، در مقایسه با سایر پیکربندی‌ها مشخص می‌شود.

دو مولفه کلیدی الگوریتم جستجو<sup>۱</sup> و الگوریتم رتبه‌بندی یا امتیازدهی<sup>۲</sup> در برنامه داکینگ وجود دارد. الگوریتم جستجو، لیگاند را در موقعیت‌ها و صورت‌بندی‌های متفاوت در جایگاه فعال پروتئین قرار می‌دهد. انتخاب الگوریتم جستجو، مشخص می‌کند که برنامه با چه صحتی در طول فرایند، موقعیت‌های ممکن مولکول را چک می‌کند و چه مدت زمانی برای آن نیاز است. قابل ذکر است که الگوریتم جستجو

<sup>۱</sup> The search algorithm

<sup>۲</sup> The scoring algorithm

صحت نتایج به دست آمده از برنامه داکینگ را مشخص نمی کند. الگوریتم امتیازدهی مسئول تعیین جهت گیری با مناسب ترین انرژی می باشد.

توصیف کننده های حاصل از داکینگ مولکولی را می توان به صورت زیر دسته بندی کرد:

- توصیف کننده های حاصل از برهم کنش لیگاند و پروتئین که به صورت انواع مختلفی از انرژی قابل محاسبه هستند، مانند:

۱. انرژی آزاد اتصال محاسبه شده<sup>۱</sup> (BE)

۲. انرژی واندروالس (vdW) + پیوند هیدروژنی (Hbond) + حلال زدایی (desolv)، (VE)

۳. انرژی بین مولکولی نهایی<sup>۲</sup> (IE)

۴. انرژی الکتروستاتیک<sup>۳</sup> (EE)

۵. مجموع انرژی داخلی نهایی<sup>۴</sup> (TI)

۶. انرژی آزاد چرخشی<sup>۵</sup> (TE)

۷. انرژی سیستم های آزاد<sup>۶</sup> (UE)

۸. ثابت بازداری<sup>۷</sup> (K<sub>i</sub>)

۹. منفی لگاریتم ثابت بازداری (-Log K<sub>i</sub>)

۱۰. انرژی درونی<sup>۸</sup> (U)

- توصیف کننده های حاصل از نوع اسیدهای آمینه که در ارتباط با لیگاند در کمپلکس قرار گرفته اند، که می توان به موارد زیر اشاره کرد:

---

<sup>۱</sup> Estimated Free Energy of Binding

<sup>۲</sup> Final Intermolecular Energy

<sup>۳</sup> Electrostatic Energy

<sup>۴</sup> Final Total Internal Energy

<sup>۵</sup> Torsional Free Energy

<sup>۶</sup> Unbound System's Energy

<sup>۷</sup> Inhibition constant

<sup>۸</sup> Internal energy



۱. هیدروفوبیک<sup>۱</sup>

۲. هیدروفیلیک<sup>۲</sup>

۳. هیدروفیلیک اسیدی<sup>۳</sup>

۴. هیدروفیلیک بازی<sup>۴</sup>

- توصیف کننده‌های حاصل از نوع اسیدهای آمینه بر اساس ساختار اسید آمینه، که شامل موارد زیر می باشد:

۱. دارای زنجیره جانبی آلیفاتیک<sup>۵</sup>

۲. دارای زنجیره‌های جانبی حاوی گروه‌های هیدرکسیلی<sup>۶</sup>

۳. دارای زنجیره‌های جانبی حاوی اتم گوگرد<sup>۷</sup>

۴. دارای زنجیره‌های جانبی حاوی گروه‌های اسیدی<sup>۸</sup> یا آمیدی

۵. دارای زنجیره جانبی حاوی گروه‌های بازی<sup>۹</sup>

۶. دارای حلقه‌های آروماتیک<sup>۱۰</sup>

۷. ایمینواسیدها<sup>۱۱</sup>

در پروژه پیش رو، سعی بر استفاده همزمان از توصیف کننده‌های ساختاری مولکول‌ها و توصیف کننده‌های استخراج شده از داکینگ مشتقات ۴-آنیلینو کوئینازولین، مشتقات با مشابهت ساختاری دی آریل پیریمیدین‌ها و مشتقات آزینی با پروتئین مربوطه بوده تا بتوان از اطلاعات حاصل از برهمکنش

---

<sup>۱</sup> Hydrophobic

<sup>۲</sup> Hydrophilic

<sup>۳</sup> Acidic Hydrophilic

<sup>۴</sup> Basic Hydrophilic

<sup>۵</sup> With aliphatic side chain

<sup>۶</sup> With side chains containing hydroxylic groups

<sup>۷</sup> With side chains containing sulfur atoms

<sup>۸</sup> With side chains containing acidic groups or their amides

<sup>۹</sup> With side chains containing basic group

<sup>۱۰</sup> Containing Aromatic Rings

<sup>۱۱</sup> Imino acids

لیگاند و پروتئین نیز استفاده نمود. بررسی این برهمکنش‌ها حائز اهمیت است زیرا می‌توان از انرژی اتصال مطلوب و پیوندهای برقرار شده با اسیدهای آمینه به اهمیت آن لیگاند پی برد.

## ۸-۵-۲ انواع داکینگ مولکولی

داکینگ مولکولی انواع متفاوتی دارد که شامل موارد زیر می‌باشد.

۱. برهمکنش لیگاند-پروتئین: برهم‌کنش یک مولکول کوچک (لیگاند) با یک ماکرو مولکول (پروتئین)

۲. برهمکنش پروتئین-پروتئین: معمولا محل اتصال سطحی بزرگتر نسبت به اتصال لیگاند-پروتئین دارد.

۳. برهمکنش پروتئین-DNA

## ۹-۵-۲ کاربردهای داکینگ مولکولی

داکینگ مولکولی کاربردهای زیادی در مطالعات مربوط به طراحی داروها دارد که شامل موارد زیر است.

۱. غربال‌گری مجازی<sup>۱</sup>

۲. کشف دارو

۳. پیش‌بینی انرژی آزاد اتصال

۴. برهمکنش پروتئین-پروتئین یا پروتئین-اسید نوکلئیک

۵. مهندسی پروتئین

متداول‌ترین کاربرد داکینگ مولکولی در مطالعات کشف دارو برای بررسی اتصال پروتئین-لیگاند است [۵۵] و در این پروژه از اتصال لیگاند-پروتئین، جهت تولید توصیف‌کننده‌های مهمی اعم از انرژی آزاد اتصال، انرژی واندروالس، انرژی بین مولکولی نهایی، انرژی پیوند هیدروژنی، انرژی الکتروستاتیک،

---

<sup>۱</sup> Virtual screening

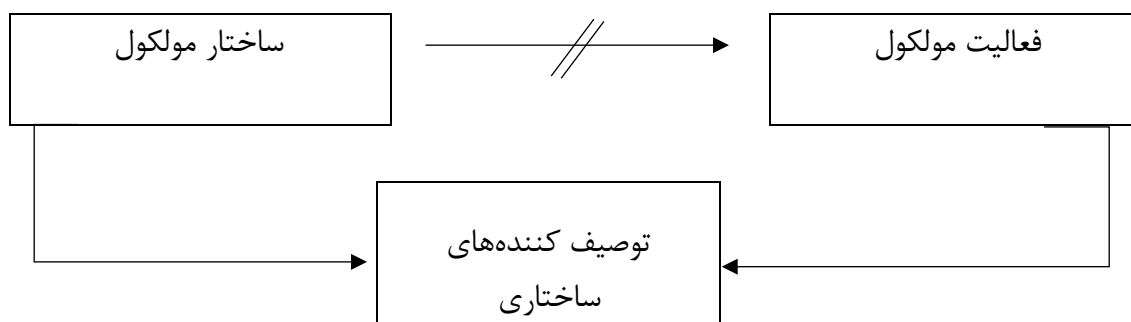
مجموع انرژی داخلی نهایی، انرژی آزاد چرخشی، انرژی درونی و ثابت بازداري تئوری و منفی لگاریتم ثابت بازداري استفاده شده است.

## ۶-۲ اصول مدل سازی QSAR

در مطالعات QSAR هرگز نمی توان به طور مستقیم ارتباط ساختارمولکول با فعالیت بیولوژیکی

آن را بیان کرد، بلکه باید مراحل زیر به ترتیب دنبال شود [۶۵، ۶۶]:

۱. جمع آوری و انتخاب سری داده ها
۲. رسم و بهینه سازی ساختار ترکیبات
۳. محاسبه توصیف کننده ها
۴. انتخاب موثرترین توصیف کننده ها
۵. ایجاد مدل های آماری
۶. تجزیه و تحلیل و ارزیابی توصیف کننده ها و انتخاب مناسب ترین مدل
۷. ارزیابی قدرت پیش بینی و اعتبار مدل انتخاب شده



شکل (۱-۲) شمای کلی از مدل سازی QSAR

دانستن این که چگونه ساختار مولکولی منجر به اثر خاصی بر روی یک سیستم بیولوژیکی می شود، نشان می دهد که کلید مهمی برای دستیابی به اطلاعات در دسترس وجود دارد. توسعه این روابط منجر به ایجاد یک مدل پیش بینی کننده می شود. اگر یک سری مواد شیمیایی وجود داشته باشد و تلاش

شود که یک رابطه بین اثرات بیولوژیکی (به‌عنوان مثال فعالیت) و ویژگی‌های شیمیایی (به‌عنوان مثال ساختار) ایجاد گردد، در این صورت یک ارتباط کمی ساختار-فعالیت یا QSAR تشکیل می‌گردد.

## ۲-۶-۱ جمع آوری و انتخاب سری داده‌ها

اولین مرحله یک مدل سازی QSAR، جمع آوری و انتخاب یک سری مولکولی از منابع قابل اعتماد و در دسترس است. کمیت مورد مدل سازی برای ترکیبات مختلف، بایستی در شرایط عملی یکسان اندازه‌گیری شده باشد تا نتیجه قابل قبول‌تر و مناسب‌تری به‌دست آید. در مدل سازی QSAR مجموعه داده‌ها به سه دسته سری آموزشی<sup>۱</sup>، سری ارزیابی<sup>۲</sup> و سری آزمون<sup>۳</sup> تقسیم می‌شود. به مجموعه داده‌ای که برای برازش مدل استفاده شده است، مجموعه داده‌های آموزشی می‌گویند. با استفاده از سری ارزیابی مدل‌های مختلف ارزیابی شده و بر اساس پارامترهای آماری به‌دست آمده برای این سری بهترین مدل انتخاب می‌شود و در نهایت یک روش برای اندازه‌گیری قدرت پیش‌بینی مدل، آزمودن آن بر روی مجموعه‌ای از داده‌ها است که در برازش آن مدل مورد استفاده قرار نگرفته باشد. در یادگیری ماشین و داده‌کاوی به چنین مجموعه‌ای، مجموعه داده‌های آزمون می‌گویند و با استفاده از سری آزمون قدرت پیش‌بینی و اعتبار مدل، مورد سنجش قرار می‌گیرد. سری آزمون به نحوی انتخاب می‌شود که نماینده جمعیت مولکول‌های سری آموزشی باشد.

## ۲-۶-۲ رسم و بهینه سازی ساختار ترکیبات

برای محاسبه برخی از توصیف‌کننده‌ها به اطلاعاتی از قبیل مختصات سه بعدی و حالت پایدار مولکول نیاز می‌باشد. بنابراین قبل از محاسبه توصیف‌کننده‌ها باید ساختار تمام مولکول‌های موجود در مجموعه داده‌ها رسم و بهینه شوند و پایدارترین حالت آن‌ها با حداقل انرژی مشخص شود. شکل سه بعدی مولکول‌ها توسط نرم افزارهایی همچون ChemDraw [۶۷]، HyperChem و غیره، رسم و سپس

---

<sup>۱</sup> Training set

<sup>۲</sup> Validation set

<sup>۳</sup> Test set

ساختار مولکول به روش‌های مختلف از جمله مکانیک مولکولی<sup>۱</sup>، دینامیک مولکولی<sup>۲</sup> و مکانیک کوانتومی<sup>۳</sup> بهینه می‌شود تا به حالت کمینه انرژی برسد. این کار معمولاً با نرم افزارهایی مانند SYBYL، MOPAC، HyperChem و Gaussian و اغلب با روش‌های مکانیک مولکولی نیمه تجربی انجام می‌پذیرد تا ساختارهای مولکولی به‌طور دقیق بهینه شوند. این روش‌ها به‌طور کلی به دو گروه روش‌های نظری محض و نیمه تجربی تقسیم می‌شوند. روش‌های نظری که امروزه به روش‌های از آغازین<sup>۴</sup> نیز معروف می‌باشند، جهت محاسبه پارامترهای الکترونی ساختاری از حل دقیق معادله شرودینگر استفاده می‌کنند. این کار علاوه بر پیچیدگی بسیار وقت‌گیر نیز هست. بنابراین استفاده از تقریب‌های مناسب بسیار مطلوب بوده و باعث صرفه‌جویی در زمان لازم برای محاسبات خواهد شد. تئوری تقریب اوربیتال مولکولی اغلب از داده‌های تجربی مربوط به اتم‌ها و سیستم‌های مولکولی جهت تخمین کمیت‌های قراردادی که پارامترهای محاسباتی محسوب می‌شوند، استفاده می‌کند. از این رو به این تکنیک‌ها، روش‌های نیمه تجربی گفته می‌شود [۶۸، ۶۹]. روش‌های نیمه تجربی با هامیلتونین PM3 و AM1 در قسمت‌هایی که ساختار الکترونی از اهمیت بیشتری برخوردارند، به دلیل زمان کمتر بهینه سازی مولکولی و کارایی بهتر، کاربرد زیادی دارند پس از رسم و بهینه سازی ساختار ترکیبات اطلاعاتی از قبیل زوایای پیوندی، طول پیوندها، زوایای چرخشی و ... را در اختیار می‌گذارد. ساختارهای بهینه شده را می‌توان به‌عنوان ورودی به سایر نرم افزارها معرفی نمود.

## ۲-۶-۳ محاسبه و غربالگری توصیف کننده‌ها

یکی از مهم‌ترین مراحل مدل سازی تعیین و انتخاب توصیف کننده‌ها می‌باشد. توصیف کننده‌های مولکولی نتیجه نهایی یک استدلال و روش ریاضی هستند که اطلاعات ساختاری مولکولی را به رمز تبدیل می‌کنند و آن‌ها را به صورت یک نماد عددی نشان می‌دهند. در صورتی که توصیف کننده‌های

---

<sup>۱</sup> Molecular mechanic

<sup>۲</sup> Molecular dynamic

<sup>۳</sup> Quantum mechanic

<sup>۴</sup> Ab initio

انتخاب شده مناسب نباشند، مدل سازی با مشکل مواجه شده و مدل مناسب به دست نخواهد آمد. بنابراین در انتخاب توصیف کننده‌ها باید دقت کرد و توصیف کننده‌هایی می‌بایست انتخاب شوند که در ارتباط با ویژگی مورد نظر مولکول بوده و فهم و تفسیر آن‌ها آسان باشد. به طور کلی توصیف کننده‌ها بر اساس چگونگی بیان خصوصیات مولکول به چند دسته تقسیم می‌شوند که معروفترین آن‌ها در اینجا آورده شده است:

۱. توصیف کننده‌های اساسی<sup>۱</sup>: این سری توصیف کننده‌ها عمومی‌ترین توصیف کننده‌های مورد استفاده می‌باشند که ساختار ترکیب را بدون هیچگونه اطلاعاتی درباره هندسه مولکولی آن در اختیار قرار می‌دهند.

۲. توصیف کننده‌های توپولوژیکی<sup>۲</sup>: توصیف کننده‌های توپولوژیکی جزء ساده‌ترین نوع توصیف کننده‌ها می‌باشند و به ساختار فضایی مولکول ارتباطی نداشته و تنها به نوع اتم‌ها، نوع پیوندها و نحوه ارتباط اتم‌ها به یکدیگر وابسته‌اند. این پارامترها را می‌توان بدون بهینه کردن ساختار مولکول محاسبه کرد.

۳. توصیف کننده‌های هندسی<sup>۳</sup>: توصیف کننده‌های هندسی نسبت به نوع توپولوژیکی از اهمیت بیشتری برخوردارند. یکی از خصوصیات بارز این توصیف کننده‌ها، ارتباط آن‌ها با ساختار سه بعدی مولکول‌ها می‌باشد. بنابراین قبل از محاسبه این توصیف کننده‌ها باید پایدارترین ساختار فضایی مولکول تعیین شود.

۴. توصیف کننده‌های الکترونی<sup>۴</sup>: توصیف کننده‌های الکترونی بیانگر خواص الکترونی مولکول‌ها می‌باشند و اطلاعات مفیدی راجع به برهمکنش‌های الکترونی بین مولکولی و درون مولکولی ارائه می‌دهند.

---

<sup>۱</sup> Constitutional descriptors

<sup>۲</sup> Topological descriptors

<sup>۳</sup> Geometrical descriptors

<sup>۴</sup> Electronic descriptors

۵. توصیف کننده‌های فیزیکو- شیمیایی<sup>۱</sup>: بیانگر بعضی از خواص فیزیکی و شیمیایی مولکول‌ها می‌باشند و به ساختار مولکول وابستگی شدیدی دارند مثل لگاریتم ضریب تقسیم آب-اکتانول، نقطه ذوب و نقطه جوش

نرم افزارهای تجاری بسیاری برای محاسبه تعداد وسیعی از توصیف کننده‌ها موجود است. DRAGON، CODESSA، Material studio، BINANA [۲۸] و غیره از جمله نرم افزارهای پرکاربرد در این زمینه هستند.

قبل از مرحله انتخاب متغیر<sup>۲</sup>، باید توصیف کننده‌های نامناسب حذف شوند. این عمل فضای متغیرها را کاهش داده و باعث کاهش زمان محاسبات و بهبود صحت پیش‌بینی‌ها می‌شود. بدین منظور ابتدا متغیرهایی را که برای ۸۰-۹۵ درصد از مولکول‌ها یکسان هستند، حذف کرده، سپس در مرحله دوم همبستگی بین متغیرها محاسبه می‌شود. ماتریس مربع ضرایب همبستگی تشکیل شده و از بین توصیف کننده‌هایی که حاوی اطلاعات مشابهی هستند و ضریب همبستگی بالایی با یکدیگر دارند (۰/۹ > R) یکی از آن‌ها که همبستگی بیشتری با پاسخ مورد نظر دارد، حفظ شده و بقیه حذف می‌شوند.

## ۲-۶-۴ انتخاب موثرترین توصیف کننده‌ها

تعداد توصیف کننده‌های حاصل بسیار زیاد است و از آنجایی که ممکن است همبستگی بالایی بین آن‌ها باشد فرایند مدل سازی را دشوار می‌سازد. برای مدل سازی بهتر باید متغیرهایی با همبستگی بالا را حذف نمود. در مرحله بعد متغیرهای مناسب باید گزینش گردد تا حاوی اطلاعات مفید و ارتباط بالا با پاسخ‌های بیولوژیکی باشند. برای حذف متغیرهایی با ضریب همبستگی بالا ابتدا ماتریس مربع همبستگی<sup>۳</sup> تشکیل داده می‌شود و از بین دو زوج با ضریب همبستگی بالا ( $R > 0/8$ ) آن که ضریب همبستگی کمتری با پاسخ بیولوژیکی دارد حذف می‌شود [۷۰]. برای تشکیل ماتریس مربع همبستگی

---

<sup>۱</sup> Physico-Chemical descriptors

<sup>۲</sup> Feature selection

<sup>۳</sup> Correlation matrix

از نرم افزار<sup>۱</sup> SPSS استفاده می‌شود. این مربع اطلاعاتی راجع به میزان همبستگی بین متغیرهای مستقل و همچنین میزان همبستگی توصیف کننده‌ها با فعالیت دارویی را در اختیار می‌گذارد [۷۱, ۷۲].

از روش‌های متفاوتی برای انتخاب موثرترین توصیف کننده‌ها استفاده می‌شود که می‌توان به روش انتخاب متغیر پیش‌رونده، انتخاب متغیر گام به گام<sup>۲</sup> و الگوریتم ژنتیک اشاره کرد که در این تحقیق از روش انتخاب متغیر رگرسیون گام به گام استفاده شده است.

### ۲-۴-۱ انتخاب متغیر به روش رگرسیون خطی

رگرسیون معادله‌ای برای ایجاد یک رابطه بین یک متغیر از یک طرف و یک یا چند متغیر از طرف دیگر می‌باشد. اگر تنها یک متغیر مستقل داشته باشد، رگرسیون را ساده و در غیر این صورت، رگرسیون را چندگانه می‌گویند [۷۳]. در بسیاری از موارد نمی‌توان تغییرات یک متغیر را فقط به مقادیر یک متغیر دیگر مربوط کرد. به عبارت دیگر برای پیش‌بینی مقادیر یک متغیر، دانستن مقادیر دو یا چند متغیر دیگر لازم است.

اما زمانی که تعداد متغیرهای مستقل مورد مطالعه زیاد باشد لازم است که تکنیکی برای انتخاب متغیرهای مهم و اثرگذار مورد استفاده قرار گیرد تا بتوان متغیرهای معنی‌دار را از متغیرهای بی‌اهمیت تمیز داد. روش‌های رگرسیون متفاوتی برای انتخاب متغیرهای مستقل وجود دارد. از جمله مهم‌ترین آن‌ها می‌توان به انتخاب پیش‌رونده<sup>۳</sup>، حذفی پس‌رونده<sup>۴</sup> و روش گام به گام اشاره کرد.

### ✓ روش انتخاب متغیر پیش‌رونده

در روش انتخاب متغیر پیش‌رونده متغیرهای مستقل بر اساس این که کدام متغیر بیشترین همبستگی را با فعالیت دارویی دارد، انتخاب می‌شود. سپس متغیر مستقلی که آماره‌ی فیشر (F) بزرگتر

<sup>۱</sup> Statistical package for the social sciences

<sup>۲</sup> Stepwise Regression

<sup>۳</sup> Forward

<sup>۴</sup> Backward



و خطای استاندارد کمتری را برای مدل ایجاد می کند وارد معادله مدل می شود.

### ✓ روش انتخاب متغیر حذفی پس رونده

در روش حذفی پس رونده با ایجاد یک رگرسیون خطی ساده که شامل تمام متغیرهای مستقل است شروع شده و سپس سهم هر عبارت در مدل مشخص می شود. متغیری که کمترین تاثیر و پایین ترین سطح معناداری را دارد کنار گذاشته شده و پارامترهای معادله رگرسیون دوباره محاسبه می گردند. به همین ترتیب متغیرهای فاقد سطح معناداری قابل قبول ادامه پیدا می کند و این روند زمانی خاتمه می یابد که خروج یک متغیر باعث کاهش قدرت پیش بینی مدل نسبت به مدل قبلی شود.

### ✓ روش انتخاب متغیر گام به گام

روش رگرسیون گام به گام ترکیبی از روش های پیش رونده و پس رونده است. در این روش متغیرهای مستقل به ترتیب و مشابه با روش پیش رونده وارد معادله می شوند و با ورود هر متغیر جدید کلیه متغیرهای موجود در مدل بررسی شده و اگر هر کدام از آنها سطح معناداری خود را از دست داده باشند قبل از ورود متغیر جدید از مدل خارج می شود. این روند تا هنگامی که ورود متغیر بعدی تغییر شگرفی در سطح معناداری ایجاد نکند ادامه می یابد.

## ۲-۶-۵ ایجاد مدل های آماری

مدل، در واقع یک رابطه ریاضی است که بیان کننده ی رابطه بین متغیر مستقل و فعالیت دارویی است و به کمک آن می توان با داشتن مقادیر متغیر مستقل، فعالیت دارویی را تخمین زد. مرحله اصلی در مطالعات QSAR ساختن مدل می باشد. مدل های QSAR را می توان بر اساس پارامترهایی که به صورت تجربی بدست آمده و یا از طریق تئوری محاسبه شده اند، به زبان ریاضی بیان نمود و اگر نتایج این نوع مطالعات همبستگی قابل قبولی را ارائه کند، علاوه بر شفاف سازی نحوه ی ارتباط بین خواص مولکول ها و ویژگی های ساختمانی آنها، به پژوهشگران در پیش بینی رفتار مولکول های جدید بر اساس رفتار مولکول های مشابه کمک می کند.

چندین روش متفاوت برای ساختن مدل QSAR وجود دارد که می توان به رگرسیون خطی

چندگانه (MLR)، رگرسیون اجزای اصلی (PCR)، کمترین توان‌های دوم جزئی (PLS)، شبکه عصبی مصنوعی (ANN) اشاره کرد. در این پژوهش از روش شبکه عصبی مصنوعی استفاده شده است.

## ۲-۶-۲ مقدمه‌ای بر شبکه عصبی مصنوعی

شبکه عصبی مصنوعی یک سامانه پردازشی داده‌ها است که از مغز انسان ایده گرفته و پردازش داده‌ها را به عهده پردازنده‌های کوچک و بسیار زیادی سپرده که به صورت شبکه‌ای به هم پیوسته و موازی با یکدیگر رفتار می‌کنند تا یک مسئله را حل نمایند. در این شبکه‌ها به کمک دانش برنامه‌نویسی، ساختار داده‌ای طراحی می‌شود که می‌تواند همانند نرون<sup>۱</sup> عمل کند. به این ساختار داده نرون گفته می‌شود. با ایجاد شبکه‌ای بین این نرون‌ها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می‌دهند. در این شبکه عصبی نرون‌ها دارای دو حالت فعال (روشن یا یک) و غیرفعال (خاموش یا صفر) می‌باشند و هر یال (سیناپس یا ارتباط بین گره‌ها) دارای یک وزن<sup>۲</sup> می‌باشد. یال‌های با وزن مثبت موجب تحریک یا فعال نمودن گره فعال بعدی می‌شوند و یال‌های با وزن منفی، گره متصل بعدی را غیرفعال یا مهار می‌کنند [۷۴].

## ۲-۶-۲-۱ ساختمان و عملکرد نرون مصنوعی

یک نرون مصنوعی در حقیقت مدل محاسباتی است که از نرون‌های عصبی واقعی انسان، الهام گرفته است. در شبکه‌های عصبی مصنوعی نیز بلوک‌های ساختاری و یا نرون‌ها، دستگاه‌های محاسباتی خیلی ساده‌ای هستند و ارتباط بین نرون‌ها عملکرد شبکه را تعیین می‌کند و هدف از آموزش شبکه‌های عصبی مصنوعی تعیین ارتباط مناسب، جهت حل مسائل مختلف است. در واقع شبکه‌های عصبی مصنوعی نیز با ایده گرفتن از رفتار شبکه عصبی بیولوژیکی شکل یافته است. برای ایجاد یک شبکه عصبی مصنوعی باید سیستمی طراحی شود که دارای تعدادی ورودی باشد و با توجه به اهمیت

---

<sup>۱</sup> Neuron

<sup>۲</sup> Weight

هر یک، آن‌ها را با یکدیگر جمع جبری کند و توسط یک تابع موسوم به تابع انتقال<sup>۱</sup>، آن‌ها را به نرون‌های دیگر ارسال نماید.

## ۲-۶-۲ تابع انتقال

زمانی که ورودی‌ها با توجه به اهمیت آن‌ها با یکدیگر جمع جبری شوند توسط تابع انتقال به نرون بعدی منتقل می‌شوند. در واقع تابع انتقال یکی از اجزای شبکه عصبی می‌باشد که برحسب کاربرد می‌تواند توابع متفاوتی باشد. دو تابع انتقال رایج در شبکه‌های عصبی که در این پژوهش برای بهینه سازی شبکه به کار گرفته شده است عبارتند از:

### -تابع انتقال لگاریتم سیگموئید<sup>۲</sup> ( $\log \text{sig}$ )

از این تابع انتقال در شبکه‌های پس انتشار استفاده می‌شود این تابع مقادیر ورودی را در محدوده منفی بی‌نهایت تا مثبت بی‌نهایت دریافت کرده و خروجی بین ۰ و ۱ تولید می‌نماید.

### -تابع انتقال تانژانت سیگموئید<sup>۳</sup> ( $\text{tansig}$ )

این تابع انتقال مقادیر ورودی را در محدوده منفی بی‌نهایت تا مثبت بی‌نهایت دریافت کرده و خروجی بین  $-1$  و  $+1$  تولید می‌کند.

## ۲-۶-۷ آموزش شبکه‌های جلو سو با تکنیک پس انتشار<sup>۴</sup>

پس انتشار خطا یک روش متداول آموزش با ناظر برای شبکه‌های جلوسو است یعنی برای به دست آوردن ارتباط بین متغیرهای ورودی و خروجی در یادگیری به الگوی آموزشی نیاز است. به طور کلی آموزش به کمک تکنیک پس انتشار طبق مراحل زیر انجام می‌شود [۷۴].

۱- انتشار ورودی‌ها از نرون‌های ورودی به سمت نرون‌های خروجی

۲- اختصاص ماتریس وزن‌های تصادفی به هریک از اتصالات

---

<sup>۱</sup> Transfer function

<sup>۲</sup> Logarithm sigmoid transfer function

<sup>۳</sup> Hyperbolic tangent transfer function

<sup>۴</sup> Back propagation

۳- مقایسه خروجی‌های شبکه با مقادیر واقعی (مقادیر هدف) و محاسبه‌ی خطای شبکه

۴- پس‌انتشار خطا از نرون‌های خروجی به سمت نرون‌های ورودی و اصلاح وزن‌ها

۵- ارزیابی عملکرد شبکه با توجه به تابع کارآیی تعیین شده

مراحل فوق تا رسیدن به حداکثر تکرار<sup>۱</sup> مجاز تکرار می‌شود و یا این که مقدار تابع کارآیی از مقداری

که تعیین شده کم‌تر باشد. شبکه عصبی انتخاب شده در این پروژه یک شبکه جلوسو با الگوریتم آموزشی

پس انتشار می‌باشد [۷۴].

## ۷-۲ ارزیابی مدل

### ۱-۷-۲ استفاده از پارامترهای آماری

برای اطمینان از این که مدل نهایی، مدل مناسبی است که توانایی پیش‌بینی نمونه‌های مختلفی

از یک جمعیت را داراست، باید مدل را ارزیابی کرد. این ارزیابی از طریق شاخص‌های کمی انجام می‌شود

که به وسیله آن‌ها صحت نتایج ارائه شده توسط مدل مورد سنجش قرار می‌گیرند. فرض کنید  $x_1, x_2, \dots, x_n$

نشان‌دهنده توصیف‌کننده‌ها  $y_1, y_2, \dots, y_n$  نشان‌دهنده پاسخ در مدل مورد بررسی می‌باشند.

برخی از این شاخص‌های کمی عبارتند از:

**ضریب همبستگی:** ساده‌ترین راه برای بررسی میزان همبستگی دو یا چند متغیر، محاسبه آماره‌ی

ضریب همبستگی آن‌هاست. ضریب همبستگی دو متغیر  $x, y$  با رابطه (۱-۲) تعریف می‌شود. مقدار این

آماره بین ۱ تا -۱ متغیر است. مقدار بزرگتر آن نشان‌دهنده‌ی این است که ارتباط خطی بیش‌تری میان

پاسخ و توصیف‌کننده‌ها وجود دارد.

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad \text{رابطه (۱-۲)}$$

**ضریب تعیین:** به‌عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و

نشان‌دهنده‌ی نسبت تغییرات فعالیت دارویی توضیح داده شده توسط متغیر مستقل است. به‌عنوان مثال

---

<sup>۱</sup> Epoch

$R^2$  برابر با ۰/۹۲۸۶ نشان می‌دهد که ۹۲/۸۶ درصد تغییرات در فعالیت دارویی می‌تواند توسط متغیر

مستقل توضیح داده شود. رابطه ریاضی مربوط به ضریب تعیین به صورت زیر است:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad \text{رابطه (۲-۲)}$$

مجموع توان‌های دوم رگرسیون ( $SSR$ )<sup>۱</sup> طبق رابطه (۳-۲) بیانگر مجموع توان‌های دوم انحراف

مقادیر پیش‌بینی شده‌ی متغیر پاسخ از میانگین مقادیر آن است.

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{رابطه (۳-۲)}$$

مجموع توان‌های دوم کامل ( $SST$ )<sup>۲</sup> طبق نشان‌گر مجموع توان‌های دوم انحراف مقادیر واقعی

متغیر پاسخ از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{رابطه (۴-۲)}$$

مجموع توان‌های دوم خطا ( $SSE$ ) نیز مبین مجموع توان‌های دوم انحراف مقادیر واقعی متغیر پاسخ از

مقادیر پیش‌بینی شده برای آن است:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{رابطه (۵-۲)}$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{رابطه (۶-۲)}$$

طبق رابطه (۶-۲) اگر تمام مشاهدات بر روی خط برازش شده قرار گرفته باشند، یعنی به ازای تمام

نقاط  $y_i = \hat{y}_i$  باشد، مقدار  $R^2$  برابر یک می‌شود و هرگونه انحرافی از این حالت باعث می‌شود که مقدار

$R^2$  از یک کوچکتر شود.

**ضریب تعیین تصحیح شده<sup>۳</sup>:** ضریب تعیین تصحیح شده یکی از پارامترهایی است که در ارزیابی و

مقایسه مدل‌های مختلف با تعداد متفاوت متغیرهای پیش‌گو استفاده می‌شود و از آنجایی که  $R^2$

(ضریب تعیین)، تعداد پارامترهای موجود در مدل را به حساب نمی‌آورد، به این منظور از ضریب تعیین

<sup>۱</sup> Sum Square Regression

<sup>۲</sup> Sum Square Total

<sup>۳</sup> Adjusted determination coefficient

تصحیح شده طبق رابطه زیر برای انتخاب مدل برتر استفاده می‌شود:

$$R_{adj}^2 = 1 - \frac{n-1}{n-p-1} \cdot \frac{SSE}{SST} = 1 - (1-R^2) \frac{n-1}{n-p-1} \quad \text{رابطه (۷-۲)}$$

که در این رابطه  $p$  تعداد متغیرهای مستقل و  $n$  تعداد ترکیبات مورد بررسی می‌باشد.

**آماره F:** آزمون F یا آزمون فیشر در واقع آزمون معنی‌دار بودن آماری در تحلیل رگرسیون ساده و چند متغیره است و برابر با نسبت میانگین توان‌های دوم رگرسیون ( $MSR$ ) به میانگین توان‌های دوم باقی‌مانده‌ها ( $MSE$ ) است. بیان ریاضی آن به صورت زیر می‌باشد:

$$F = \frac{MSR}{MSE} = \frac{SSR/df_m}{SSE/df_{res}} \quad , \quad df_m = p \quad , \quad df_{res} = n - p - 1 \quad \text{رابطه (۸-۲)}$$

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

**مجموع توان دوم باقی‌مانده‌ها<sup>۱</sup> (PRESS):** برابر مجموع توان دوم تفاوت بین مقدار کمیت مشاهده شده ( $y_i$ ) و مقدار برآورد شده ( $\hat{y}_i$ ) است.

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{رابطه (۹-۲)}$$

**خطای استاندارد پیش‌بینی<sup>۲</sup> (SEP)**

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{رابطه (۱۰-۲)}$$

**خطای مطلق میانگین<sup>۳</sup> (MAE)**

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad \text{رابطه (۱۱-۲)}$$

**خطای نسبی پیش‌بینی<sup>۴</sup> (REP)**

<sup>۱</sup> Predictive Residual Sum of Squares

<sup>۲</sup> Standard Error of Prediction

<sup>۳</sup> Mean Absolute Error

<sup>۴</sup> Relative Error of Prediction

$$\text{REP}(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n}} \quad \text{رابطه (۱۲-۲)}$$

میانگین توان‌های دوم خطا<sup>۱</sup> (MSE)

$$\text{MSE} = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n} \quad \text{رابطه (۱۳-۲)}$$

میانگین خطای نسبی<sup>۲</sup> (MRE)

$$\text{MRE} = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right|}{n} \times 100 \quad \text{رابطه (۱۴-۲)}$$

تولرانس<sup>۳</sup>: همان‌طور که می‌دانیم وجود یا عدم وجود هم‌خطی<sup>۴</sup> در بین متغیرهای پیش‌بینی‌کننده مسئله مهمی است که باید بررسی شود. پدیده هم‌خطی می‌تواند به یکی از دو دلیل زیر باشد:

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

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

لذا برای کشف پدیده هم‌خطی برای هر متغیر مستقل مقداری محاسبه می‌شود که به آن تولرانس گفته می‌شود. تولرانس برای هر متغیر  $x_i$  از رابطه زیر به دست می‌آید [۷۳]:

$$T_i = 1 - R_i^2 \quad i = 1, 2, 3, \dots, p \quad \text{رابطه (۱۵-۲)}$$

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

<sup>۱</sup> Mean Square Error

<sup>۲</sup> Mean Relative Error

<sup>۳</sup> Tolerance

<sup>۴</sup> Co-linearity

عامل تورم واریانس<sup>۱</sup>: در برخی مواقع از عامل تورم واریانس به جای تلرانس استفاده می‌شود که رابطه آن با تلرانس به صورت زیر است:

$$VIF = \frac{1}{T_i} = \frac{1}{1 - R_i^2} \quad i = 1, 2, 3, \dots, p \quad \text{رابطه (۱۶-۲)}$$

واضح است که اگر  $x_i$  رابطه خطی با سایر متغیرها داشته باشد، آن‌گاه  $R_i^2$  نزدیک به یک است و عامل تورم واریانس بزرگ می‌شود. مقادیر  $VIF_i$  بین محدوده ۱ تا ۱۰ قابل قبول بوده و مقادیر بیشتر از ۱۰ اغلب به‌عنوان علامتی از این‌که داده‌ها مشکل هم‌خطی دارند، تلقی می‌شود. اگر هیچ‌گونه رابطه خطی بین متغیرهای پیش‌بینی کننده نباشد، آن‌گاه  $R_i^2$  صفر و  $VIF_i$  برابر یک است [۷۵].

علاوه بر پارامترهای آماری نام برده، برخی از پارامترهای دیگر نیز برای ارزیابی مدل مورد استفاده قرار می‌گیرند. پارامتر  $R_0^2$  ضریب تعیین مدل با عرض از مبدا صفر است زمانی که نمودار فعالیت پیش‌بینی شده بر حسب فعالیت تجربی رسم شده است.  $R_0'^2$  ضریب تعیین مدل با عرض از مبدا صفر است زمانی که نمودار فعالیت تجربی بر حسب فعالیت پیش‌بینی شده رسم شده است. پارامترهای  $R_m^2$  و  $R_m^2$  اختلاف بین  $R^2$  و  $R_0^2$  را در مدل محدود می‌کند، به طوری که مدل با  $R_m^2$  بیشتر از ۰/۵ قابل قبول بوده و مدل توسعه یافته از قدرت پیش‌بینی بالاتری برخوردار است. معادله ریاضی این پارامتر در جدول (۲-۱) آمده است.  $R_{0,R}^2$  و  $R_{0,R}'^2$  پارامترهایی هستند که به ترتیب نشان‌دهنده‌ی نزدیکی دو پارامتر  $R_0^2$ ،  $R^2$  و  $R_0'^2$  می‌باشند. هر چه این مقادیر به هم نزدیک‌تر باشند و نسبت تفاضل آنها به  $R^2$  کوچکتر از ۰/۱ باشد مدل دارای قدرت پیش‌بینی بالاتری است.  $k$  شیب معادله رگرسیون برای نمودار فعالیت پیش‌بینی شده بر حسب فعالیت تجربی با عرض از مبدا صفر می‌باشد.  $k'$  شیب معادله رگرسیون برای نمودار مقادیر تجربی بر حسب مقادیر پیش‌بینی شده با عرض از مبدا صفر می‌باشد و پارامتر  $k$  و  $k'$  باید در محدوده ۰/۸۵ تا ۱/۱۵ باشند تا مدل توسعه یافته از نظر آماری مورد تایید واقع گردد. سایر پارامترها نیز به همراه نام و فرمول ریاضی آنها در جدول (۲-۱) خلاصه شده است.

<sup>۱</sup> Variance Inflation Factor (VIF)



جدول ( ۲-۱) پارامترهای آماری

شماره	پارامتر	فرمول محاسبه	محدوده قابل قبول
۱	$R_0^2$	-	$R^2$ [۷۶] نزدیک به $R^2$
۲	$R_{0,R}^2$	$R_{0,R}^2 = \frac{(R^2 - R_0^2)}{R^2}$	$< 0.1$ [۷۷]
۳	$R_m^2$	$R^2 \times [1 - (R^2 - R_0^2)^{\frac{1}{2}}]$	$> 0.5$ [۷۸, ۷۷]
۴	$R_0'^2$	-	$R^2$ [۷۶] نزدیک به $R^2$
۵	$R_{0,R}'^2$	$R_{0,R}'^2 = \frac{(R^2 - R_0'^2)}{R^2}$	$< 0.1$ [۷۷]
۶	$R_m'^2$	$R^2 \times [1 - ((R^2 - R_0'^2))^{\frac{1}{2}}]$	$> 0.5$ [۷۸, ۷۷]
۷	$\Delta R_m^2$	$ R_m^2 - R_m'^2 $	$< 0.2$ [۷۹]
۸	k	-	$0.85 < k < 1.15$ [۷۷]
۹	k'	-	$0.85 < k' < 1.15$ [۷۷]

### ۲-۷-۲ استفاده از نمودار خطای باقی مانده‌ها

منظور از عبارت خطای باقی مانده، اختلاف بین مقادیر پیش‌بینی شده و مقادیر تجربی است. اگر مقادیر در دو طرف محور صفر توزیع پراکنده‌ای داشته باشد، این امر نشان‌دهنده‌ی تصادفی بودن خطاست. ولی اگر عمده‌ی نقاط در این نمودار، در یک طرف خط صفر باشند، این بدان معناست که خطای جهت‌داری رخ داده است.

### ۳-۷-۲ استفاده از سری آزمون

اعتبار و اهمیت مدل‌های پیش‌بینی کننده وقتی مشخص می‌شود که بتوانند خاصیت مورد نظر را برای مولکول‌های جدیدی که در دسته‌ی آموزش موجود نیستند پیش‌بینی کنند. بدین منظور از ابتدای کار تعدادی از مولکول‌ها به‌عنوان سری آزمون خارجی کنار گذاشته می‌شود و در نهایت از مدل منتخب بهینه شده، جهت پیش‌بینی خاصیت مورد نظر مولکول‌های موجود در سری تست که در مدل سازی استفاده نشده‌اند، استفاده می‌شود.

### ۴-۷-۲ تکنیک رد مرحله‌ای تک تک<sup>۱</sup>

در روش رد مرحله‌ای تک تک (LOO)، یک ترکیب به طور تصادفی از مجموعه داده‌ها خارج و سپس مدل شبکه عصبی بهینه با استفاده از ترکیبات باقی مانده، برای پیش‌بینی فعالیت دارویی ترکیب

<sup>۱</sup> Leave one out

حذف شده در شرایط بهینه ساخته می‌شود. با استفاده از این فرایند، فعالیت دارویی همه ترکیبات موجود در مجموعه داده‌ها یک بار توسط مدل بهینه پیش بینی می‌شود.

## ۲-۷-۵ آزمون پراکندگی<sup>۱</sup>

آنالیز پراکندگی برای بررسی توزیع داده‌های سری آموزش، ارزیابی و آزمون در فضای شیمیایی استفاده می‌شود. سری داده‌های آزمون باید نماینده مناسبی از مجموعه داده‌های آموزش باشد، از این رو این مفهوم با استفاده از آزمون پراکندگی مشخص می‌شود. به این منظور، فاصله اقلیدسی<sup>۲</sup> مطابق رابطه (۲-۱۷) برای دو ترکیب p و q، با استفاده از مقادیر توصیف کننده‌های مدل محاسبه می‌شود.

$$r_{ij} = \|P_i - Q_j\| = \sqrt{\sum_{k=1}^n (p_{ij} - q_{ij})^2} \quad \text{رابطه (۲-۱۷)}$$

که در معادله فوق  $P_{ij}$  و  $Q_{ij}$  مقادیر توصیف کننده دو ترکیب p و q هستند. میانگین فواصل اقلیدسی برای هر ترکیب  $(P_{ij})$  بین ۰ و ۱ محاسبه و نرمال می‌شود. داده‌های نرمال شده  $r$  برای تمام ترکیبات سری آموزش، ارزیابی و آزمون بر حسب فعالیت‌های تجربی مربوطه رسم می‌شود. توزیع مناسب داده‌ها در کل فضای شیمیایی نشان دهنده ثبات مدل‌های ساخته شده بر اساس این تقسیم بندی می‌باشد [۸۰]

## ۲-۷-۶ قلمرو کاربرد<sup>۳</sup>

یکی دیگر از مراحل ارزیابی مدل، تعیین قلمرو کاربرد آن می‌باشد [۸۱]. قلمرو کاربرد یک محدوده تئوری در فضای شیمیایی است که به وسیله توصیف کننده‌های منتخب و متغیر پاسخ تعریف می‌شود. با استفاده از این روش امکان حضور یا عدم حضور یک ماده شیمیایی در درون قلمرو ساختاری

<sup>۱</sup> Diversity test

<sup>۲</sup> Euclidean distance

<sup>۳</sup> Applicability domain

تعیین می‌شود. به این منظور ابتدا مقادیر انحراف<sup>۱</sup> ( $h$ ) برای هر ترکیب با استفاده از فعالیت پیش‌بینی شده آن توسط مدل QSAR بهینه مطابق با رابطه (۱۸-۲) محاسبه می‌شود.

$$h_i = x_i^T (X^T X)^{-1} x_i \quad \text{رابطه (۱۸-۲)}$$

در معادله فوق  $x$  بردار توصیف کننده ترکیب مورد نظر و  $X$  یک ماتریس  $k \times n$  است،  $k$  تعداد توصیف کننده‌ها و  $n$  تعداد ترکیبات در سری آموزش می‌باشد و علامت  $T$  نشان‌دهنده وارون ماتریس است. حد آستانه انحراف ( $h^*$ ) نیز برابر با  $3p/n$  می‌باشد که  $n$  تعداد ترکیبات در سری آموزش و  $p$  تعداد متغیرهای مدل به علاوه یک است. این مفهوم از رسم مقادیر استاندارد شده باقیمانده‌ها بر حسب مقادیر حد انحراف مشخص شده و برای تمام ترکیبات سری آموزش، مقادیر استاندارد شده باقیمانده بر حسب مقادیر انحراف رسم می‌شود (نمودار ویلیامز<sup>۲</sup>). با رسم این شکل داده‌های دور افتاده (ترکیبات با باقیمانده استاندارد شده بزرگتر از سه واحد انحراف استاندارد) و ترکیبات خارج از قلمرو کاربرد مدل ( $h > h^*$ ) تعیین می‌شوند. ترکیبات با حد انحراف بالا در سری آموزش تا حد زیادی خط رگرسیون را تحت تأثیر قرار داده و باعث می‌شوند خط رگرسیون نزدیک به مقدار مشاهده شده برازش شود. اگر یک ترکیب در سری پیش‌بینی دارای مقدار  $h$  بزرگتر از  $h^*$  باشد به این معنی است که نتیجه پیش‌بینی شده به‌طور قابل توجهی برون یابی شده و بنابراین ممکن است قابل اعتماد نباشد. بنابراین، تلفیق حد انحراف و باقیمانده استاندارد شده برای تشخیص قلمرو کاربرد مدل QSAR استفاده می‌شود.

## ۷-۷-۲ بررسی میزان مشارکت توصیف کننده‌های منتخب در شبکه عصبی

میزان مشارکت توصیف کننده‌های منتخب به صورت زیر تعیین شد:

۱- توصیف کننده مورد نظر به همراه وزن‌های مربوطه‌اش از شبکه بهینه حذف گردید.

۲- با استفاده از بقیه توصیف کننده‌ها مقدار متغیر وابسته برای هر ترکیب سری آموزش به روش

ارزیابی تقاطعی پیش‌بینی شد.

<sup>۱</sup> Leverage

<sup>۲</sup> Williams plot

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

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

۵- سرانجام درصد مشارکت هر توصیف کننده توسط رابطه (۱۲-۳) برآورد شد [۸۲].

$$C_i = \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad \text{رابطه (۲-۱۹)}$$

در این رابطه  $C_i$  درصد مشارکت توصیف کننده حذف شده  $i$  ام،  $N$  تعداد توصیف کننده‌های مدل

و  $\Delta m_i$  میانگین خطای مطلق حاصل از ارزیابی تقاطعی سری آموزش در غیاب توصیف کننده  $i$  ام را

نشان می‌دهد.

فصل سوم

روش

### ۱-۳ نرم افزارهای مورد استفاده

در علم کمومتریکس بسته‌های نرم افزاری متفاوت برای انجام تمام مراحل مدل سازی به کار گرفته می‌شود که در ادامه به اختصار بسته‌های نرم افزاری استفاده شده در این تحقیق معرفی می‌شوند.

#### ۱-۱-۳ بسته نرم افزاری Hyperchem

از بسته نرم افزاری Hyperchem برای رسم شکل مولکول‌ها و بهینه سازی ساختارها با استفاده از روش‌های کوانتومی و مکانیکی، استفاده می‌شود. به کمک این برنامه می‌توان طول پیوند، زاویه پیوندی و زوایای پیچشی را در مولکول تعیین کرد. داده‌های حاصل از این نرم افزار را می‌توان به‌عنوان ورودی به سایر نرم افزارها معرفی نمود. همچنین به کمک این نرم افزار می‌توان تعدادی از توصیف کننده‌ها از جمله حجم مولی و قطبش پذیری را محاسبه کرد.

#### ۲-۱-۳ نرم افزار DRAGON

نرم افزار DRAGON5.5 که توسط گروه تحقیقاتی کمومتریکس میلانو ارائه شده، امکان محاسبه ۳۲۲۴ توصیف کننده مختلف را فراهم می‌کند. جهت محاسبه توصیف کننده به کمک این نرم افزار لازم است ساختار هندسی بهینه مولکول مورد استفاده قرار گیرد. برای این منظور می‌توان ساختار بهینه مولکول‌ها را به صورت فایل‌هایی با فرمت mol، solf، hin ... به‌عنوان اطلاعات ورودی به کار برد.

#### ۳-۱-۳ نرم افزار AutoDock4.2

نرم افزار AutoDock4.2 یک مجموعه از نرم افزار مدل سازی مولکولی است که قابل اجرا بر روی سیستم عامل ویندوز می‌باشد [۸۳]. این نرم افزار رایگان و برای عموم در دسترس می‌باشد. این نرم افزار برای پیش‌بینی چگونگی اتصال مولکول‌هایی به‌عنوان سوبسترا یا دارو به یک مولکول گیرنده با ساختار سه بعدی معلوم طراحی شده است و از انواع الگوریتم‌ها برای ارزیابی انرژی و یافتن موقعیت‌های مناسب اتصال استفاده می‌کند.

نرم افزار Autodock4.2 شامل یک رابط کاربر است که ساختار گرافیکی مولکول‌ها را نمایش می‌دهد.

این نرم افزار قادر به ساخت مولکول‌ها با استفاده از الگوهای ساخت، تغییر نوع اتم و پیوند و ایجاد ساختارهای جدید از ساختارهای موجود است. کار با این نرم افزار آسان است و در ضمن دارای قابلیت طراحی مولکول‌ها و به حداقل رساندن انرژی آن‌ها می‌باشد که باعث بالا رفتن سرعت کار می‌شود.

### ۳-۱-۴ نرم افزار ViewerLite5.0

این نرم افزار امکان بررسی جزئی ماکرومولکول و لیگاند را به کاربر می‌دهد، از جمله می‌توان ماکرومولکول را مشاهده و ویرایش کرد. همین‌طور می‌توان جایگاه فعال پروتئین را مشاهده و تعیین مختصات نمود. علاوه بر این می‌توان طول پیوندهای هیدروژنی و هیدروفوبی مربوط به لیگاند و اسیدهای آمینه موجود در جایگاه فعال را نیز مورد بررسی قرار داد [۸۴]. در این پایان نامه برای آماده سازی لیگاند و پروتئین جهت انجام داکینگ، بررسی حضور پیوندهای هیدروژنی و محاسبه طول پیوندهای هیدروژنی، بررسی اسیدهای آمینه و جدا کردن آن‌ها بر اساس انواع هیدروفیلی و هیدروفوبی و ... از نرم افزار ViewerLite5.0 استفاده شد. از این نرم افزار هم قبل از انجام داکینگ برای آماده سازی لیگاند و پروتئین و هم بعد از داکینگ برای بررسی برهم کنش‌های لیگاند-پروتئین استفاده می‌شود.

### ۳-۱-۵ نرم افزار Notepad++

Notepad++ برنامه‌ای برای ویرایش کد اصلی برنامه‌ها و جایگزینی مناسب برای Notepad++ ویندوز است. این نرم افزار، زبان‌های برنامه نویسی متعددی را پشتیبانی می‌نماید. امکانات این برنامه شامل تفکیک و قالب بندی کدها، تکمیل خودکار کدها، ویرایش همزمان چند فایل، قابلیت جستجوی کامل و تبدیل فرمت‌های متفاوت عکس به مختصات عددی و ... می‌باشد. در این پایان نامه خروجی داکینگ که به فرمت 'dlg' می‌باشد با نرم افزار Notepad++ نمایش داده شد و تمامی توصیف کننده‌ها با استفاده از این نرم افزار بازخوانی و استخراج شد.

### ۳-۱-۶ نرم افزار SPSS

SPSS [۷۲] که نخستین نسخه آن در سال ۱۹۷۰ توسط جمعی از فارغ التحصیلان دانشگاه استانفورد آمریکا ارائه شده، امکان تجزیه و تحلیل آماری داده‌ها را فراهم می‌آورد. برخی از قابلیت‌های این بسته نرم افزاری عبارتند از:

- محاسبه میانگین ساده برای داده‌ها
  - نمایش اطلاعات به صورت متنوع در قالب نمودار و جدول
  - انجام رگرسیون تک توصیف کننده و چند توصیف کننده
- در این پایان نامه برای محاسبه همبستگی و هم‌خطی موجود بین توصیف کننده‌ها و برای انجام عملیات انتخاب توصیف کننده به روش رگرسیون گام به گام (SR) از SPSS v.21 استفاده شد.

### ۳-۱-۷ نرم افزار MATLAB

نرم افزار MATLAB [۸۵] یکی از جامع‌ترین و کارآمدترین نرم افزارهای علمی و محاسباتی است که طی چند سال گذشته تهیه و به بازار عرضه شد و در سال‌های اخیر با تدوین نسخه‌های جدیدتر و کامل‌تر روز به روز بر غنای آن افزوده می‌شود.

MATLAB به معنای آزمایشگاه ماتریس است. ورودی‌ها اساساً به صورت ماتریس در نظر گرفته می‌شوند و هیچ نیازی به مشخص کردن ابعاد ماتریس نمی‌باشد. در MATLAB حتی اعداد اسکالر، ماتریس‌های (x) به حساب می‌آیند و بردارها، حالت خاصی از ماتریس‌های سطری یا ستونی در نظر گرفته می‌شوند. از جمله کاربردهای جالب آن، که مورد توجه شیمیدانان قرار گرفته است استفاده از جعبه ابزار شبکه عصبی مصنوعی و جنگل‌های تصادفی است که در این پژوهش از شبکه عصبی مصنوعی استفاده شده است. اطلاعات شیمیایی به‌عنوان ورودی و به صورت یک ماتریس به نرم افزار داده می‌شود. در محیط این برنامه با استفاده از آرایه‌ها و فرامین موجود امکان مدل سازی غیرخطی فعالیت دارویی



با ساختار ترکیبات فراهم می‌گردد.

در این بخش به نحوه‌ی رسم و بهینه‌سازی ساختارها، استخراج توصیف‌کننده‌های ساختاری و داکینگ مولکولی و مراحل بهینه‌سازی مدل شبکه عصبی مصنوعی اشاره شده است.

### ۲-۳ رسم و بهینه‌سازی ساختار ترکیبات مورد مطالعه

به منظور محاسبه و استخراج توصیف‌کننده‌های ساختاری، ساختار مولکولی ترکیبات در نرم افزار Hyperchem رسم شدند و با استفاده از روش نیمه تجربی کوانتومی AM1 تا رسیدن جذر میانگین مربعات انرژی به  $0.001$  کیلو کالری بر مول بهینه‌سازی شدند. ساختارهای بهینه شده با دو فرمت (\*.hin) و (\*.mol) ذخیره شدند. از ساختارهای بهینه شده با فرمت \*.hin\* در نرم افزار DRAGON برای محاسبه توصیف‌کننده‌های ساختاری استفاده شد.

### ۳-۳ استخراج توصیف‌کننده‌های حاصل از داکینگ مولکولی

به منظور استخراج توصیف‌کننده‌هایی از نوع برهمکنش لیگاند-گیرنده، شرایط بهینه داکینگ با استفاده از داکینگ لیگاند کریستالوگرافی (کوگنیت لیگاند) و گیرنده به روش اعتبارسنجی به‌دست آمد. سپس در شرایط بهینه داکینگ، کلیه لیگاندها به‌طور جداگانه در جایگاه فعال پروتئین مربوطه داک شدند. در نهایت توصیف‌کننده‌های داکینگ مولکولی با استفاده از اطلاعات برهمکنش پایدارترین کمپلکس لیگاند-گیرنده استخراج شدند. در ادامه روش کار به تفصیل آورده شده است.

### ۳-۳-۱ استخراج و آماده‌سازی لیگاند موجود در پروتئین

انجام فرایند اعتبارسنجی جهت یافتن شرایط بهینه قبل از اجرای داکینگ مولکولی لیگاندها- پروتئین ضروری است. اعتبارسنجی داکینگ به منظور رسیدن به شرایط بهینه با ورود و خروج متوالی لیگاند کریستالوگرافی (لیگاند موجود در ساختار پروتئین) به جایگاه فعال پروتئین حاصل می‌گردد. لذا لازم است لیگاند کریستالوگرافی و پروتئین جهت انجام فرایند اعتبارسنجی آماده شوند. استخراج و آماده‌سازی لیگاند کریستالوگرافی به صورت زیر انجام شد:

با استناد به گزارش‌های موجود و با توجه به مشابهت ساختاری بین لیگاند موجود در جایگاه فعال پروتئین و لیگاندهای هدف، ساختار کریستالوگرافی پروتئین با فرمت pdb از بانک اطلاعاتی پروتئین [۶۲] دریافت شد. برای استخراج و ذخیره‌سازی لیگاند کریستالوگرافی، ساختار کریستالوگرافی پروتئین در نرم افزار ViewerLite5.0 فرا خوانده شد و همه زنجیره‌های اسید آمینه‌ای، مولکول‌های آب و کوفاکتورها حذف شدند و لیگاند باقی‌مانده با فرمت pdb ذخیره شد. سپس با توجه به جهت‌گیری ویژه لیگاند کریستالوگرافی در جایگاه فعال پروتئین، مختصات مرکز ثقل<sup>۱</sup> لیگاند به‌عنوان مختصات جایگاه فعال پروتئین مربوطه در نظر گرفته شد و در مراحل بعدی داکینگ مولکولی مورد استفاده قرار گرفت.

### ۳-۳-۲ آماده سازی پروتئین

به منظور آماده سازی پروتئین جهت استفاده در فرایند اعتبارسنجی و داکینگ لیگاندها، ساختار کریستالوگرافی با فرمت pdb از بانک اطلاعاتی پروتئین [۶۲] دریافت شد. لازم به ذکر است که مقدار ارزش تفکیک<sup>۲</sup> کمتر از  $3\text{\AA}$  برای ساختارهای کریستالوگرافی پروتئین در مطالعات داکینگ مولکولی قابل قبول هستند [۶۲]. پس از فراخوانی مجزای ساختار کریستالوگرافی در نرم افزار ViewerLite5.0، همه زنجیره‌ها به جز زنجیره اصلی (زنجیره حاوی لیگاند کریستالوگرافی)، کوفاکتورها، مولکول‌های آب و لیگاندهای کریستالوگرافی پروتئین حذف شدند و باقی‌مانده ساختار که شامل جایگاه فعال می‌باشد مجدداً با فرمت pdb ذخیره شد. آماده سازی بیشتر پروتئین در نرم افزار Autodock4.2 انجام شد.

### ۳-۳-۳ فرایند اعتبارسنجی (داک-ریداک)

تعیین اعتبار داکینگ مولکولی با استفاده از فرایند اعتبارسنجی انجام می‌شود به طوری که ساختار لیگاند کریستالوگرافی به‌طور متوالی و در شرایط متفاوت در جایگاه فعال پروتئین داک می‌شود و با استفاده از آنالیز نتایج خروجی داکینگ لیگاند کریستالوگرافی-پروتئین، شرایط بهینه تعیین می‌گردد.

<sup>۱</sup> Center box

<sup>۲</sup> Resolution value

انرژی اتصال، تعداد خوشه‌ها، تعداد کنفورماسیون‌های موجود در خوشه برتر و زمان اجرا از جمله معیارهای تعیین شرایط بهینه داکینگ مولکولی هستند [۸۶]. نحوه انجام فرایند اعتبارسنجی در نرم افزار Autodock4.2 به شرح زیر می‌باشد.

### ۱-۳-۳ فراخوانی<sup>۱</sup> لیگاند و پروتئین

در این مرحله پس از استخراج لیگاند کریستالوگرافی و آماده سازی پروتئین (بخش ۲-۳-۳)، ساختار پروتئین و لیگاند در نرم افزار AutoDock4.2 فراخوانی شدند. سپس همه هیدروژن‌ها برای جبران کمبود هیدروژن ساختار کریستالوگرافی اضافه شدند [۶۳, ۶۴, ۸۷]. هیدروژن‌های غیر قطبی متصل به هر اتم کربن ادغام شدند و بار الکتریکی کلمن<sup>۲</sup> نیز اضافه گردید.

### ۲-۳-۳ تنظیم فاصله بین نقاط شبکه، ابعاد شبکه و مختصات مرکز ثقل

برای رسیدن به یک داکینگ قابل قبول، باید شبکه‌ای با ابعاد مشخص در راستای محورهای سه‌گانه مختصات و در برگیرنده جایگاه فعال گیرنده در بخش grid نرم افزار تعریف شود. لذا در این پژوهش با توجه به دستورالعمل‌های اجرای داکینگ مولکولی موجود در مقالات، شبکه‌ای با ابعاد  $A^\circ$   $60 \times 60 \times 60$  و فاصله بین نقاط  $A^\circ$   $0.375$  در حدود یک چهارم طول پیوند یگانه کربن-کربن در نظر گرفته شد [۸۸] و با توجه به توضیحات بخش (۱-۳-۳)، مختصات مرکز شبکه برابر با مختصات مرکز ثقل لیگاند کریستالوگرافی تعریف شد.

### ۳-۳-۳ انتخاب الگوریتم جستجو در داکینگ مولکولی

فرایند داکینگ مولکولی، با استفاده از الگوریتم‌های متفاوت ژنتیک الگوریتم لامارکین<sup>۳</sup>، ژنتیک الگوریتم، جستجوی محلی و ارزیابی انرژی اجرا می‌شود. در الگوریتم ژنتیک لامارکین به دلیل این که لیگاند درجه آزادی بالاتری نسبت به سایر الگوریتم‌های جستجو دارد، روش اجرایی کارآمدتری نسبت به سایر الگوریتم‌ها در جستجوی کنفورماسیونی در Autodock4.2 می‌باشد. علاوه بر این، این الگوریتم

<sup>۱</sup> Call

<sup>۲</sup> Kollman

<sup>۳</sup> Lamarckian

نتایج داکینگ تکرار پذیرتری را نسبت به سایر الگوریتم‌ها ارائه می‌دهد [۶۴]. در این پروژه نیز از الگوریتم ژنتیک لامارکین استفاده شده است.

### ۳-۳-۴ تنظیم پارامترهای الگوریتم ژنتیک لامارکین

به منظور یافتن بهترین حالت اتصال بین لیگاند-گیرنده، پارامترهای جستجوی الگوریتم ژنتیک برای فرایند اعتبارسنجی و داکینگ لیگاندها-پروتئین باید تنظیم گردد. افزایش تعداد دفعات اجرای الگوریتم ژنتیک باعث افزایش دقت و صحت محاسبات می‌گردد. از طرفی با افزایش سطح این فاکتور زمان اجرای شبیه‌سازی داکینگ مولکولی نیز افزایش می‌یابد. لذا برای حصول به دقت و صحت بالا و زمان مناسب، تعداد دفعات اجرای این فاکتور در عملیات اعتبارسنجی (داک-ریداک) بهینه می‌شود. به این منظور، حداکثر تعداد ارزیابی انرژی<sup>۱</sup> در نرم افزار Autodock4.2 در مقدار متوسط ۲۵۰۰۰۰۰ [۸۹]- [۹۱] و تعداد اجرای الگوریتم ژنتیک در سه مقدار ۱۰۰، ۱۵۰ و ۲۰۰ قرار داده شد. سایر پارامترهای جستجوی الگوریتم در حالت پیش فرض برنامه قرار داده شد.

### ۳-۳-۵ تنظیم گزینه‌های اجرای داکینگ

در این بخش گزینه‌های میزان جابجایی لیگاند در فضا در راستای محور  $x$ ،  $y$  و  $z$  (انتقال<sup>۲</sup>)، دوران در فضای سه بعدی حول محوره‌های  $x$ ،  $y$  و  $z$  (چهارگان<sup>۳</sup>) و میزان چرخش<sup>۴</sup> در مقادیر پیش فرض قرار داده شد. پس از اعمال تمامی گزینه‌های مربوط به انجام فرایند داکینگ مولکولی، شرایط بهینه با آنالیز فایل خروجی (dlg) به دست آمد. با توجه به معیارهای اشاره شده در فرایند اعتبارسنجی (بخش ۳-۳-۳) رسیدن به RMSD کمتر از  $2\text{\AA}$ ، انرژی اتصال منفی‌تر، تعداد خوشه کمتر، حداکثر تعداد کنفورماسیون در خوشه برتر با انرژی منفی‌تر [۸۶، ۹۲] و در نهایت زمان اجرای کمتر مطلوب می‌باشد. تمامی پارامترهای ذکر شده برای تعداد اجراهای متفاوت الگوریتم ژنتیک لامارکین به دست آمد و با توجه به

<sup>۱</sup> Maximum Number of evaluations

<sup>۲</sup> Translation

<sup>۳</sup> Quaternion

<sup>۴</sup> Torsion

نتایج، حالت بهینه برای ادامه فرایند داکینگ، انتخاب گردید.

### ۳-۳-۴ داکینگ لیگاند های مورد مطالعه در پروتئین های مربوطه

به منظور استخراج توصیف کننده های داکینگ مولکولی، لیگاندهای مورد مطالعه باید در شرایط بهینه حاصل از فرایند اعتبارسنجی در پروتئین مربوطه داک شوند. از این رو جهت انجام فرایند داکینگ مولکولی، ساختارهای لیگاندها و پروتئین مربوطه نیز آماده سازی شدند. در ادامه مراحل مورد نیاز جهت استخراج توصیف کننده های داکینگ مولکولی توضیح داده می شود.

### ۳-۳-۴-۱ آماده سازی لیگاندهای مورد مطالعه

لیگاندهای مورد مطالعه مطابق با مراحل ذکر شده در بخش (۳-۲) رسم و بهینه سازی شدند. ساختارهای بهینه شده با فرمت MOLfile و با پسوند mol ذخیره شدند و پس از فراخوانی در نرم افزار ViewerLite5.0 با پسوند pdb ذخیره شدند و در نهایت به عنوان ورودی های نرم افزار AutoDock4.2 مورد استفاده قرار گرفتند.

### ۳-۳-۴-۲ اجرای عملیات داکینگ مولکولی لیگاندها و پروتئین

با توجه به کلیه نکات اشاره شده در بخش (۳-۳-۲)، پروتئین آماده شد و در نرم افزار داکینگ فراخوانی شد و تمامی تنظیمات مربوط به افزایش بارکلمن، افزایش هیدروژن، تعریف مختصات جایگاه فعال و شبکه روی آن اعمال گردید و در مرحله نهایی الگوریتم ژنتیک لامارکین برای اجرای فرایند داکینگ با استفاده از شرایط بهینه مورد استفاده قرار گرفت. تمامی لیگاندها به صورت جداگانه با پروتئین مربوطه در نرم افزار Autodock4.2 فراخوانی شده و با شرایط یکسان تعیین و توسط فرایند اعتبارسنجی داک شدند. پس از اتمام اجرای داکینگ مولکولی، نتایج خروجی با فرمت dlg ذخیره شدند. در نهایت توصیف کننده های عددی با اطلاعات ارزشمند حاصل از برهمکنش لیگاند-پروتئین در خوشه ی برتر با بیشترین صورت بندی و کمترین انرژی اتصال استخراج و آنالیز شدند.

### - استخراج توصیف کننده های حاصل از داکینگ مولکولی

پس از آنالیز خروجی های داکینگ لیگاندها-پروتئین، توصیف کننده هایی از نوع انرژی با استفاده

از نرم افزار Notepad++ به صورت عددی استخراج شدند. این توصیف کننده‌ها عبارتند از:

۱. انرژی آزاد اتصال محاسبه شده (BE)

۲. انرژی واندروالس (vdW) + پیوند هیدروژنی (Hbond) + حلال زدایی (desolv)، (VE)

۳. انرژی بین مولکولی نهایی (IE)

۴. انرژی الکتروستاتیک (EE)

۵. مجموع انرژی داخلی نهایی (TI)

۶. انرژی آزاد چرخشی (TE)

۷. انرژی سیستم‌های آزاد (UE)

۸. ثابت بازداری (Ki)

۹. منفی لگاریتم ثابت بازداری (-Log Ki)

۱۰. انرژی درونی (U)

### ۳-۴ انتخاب موثرترین توصیف کننده‌ها

مهمترین مسئله در توسعه‌ی مدل‌های QSAR انتخاب توصیف کننده‌هایی است که ارتباط بسیار زیادی با فعالیت مورد نظر داشته باشند. این احتمال وجود دارد که تعدادی از توصیف کننده‌های محاسبه شده دارای همبستگی بالا با یکدیگر باشند و اثر یکسان بر روی فعالیت بازدارندگی داشته باشند. همچنین بدیهی است که هر چه تعداد توصیف کننده‌های مستقل کمتر باشد دسترسی به یک مدل مطلوب آسان‌تر است و از پیچیدگی محاسبات جلوگیری به عمل می‌آید [۷۰]. از این رو برای حذف توصیف کننده‌های همبسته از نرم افزار SPSS استفاده شد، به طوری که ماتریس مربع ضرایب همبستگی تشکیل شد و پس از بررسی همبستگی بین متغیرها، از میان توصیف کننده‌هایی با ضریب همبستگی بزرگتر از ۰/۹، آن توصیف کننده‌ای که همبستگی کمتری با فعالیت دارویی داشت، حذف گردید. در این مطالعه توصیف کننده‌های محاسبه شده به عنوان توصیف کننده‌های مستقل و منفی لگاریتم فعالیت

دارویی به‌عنوان متغیر پاسخ در نظر گرفته شد. لازم به ذکر است که انتخاب مهمترین توصیف کننده با استفاده از داده های سری آموزش انجام شد.

### ۵-۳ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه

بعد از جمع آوری و بهینه سازی ساختار مجموعه داده‌ها، توصیف کننده‌ها محاسبه شدند. نظر به این که پدیده‌های بیولوژیکی اغلب دارای ویژگی های غیر خطی هستند [۹۳]، لذا در این پژوهش از مدل ANN به‌عنوان یک روش غیر خطی، برای کشف رابطه ممکن بین توصیف کننده‌های ورودی و فعالیت دارویی ترکیبات مورد مطالعه استفاده شدند. در ادامه مراحل ساخت و بهینه سازی مدل ANN به تفصیل شرح داده می‌شود.

#### مدل شبکه عصبی مصنوعی

شبکه طراحی شده در این پژوهش تحقیقاتی، شبکه پیشرو با الگوریتم آموزش پس انتشار می‌باشد که الگوریتم آن با برنامه متلب نوشته شد. با توجه به نبود هیچ قاعده‌ی تئوری یا تجربی که بتوان به وسیله‌ی آن تعداد لایه‌های پنهان در یک شبکه را تعیین نمود، انتخاب یک لایه پنهان نتایج رضایت‌بخشی را در مطالعات محاسباتی شیمی ایجاد می‌کند [۹۴, ۹۵]. لذا در این تحقیق شبکه عصبی سه لایه متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی به کار گرفته شد.

برای یافتن شرایط بهینه و رسیدن به بهترین معماری ANN با صحت پیش‌بینی بالا، از توابع آموزشی لونبرگ-مارکوارت<sup>۱</sup> (LM)، تنظیم بیزی<sup>۲</sup> (BR) و دوتابع انتقال متداول تانژانت سیگموئید و لگاریتم سیگموئید در لایه پنهان استفاده شد. در هر یک از شبکه‌های عصبی طراحی شده، بهینه سازی همزمان پارامترهای شبکه مانند تعداد ورودی، تعداد نرون‌های لایه پنهان (گره) و تعداد دور آموزشی انجام شد. معیار بهینه سازی، حداقل میانگین مربعات خطای (MSE) فعالیت پیش‌بینی شده به وسیله مدل و فعالیت تجربی قرار گرفت. پس از تقسیم‌بندی داده‌ها، توصیف کننده‌ها به‌عنوان ورودی شبکه و

---

<sup>۱</sup> Levenberg-Marquardt

<sup>۲</sup> Bayesian regularization

فعالیت دارویی به‌عنوان خروجی شبکه در نظر گرفته شدند. به منظور انجام فرایند بهینه‌سازی هر شبکه، تعداد ورودی‌ها از ۲ تا تعداد توصیف‌کننده‌های انتخاب شده، تعداد گره در لایه میانی از ۲ تا ۱۰ با گام ۱ و تعداد دور آموزشی از ۲ تا ۵۰ با گام ۲ به‌طور همزمان بهینه شد. شبکه عصبی بهینه در حالات متفاوت با توجه به حداقل مقدار MSE به‌دست آمد و فعالیت دارویی سری آزمون پیش‌بینی شد. در نهایت مدل توسعه یافته با روش‌ها و پارامترهای آماری اشاره شده در بخش (۲-۷)، مورد ارزیابی قرار گرفت.

### ۳-۶ ارزیابی مدل

پس از ساخت و بهینه‌سازی مدل شبکه عصبی، بذای ارزیابی مدل پارامترهای آماری متفاوت برای سری آزمون و روش رد مرحله‌ای تک تک برای همه ترکیبات محاسبه می‌شود. نمودار خطای باقیمانده و بررسی توزیع پراکندگی نتایج حول محور صفر و نمودار گستره کاربرد نیز برای مدل توسعه یافته رسم می‌شود.



فصل چہارم

نتیجہ تجربی

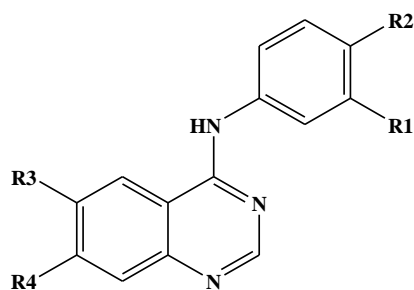
## ۴-۱ مدل سازی فعالیت دارویی مشتقات ۴-آنیلینوکوئینازولین به عنوان ترکیبات فعال

### دارویی ضد سرطان

در این تحقیق، از شبکه‌ی عصبی مصنوعی به عنوان یک روش مدل سازی در مطالعه‌ی ارتباط ساختار-فعالیت برای پیش‌بینی قدرت بازدارندگی ( $-\log(\text{IC}_{50})$ ) مشتقات ۴-آنیلینوکوئینازولین در برابر تیروزین کیناز استفاده شده است. توصیف کننده‌های مولکولی حاصل از برهمکنش‌های موثر لیگاند-پروتئین با استفاده از داکینگ مولکولی و بهترین موقعیت لیگاند در جایگاه فعال پروتئین، استخراج و به عنوان توصیف کننده در توسعه مدل QSAR به کار گرفته شدند. با توجه به اینکه توصیف کننده‌های داکینگ مولکولی فاقد اطلاعات ساختاری هستند، ارتباط مناسب بین ساختار-فعالیت وجود ندارد و اغلب مدل‌های ساخته شده با توصیف کننده‌های داکینگ مولکولی دارای پارامترهای آماری قابل قبولی نمی‌باشند از این رو، در راستای بهبود نتایج مدل سازی از توصیف کننده‌های تابع توزیع شعایی (RDF)، که حاوی ویژگی‌های شیمیایی ساختار لیگاندها می‌باشند، استفاده شد. لازم به ذکر است که این گروه از توصیف کننده‌ها حاوی اطلاعات اساسی در مورد فاصله‌های بین اتمی در طول مولکول، طول پیوند، نوع حلقه و صفحه‌ای و غیر صفحه‌ای بودن سیستم مولکولی می‌باشند و به همین دلیل به تنهایی برای ساخت مدل های QSAR استفاده می‌شوند [۹۶]. با این حال، توصیف کننده‌های RDF فاقد اطلاعات مربوط به برهم کنش های لیگاند-پروتئین می‌باشند. بنابراین در این مطالعه تلاش شد که با ترکیب توصیف کننده های RDF و توصیف کننده‌های داکینگ مولکولی، مدل شبکه عصبی با قدرت پیش‌بینی بالا توسعه یابد و تأثیر این توصیف کننده ها بر عملکرد مدل QSAR بررسی شد. در نهایت، مدل توسعه یافته با استفاده از پارامترهای آماری ارزیابی شد.

#### ۱-۱-۴ مجموعه داده‌ها

۴۳ مولکول مورد استفاده در این تحقیق از مطالعه نولوی<sup>۱</sup> و همکاران استخراج شده است [۹۷]. ساختار اصلی این ترکیبات در شکل (۱-۴) و مقادیر عددی  $pIC_{50}$  و جزئیات استخلاف‌های هر ترکیب در جدول (۱-۴) نشان داده شده است. غلظت بازدارندگی ( $IC_{50}$ ) غلظتی از دارو است که منجر به ۵۰٪ بازدارندگی در محیط کشت مصنوعی شود. در این تحقیق سعی شده است با ارائه یک مدل پیشگو، قدرت بازدارندگی هر ترکیب با هر نوع استخلاف در هر موقعیتی، بدون نیاز به سنتز و آزمایش ترکیب پیش‌بینی شود.



شکل (۱-۴) اسکلت اصلی ترکیبات مورد مطالعه

---

<sup>۱</sup> Noolvi

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

شماره ترکیب	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	$pIC_{50}$
۱	H	H	H	H	۶/۴۶
۲ <sup>t</sup>	CH <sub>3</sub>	H	H	H	۶/۰۴
۳ <sup>t</sup>	Cl	H	H	H	۷/۶۳
۴	Br	H	H	H	۷/۵۶
۵	I	H	H	H	۷/۰۹
۶	CF <sub>3</sub>	H	H	H	۶/۲۳
۷ <sup>t</sup>	Br	H	NO <sub>2</sub>	H	۶/۰۴
۸	Br	H	OCH <sub>3</sub>	H	۷/۵۲
۹	Br	H	H	NO <sub>2</sub>	۶/۰۰
۱۰ <sup>t</sup>	Br	H	H	OCH <sub>3</sub>	۸/۰۰
۱۱	Br	H	OH	OH	۹/۷۶
۱۲	Br	H	NH <sub>2</sub>	NH <sub>2</sub>	۹/۹۲
۱۳ <sup>t</sup>	F	H	H	H	۷/۲۵
۱۴	H	H	OCH <sub>3</sub>	H	۷/۲۴
۱۵	H	H	NH <sub>2</sub>	H	۶/۱۱
۱۶	CF <sub>3</sub>	H	NH <sub>2</sub>	H	۶/۲۴
۱۷	H	H	H	NH <sub>2</sub>	۷/۰۰
۱۸	CF <sub>3</sub>	H	H	NH <sub>2</sub>	۸/۴۸
۱۹	F	H	H	NO <sub>2</sub>	۵/۲۱
۲۰	Cl	H	H	NO <sub>2</sub>	۶/۰۹
۲۱	I	H	H	NO <sub>2</sub>	۶/۲۶
۲۲ <sup>t</sup>	H	H	OCH <sub>3</sub>	OCH <sub>3</sub>	۷/۵۳
۲۳	F	H	OCH <sub>3</sub>	OCH <sub>3</sub>	۸/۴۲
۲۴	Cl	H	OCH <sub>3</sub>	OCH <sub>3</sub>	۹/۵
۲۵	I	H	OCH <sub>3</sub>	OCH <sub>3</sub>	۹/۰۵
۲۶	CF <sub>3</sub>	H	OCH <sub>3</sub>	OCH <sub>3</sub>	۹/۶۱
۲۷	Br	H	NHCH <sub>3</sub>	H	۸/۳۹
۲۸	Br	H	N(CH <sub>3</sub> ) <sub>2</sub>	H	۷/۰۷
۲۹	Br	H	NHCOOCH <sub>3</sub>	H	۷/۹۲
۳۰	Br	H	H	OH	۸/۳۲
۳۱	Br	H	H	NHCOCH <sub>3</sub>	۷/۳۹
۳۲	Br	H	H	NHCH <sub>3</sub>	۸/۱۵
۳۳	Br	H	H	NHC <sub>2</sub> H <sub>5</sub>	۷/۹۲
۳۴	Br	H	H	N(CH <sub>3</sub> ) <sub>2</sub>	۷/۹۵
۳۵	Br	H	NH <sub>2</sub>	NHCH <sub>3</sub>	۹/۱۶
۳۶	Br	H	NH <sub>2</sub>	N(CH <sub>3</sub> ) <sub>2</sub>	۶/۷۹
۳۷	Br	H	NH <sub>2</sub>	OCH <sub>3</sub>	۸/۴۲
۳۸	Br	H	NH <sub>2</sub>	Cl	۸/۱۸
۳۹	Br	H	NO <sub>2</sub>	NHCH <sub>3</sub>	۷/۱۶
۴۰ <sup>t</sup>	Br	H	NO <sub>2</sub>	OCH <sub>3</sub>	۷/۸۲
۴۱	Br	H	OC <sub>2</sub> H <sub>5</sub>	OC <sub>2</sub> H <sub>5</sub>	۱۱/۲۲
۴۲ <sup>t</sup>	Br	H	O(CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub> )	O(CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub> )	۹/۷۶
۴۳	H	Br	OCH <sub>3</sub>	OCH <sub>3</sub>	۹/۰۱

\* نماد t بیانگر داده های سری آزمون بوده و سایر داده ها در سری آموزش قرار دارند.

#### ۲-۱-۴ رسم و بهینه سازی ساختار ترکیبات مورد مطالعه

ساختار مولکول‌های مورد مطالعه، مطابق با بخش (۲-۳) رسم و بهینه سازی شدند.

#### ۳-۱-۴ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه با استفاده از توصیف کننده‌های حاصل

##### از داکینگ مولکولی

در این تحقیق توصیف کننده‌های داکینگ مولکولی برای ترکیبات مجموعه داده‌ها استخراج شده و پس از انتخاب مهم‌ترین توصیف کننده‌ها، از آن‌ها به‌عنوان ورودی شبکه عصبی استفاده شد. پس از ساخت مدل شبکه عصبی مصنوعی، کارایی آن با استفاده از روش‌های مختلف ارزیابی شد.

#### ۱-۳-۱-۴ استخراج توصیف کننده‌های حاصل از داکینگ مولکولی

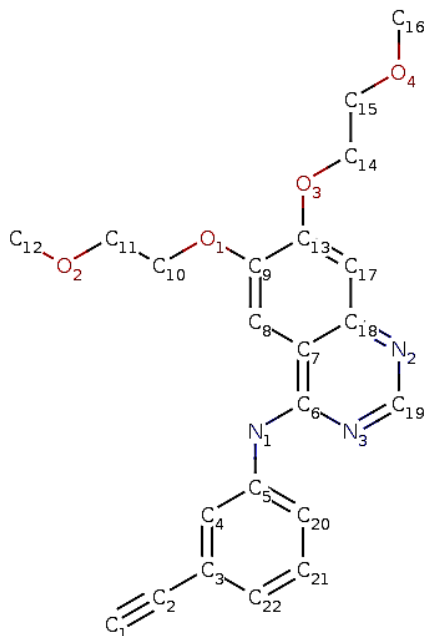
برای ایجاد ارتباط بین ساختار-فعالیت، ۱۰ توصیف کننده از نوع برهمکنش لیگاند-گیرنده با استفاده از داکینگ لیگاندهای مورد مطالعه در ساختار گیرنده فاکتور رشد اپیدرمی تیروزین کیناز 1M17 در شرایط بهینه اعتبارسنجی استخراج گردید. در بخش‌های زیر، آماده سازی لیگاندها (لیگاند کریستالوگرافی و ترکیبات مورد مطالعه)، آماده سازی پروتئین، تعیین بهترین شرایط داکینگ به روش اعتبارسنجی و در نهایت داکینگ لیگاندها در ساختار پروتئین به منظور استخراج توصیف کننده‌های داکینگ مولکولی مورد بحث واقع شده است.

#### - استخراج و آماده سازی لیگاند موجود در پروتئین

برای یافتن شرایط بهینه داکینگ مولکولی، فرایند اعتبارسنجی صورت گرفت. به این منظور، باتوجه به گزارش موجود [۹۷] ساختار کریستالوگرافی پروتئین 1M17، از بانک اطلاعاتی پروتئین [۶۲] دریافت شد و با فرمت بانک اطلاعاتی پروتئین ذخیره گردید. در ادامه استخراج و آماده سازی لیگاند کریستالوگرافی برای فرایند اعتبارسنجی مطابق با بخش (۳-۳-۱) صورت گرفت. ساختار لیگاند کریستالوگرافی موجود در جایگاه فعال در شکل (۴-۲) نمایش داده شده است. مختصات مرکز ثقل لیگاند کریستالوگرافی پروتئین ( $x=۲۲/۰۳۷۰$ ،  $y=-۰/۳۰۹۰$ ،  $z=۵۳/۵۷۲۰$ ) که با نرم افزار

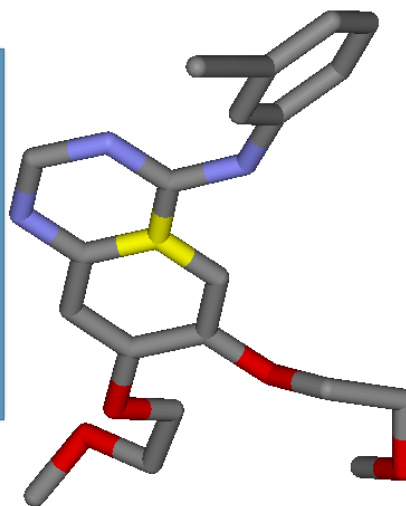
ViewerLite5.0 به دست آمده و در شکل (۳-۴) مشاهده می شود. اطلاعات استخراج شده، در مراحل

بعدی داکینگ مولکولی استفاده گردید.



شکل (۲-۴) ساختار لیگاند موجود در پروتئین 1m17

PROPERTY	VALUE
Name	C7
ID	2520
Color	144 144 144
Parent	A04999
XYZ	22.0370 -0.3090 53.5720
Element	Carbon
Hybridization	Sp 2
PartialCharge	0.040
FormalCharge	
Chirality	
Isotropic Displacement	52.23



شکل (۳-۴) مختصات جایگاه فعال به دست آمده از جایگاه لیگاند در پروتئین 1M17

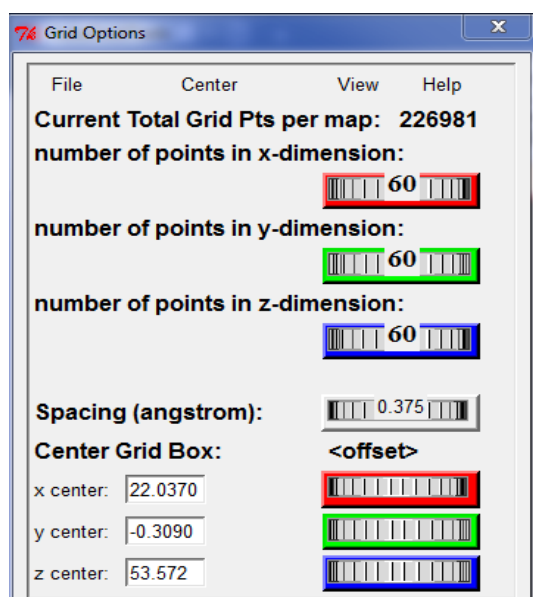
- آماده سازی پروتئین

برای آماده سازی پروتئین، فایل دانلود شده ساختار کریستالوگرافی 1M17 با ارزش تفکیک  $A^\circ$

۲/۶ در نرم افزار ViewerLite5.0 فراخوانی شد و سپس مطابق با بخش (۳-۳-۲) آماده و با فرمت pdb ذخیره شد.

#### - فرایند اعتبارسنجی (داک-ریداک)

به منظور انجام فرایند اعتبارسنجی، لیگاند کریستالوگرافی و پروتئین 1M17 در نرم افزار Autodock4.2 فراخوانی شدند. سپس مطابق بخش (۳-۳-۱) آماده سازی بیشتر بر روی پروتئین انجام شد. در ادامه، ابعاد و مختصات مرکز جایگاه فعال شبکه مطابق بخش (۳-۳-۲) تعیین شد. شکل (۴-۴) روند تعریف ابعاد شبکه و مختصات جایگاه فعال را نشان می دهد. به منظور اجرای عملیات داکینگ، الگوریتم ژنتیک لامارکین انتخاب گردید. تعداد اجراهای الگوریتم ژنتیک در مقادیر ۱۰۰، ۱۵۰ و ۲۰۰ و سایر پارامترهای الگوریتم ژنتیک و پارامترهای داکینگ در مقادیر پیش فرض الگوریتم Autodock4.2 قرار گرفتند. پس از تنظیم همه پارامترها، داکینگ لیگاند کریستالوگرافی در جایگاه فعال پروتئین 1M17 اجرا شد. خلاصه نتایج هر سه اجرا در جدول (۴-۲) آمده است. با توجه به این نتایج و موارد ذکر شده در بخش (۳-۳-۳)، داکینگ مولکولی با تعداد اجرای ژنتیک الگوریتم ۱۵۰ دارای RMSD معتبر (ارزش تفکیک  $>2A^\circ$ )، کمترین تعداد خوشه، بیشترین تعداد کنفورماسیون در خوشه برتر [۹۲] و زمان اجرای قابل قبول است. بنابراین تعداد اجرای ژنتیک الگوریتم ۱۵۰ به عنوان تعداد اجرای بهینه برای اجرای داکینگ مولکولی همه مولکولها در مجموعه دادهها مورد استفاده قرار گرفت.



شکل (۴-۴) تعیین مختصات شبکه در بخش اتوگرید نرم افزار

جدول (۲-۴) نتایج اجراهای الگوریتم ژنتیک متفاوت برای پروتئین با شناسه کریستالوگرافی 1M17

تعداد اجرا (دقیقه)	تعداد کنفورماسیون‌های موجود در خوشه اول	تعداد خوشه‌های چند کنفورماسیونه	میانگین انرژی اتصال (kcal/mol)	RMSD (Å)	تعداد اجرا الگوریتم ژنتیک
۴۴	۴۴	۲۷	-۷/۴۸	۰/۵۷	۱۰۰
۶۴	۱۰۲	۱۳	-۷/۵۱	۰/۵۶	۱۵۰
۱۴۵	۸۱	۳۲	-۷/۳۲	۰/۸۷	۲۰۰

#### - داکینگ لیگاندهای مورد مطالعه در پروتئین مربوطه

بعد از آماده سازی لیگاندهای مورد مطالعه و پروتئین کریستالوگرافی، همه لیگاندها در شرایط بهینه فرایند اعتبارسنجی داک شدند و ۱۰ توصیف کننده مولکولی حاصل از برهمکنش لیگاند-پروتئین استخراج گردید. سپس از بین این ۱۰ توصیف کننده، موثرترین توصیف کننده‌ها انتخاب شدند و از آنها برای ساخت مدل‌های QSAR استفاده شد.

#### ❖ آماده سازی لیگاندهای مورد مطالعه

برای آماده سازی لیگاندهای مورد مطالعه در عملیات داکینگ مولکولی، همه لیگاندها مطابق بخش (۲-۳) رسم و بهینه سازی و با فرمت \*.mol ذخیره شدند. در ادامه، ترکیبات بهینه شده پس از فراخوانی در نرم افزار ViewerLite5.0 با پسوند pdb ذخیره شدند.



## ❖ اجرای عملیات داکینگ مولکولی لیگاندها و پروتئین

به منظور استخراج توصیف کننده‌های داکینگ مولکولی، پروتئین 1M17 در نرم افزار Autodock4.2 فراخوانی شد. آماده سازی بیشتر پروتئین در نرم افزار Autodock4.2 مطابق بخش (۳-۳-۱) انجام شد و با فرمت pdb ذخیره گردید. مختصات جایگاه فعال و ابعاد شبکه برای پروتئین مطابق بخش (۳-۳-۲) تعیین شد. در ادامه، تک تک لیگاندها به‌طور جداگانه در نرم افزار Autodock4.2 فراخوانی شدند. برای انجام عملیات داکینگ لیگاندها در جایگاه فعال پروتئین از الگوریتم ژنتیک لامارکین با شرایط بهینه انتخاب شده در جدول (۴-۲) استفاده شد. پس از اجرای عملیات داکینگ نتایج خروجی با فرمت dlz ذخیره شد. در نهایت ۱۰ توصیف کننده‌ی مولکولی حاصل از برهم کنش لیگاند-پروتئین از خوشه‌ی برتر که شامل بیشترین تعداد صورت‌بندی و انرژی اتصال منفی‌تر می‌باشد استخراج گردید.

## - استخراج توصیف کننده‌های حاصل از داکینگ مولکولی

پس از داکینگ لیگاندها در پروتئین مربوطه، برهمکنش‌های بین لیگاندها با پروتئین آنالیز شد و به منظور برقراری ارتباط ریاضی بین ساختار-فعالیت، ۱۰ توصیف کننده داکینگ مولکولی از نوع انرژی با استفاده از نرم افزار Notepad++ به صورت عددی استخراج شدند. نام این توصیف کننده‌ها به همراه علائم اختصاری آن‌ها در بخش (۲-۵-۷) ذکر شده است. از آنجایی که این توصیف کننده‌ها حاوی اطلاعات مفیدی از برهمکنش لیگاندها-پروتئین هستند، بسیار مورد توجه قرار می‌گیرند. همه توصیف کننده‌ها برای ۴۳ لیگاند مورد مطالعه محاسبه و سپس به‌عنوان ورودی در مدل‌سازی‌های QSAR مورد استفاده قرار گرفت.

## ۴-۱-۲ انتخاب موثرترین توصیف کننده‌های داکینگ مولکولی

با توجه به این که تعداد زیاد توصیف کننده‌ها باعث پیچیدگی محاسبات می‌شود، باید تعداد توصیف کننده‌ها کاهش یابد. به همین منظور، از بین دو توصیف کننده با همبستگی بزرگتر از ۰/۹،

توصیف کننده‌ای که همبستگی کمتری با متغیر پاسخ داشت حذف گردید. به منظور انتخاب موثرترین توصیف کننده‌ها مطابق بخش (۳-۴)، از نرم افزار SPSS استفاده شد و پس از اجرای روش رگرسیون گام به گام ۲ توصیف کننده به‌عنوان موثرترین توصیف کننده‌ها انتخاب شدند. نام و نوع این دو توصیف کننده در جدول (۳-۴) نشان داده شده است و جدول (۴-۴) نیز عدم وجود همبستگی بین این دو توصیف کننده را نشان می‌دهد. با توجه به جدول (۵-۴) مقادیر VIF کمتر از ۱۰ نیز، عدم وجود هم‌خطی بین دو توصیف کننده را نشان می‌دهد.

جدول (۳-۴) توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

No	Symbol	Meaning
۱	VE	(vdW + Hbond + desolv) Energy
۲	U	Internal energy

جدول (۴-۴) ماتریس همبستگی توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

	VE	U
VE	۱	
U	۰/۲۲	۱

جدول (۵-۴) مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

توصیف کننده	VE	U
VIF	۱/۰۵	۱/۰۵

#### ۳-۳-۱-۴ مدل شبکه عصبی مصنوعی

شبکه عصبی انتخاب شده در این مطالعه یک شبکه پیشخور سه لایه با الگوریتم آموزشی پس انتشار است. ۲ توصیف کننده‌ی داکینگ مولکولی انتخاب شده به‌عنوان توصیف کننده ورودی و فعالیت متناظر آن‌ها به‌عنوان توصیف کننده هدف در نظر گرفته شدند تا پاسخ شبکه با آن سنجیده شود. به دلیل تعداد کم مجموعه داده‌ها و به منظور جلوگیری از بیش برآزش شبکه، از تابع آموزش تنظیم بیزی استفاده شد. مجموعه داده‌ها به‌طور تصادفی به مجموعه آزمون (۸ ترکیب) و مجموعه آموزش (۳۵ ترکیب) تقسیم شدند. سری آموزش برای محاسبه گرادیان و به روز کردن وزن‌ها و بایاس‌های شبکه

استفاده شد. برای یافتن شرایط بهینه شبکه عصبی، پارامترهای مؤثر به صورت همزمان بهینه شدند. هر شبکه با تعداد ورودی ۲ و با تابع آموزش تنظیم بیزی و دو تابع انتقال لگاریتم سیگموئید و تانژانت سیگموئید و تعداد گره از ۲ تا ۱۰ با گام ۱ و تعداد دوره‌های آموزش ۲ تا ۵۰ با گام ۲ آموزش داده شد. همچنین از تابع انتقال خطی در لایه خروجی استفاده گردید. در روند بهینه سازی پارامترهای فوق، کمینه سازی میانگین مربع خطای داده‌های سری آموزش، به‌عنوان معیار انتخاب شد. نتایج بهینه سازی این پارامترها در جدول (۴-۶) گردآوری شده است. با توجه به این نتایج، شبکه با ۲ ورودی در لایه اول، ۵ گره در لایه میانی، ۴۲ دور آموزش، تابع انتقال تانژانت سیگموئیدی و تابع آموزش تنظیم بیزی با توجه به سادگی بیشتر به‌عنوان شبکه برتر انتخاب شد.

جدول (۴-۶) توابع و پارامترهای شبکه‌های بهینه شده با توصیف کننده‌های انتخاب شده داکینگ مولکولی

MSE	تعداد دور آموزش	تعداد گره	تابع انتقال	تابع آموزش	تعداد توصیف کننده
۰/۶۴	۱۰	۹	لگاریتم-سیگموئید	تنظیم بیزی	۲
۰/۶۴	۴۲	۵	تانژانت-سیگموئید	تنظیم بیزی	۲

#### ۴-۱-۳-۴ ارزیابی مدل شبکه عصبی مصنوعی

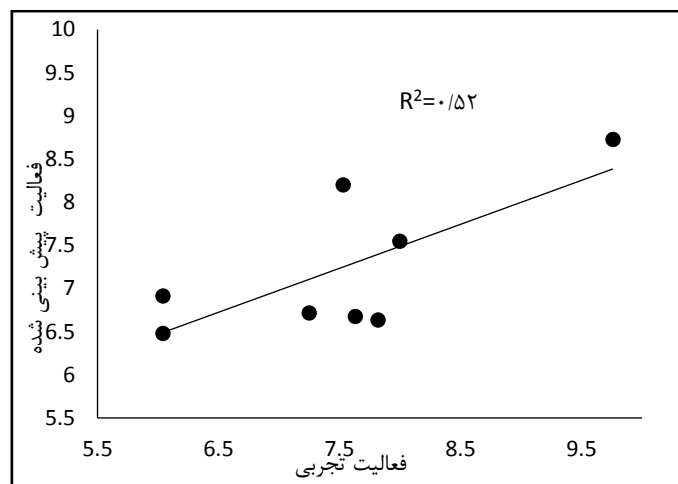
اساسی‌ترین مرحله در مطالعات رابطه کمی ساختار-فعالیت، ارزیابی قدرت پیش‌بینی مدل است که با بررسی توانایی آن در ارائه پیش‌بینی‌های رضایت‌بخش برای مولکول‌هایی که در سری آموزش وجود نداشته‌اند، تعریف می‌شود. در این تحقیق ارزیابی مدل با استفاده از سری آزمون مورد بررسی قرار گرفت و فعالیت ضد سرطان ترکیبات مورد مطالعه پیش‌بینی شد.

بعد از انتخاب شبکه عصبی بهینه، قدرت پیش‌بینی مدل توسعه یافته، به وسیله ۸ ترکیب سری آزمون که در آموزش شبکه استفاده نشده‌اند، مورد بررسی قرار گرفت. نتایج حاصله در جدول (۴-۷) خلاصه شد. نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون در شکل (۴-۵) نشان داده شده است. ضریب تعیین ( $R^2$ ) برای سری آزمون ۰/۵۲ به‌دست آمده که از مقدار قابل قبول ارائه شده [۷۷] (۰/۶) کوچک‌تر است، می‌توان نتیجه گرفت که مدل شبکه عصبی

مصنوعی ساخته شده با توصیف کننده‌های داکینگ مولکولی قدرت پیش‌بینی خوبی ندارد.

جدول (۷-۴) نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از داده‌های سری آزمون

شماره ترکیب	pIC <sub>50</sub>		درصد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۲	۶/۰۴	۶/۹۱	۱۴/۴۰
۳	۷/۶۳	۶/۶۸	-۱۲/۴۵
۷	۶/۰۴	۶/۴۸	۷/۲۸
۱۰	۸/۰۰	۷/۵۵	-۵/۶۳
۱۳	۷/۲۵	۶/۷۲	-۷/۳۱
۲۲	۷/۵۳	۸/۲۰	۸/۹۰
۴۰	۷/۸۲	۶/۶۴	-۱۵/۰۹
۴۲	۹/۷۶	۸/۷۳	-۱۰/۵۵



شکل (۷-۴) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون

#### ۴-۱-۴ مدل سازی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

در راستای بهبود مدل شبکه عصبی مصنوعی، با استفاده از ساختار بهینه شده لیگاندها و بهره‌گیری از نرم افزار دراگون، ۱۵۰ توصیف کننده‌ی تابع توزیع شعاعی (RDF) استخراج گردید. مدل شبکه عصبی مصنوعی با استفاده از موثرترین توصیف کننده‌ها از مجموعه توصیف کننده‌های داکینگ مولکولی و توصیف کننده‌های RDF (۱۶۰ توصیف کننده) توسعه یافت. نتایج شبکه، بهبود مدل با توصیف کننده‌های منتخب این دو دسته توصیف کننده را نشان می‌دهد.

#### ۱-۴-۱-۴ توصیف کننده‌های تابع توزیع شعاعی

توصیف کننده‌های RDF مجموعه‌ای از کدهای ساختاری هستند که مختصات سه بعدی اتم‌های یک مولکول را بدون در نظر گرفتن اندازه مولکول پوشش می‌دهند. در حقیقت، برای یک مولکول با  $N$  اتم، RDF نشان دهنده توزیع احتمالی برای یافتن اتم  $N$  ام در یک کره با شعاع  $r$  است. فرم کلی این تابع به صورت رابطه زیر است:

$$g(r) = f \sum_i^{N-1} \sum_j^N A_i \cdot A_j \cdot e^{-\beta \cdot (r-r_{ij})^2} \quad \text{رابطه (۱-۴)}$$

که در این رابطه  $N$  تعداد اتم‌ها،  $A_i$  و  $A_j$  خاصیت اتمی (الکترونگاتیویته، جرم اتمی، قطبش‌پذیری اتمی و غیره)،  $r_{ij}$  شعاع بین اتم  $i$  و  $j$ ،  $f$  فاکتور مقیاس و  $\beta$  فاکتور دما است که باعث جابه‌جایی و حرکت یون‌ها می‌باشد. این توصیف کننده‌ها مستقل از تعداد اتم‌ها (اندازه مولکولی)، ترتیب سه بعدی اتم‌ها و انتقال یا چرخش مولکول [۱۹، ۲۰] می‌باشند و حاوی اطلاعات پایه‌ای در مورد فاصله بین اتمی مولکول، طول پیوند، نوع حلقه و سیستم‌های صفحه‌ای یا غیر صفحه‌ای هستند [۹۸]. اهمیت این توصیف کننده‌ها به دلیل اختلاف در توزیع اتم‌ها در مولکول و اثر این توزیع بر روی فعالیت بازدارندگی ترکیبات می‌باشد. در این تحقیق استخراج توصیف کننده‌های RDF که بیشتر از توصیف کننده‌های داکینگ مولکولی بیانگر ویژگی‌های شیمیایی ساختار می‌باشند، مطابق بخش (۳-۳) انجام شد. در ادامه مدل شبکه عصبی مصنوعی با استفاده از ترکیب توصیف کننده‌های داکینگ مولکولی و توصیف کننده‌های RDF توسعه یافت و قدرت پیش‌بینی شبکه در فعالیت بازدارندگی ترکیبات مورد بررسی مورد مطالعه قرار گرفت.

#### ۱-۴-۲-۱-۴ انتخاب موثرترین توصیف کننده‌ها از مجموع توصیف کننده‌های داکینگ مولکولی و

##### ساختاری

انتخاب توصیف کننده‌هایی با بیشترین ارتباط بین فعالیت دارویی - ساختار مولکولی، در مدل سازی بسیار حائز اهمیت است. بنابراین در انتخاب توصیف کننده‌ها باید دقت کرد و توصیف کننده‌هایی

می‌بایست انتخاب شوند که در ارتباط با ویژگی مورد نظر مولکول بوده، فهم و تفسیر آن‌ها آسان باشد، هم چنین تغییرات کوچکی در ساختار مولکول، بایستی منجر به تغییرات محسوسی در مقادیر توصیف کننده شود. به این منظور برای انتخاب توصیف کننده‌های مناسب با بیشترین ارتباط با مقادیر فعالیت بازدارندگی، مطابق با بخش (۳-۴) عمل شد و از میان ۱۶۰ توصیف کننده، ۸ توصیف کننده به‌عنوان موثرترین توصیف کننده‌ها انتخاب شدند. نام و نوع توصیف کننده‌های انتخاب شده در جدول (۴-۸) خلاصه شده است.

جدول (۴-۸) توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی به روش SR

NO	Symbol	Meaning	Class
۱	RDF025u	/unweighted $2/5$ Radial Distribution Function-	RDF descriptors
۲	RDF095u	/unweighted $9/5$ Radial Distribution Function-	RDF descriptors
۳	RDF070e	/weighted by $7/0$ -Radial Distribution Function-atomic Sanderson electronegativities	RDF descriptors
۴	RDF065v	/weighted by $6/5$ Radial Distribution Function-atomic Van der Waals volumes	RDF descriptors
۵	RDF105v	/weighted by $10/5$ Radial Distribution Function-atomic Van der Waals volumes	RDF descriptors
۶	RDF115p	/weighted by $11/5$ Radial Distribution Function-atomic polarizabilities	RDF descriptors
۷	VE	VdW + Hbond + desolv Energy	Docking descriptors
۸	U	Internal energy	Docking descriptors

جدول (۴-۹) ماتریس همبستگی این توصیف کننده‌ها را نشان می‌دهد. مقادیر موجود در جدول عدم همبستگی قابل ملاحظه بین توصیف کننده‌ها را نشان می‌دهد. به منظور بررسی وجود یا عدم وجود هم‌خطی در بین توصیف کننده‌های انتخاب شده، مقدار VIF برای توصیف کننده‌ها محاسبه شد (جدول (۴-۱۰)). با توجه به اینکه مقدار این کمیت برای همه توصیف کننده‌ها از ۱۰ کوچکتر است، عدم وجود هم‌خطی در بین توصیف کننده‌ها تایید می‌شود.

جدول ( ۴-۹ ) ماتریس همبستگی توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR

	VE	U	RDF115p	RDF025u	RDF070e	RDF105v	RDF095u	RDF065v
VE	۱							
U	۰/۲۲	۱						
RDF115p	-۰/۴۹	-۰/۲۰	۱					
RDF025u	-۰/۶۲	-۰/۴۵	۰/۵۶	۱				
RDF070e	-۰/۵۶	-۰/۲۳	۰/۵۴	۰/۷۶	۱			
RDF105v	۰/۵۷	-۰/۱۱	۰/۷۰	۰/۷۳	۰/۵۵	۱		
RDF095u	-۰/۴۵	۰/۱۱	۰/۶۵	۰/۶۴	۰/۶۳	۰/۶۱	۱	
RDF065v	-۰/۶۲	-۰/۲۱	۰/۷۱	۰/۸۲	۰/۷۲	۰/۸۰	۰/۷۶	۱

جدول ( ۴-۱۰ ) مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با روش SR

توصیف کننده	VE	U	RDF115p	RDF025u	RDF070e	RDF105v	RDF095u	RDF065v
VIF	۱/۸۳	۱/۶۴	۲/۵۸	۴/۷۷	۳/۱۳	۳/۵۱	۳/۶۱	۶/۹۱

#### ۴-۱-۳ مدل سازی شبکه عصبی مصنوعی با استفاده از مجموع توصیف کننده‌های داکینگ

##### مولکولی و ساختاری

شبکه‌ی انتخاب شده یک شبکه عصبی پیشخور با الگوی پس انتشار خطا و الگوریتم آموزشی تنظیم بیزی می‌باشد. به منظور بهینه‌سازی و شناسایی بهترین پارامترهای شبکه (تعداد توصیف کننده‌های ورودی شبکه، تعداد گره در لایه پنهان، تعداد دوره‌های آموزشی و تابع انتقال)، مجموعه داده‌ها به‌طور تصادفی به مجموعه آزمون (۸ ترکیب) و مجموعه آموزش (۳۵ ترکیب) تقسیم شدند. برای یافتن مقادیر بهینه پارامترهای مؤثر، هر شبکه با ورودی‌های از ۲ تا ۸ با گام ۱، تعداد گره‌های ۲ تا ۱۰ با گام ۱ در لایه پنهان و تعداد دوره‌های آموزشی از ۲ تا ۵۰ با گام ۲ و دو تابع انتقال لگاریتم سیگموئید و تانژانت سیگموئید آموزش داده شد. ملاک بهینه سازی کمینه سازی میانگین مربع خطاهای حاصل از داده‌های سری آموزش در نظر گرفته شد. با توجه به نتایج جدول ( ۴-۱۱ ) مدل با ۸ توصیف کننده ورودی و ۲ گره در لایه پنهان و ۳۸ دور آموزشی و تابع انتقال تانژانت سیگموئید از سادگی و قدرت پیش‌بینی مناسبی برخوردار است.

جدول ( ۴-۱۱) توابع و پارامترهای شبکه‌های بهینه با توصیف کننده‌های منتخب از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR

MSE	تعداد دور آموزش	تعداد گره	تابع انتقال	تابع آموزش	تعداد توصیف کننده
۰/۱۸	۵۰	۶	لگاریتم - سیگموئیدی	تنظیم بیزی	۸
۰/۱۷	۳۸	۲	تانژانت - سیگموئیدی	تنظیم بیزی	۸

#### ۴-۱-۴-۱-۴ ارزیابی مدل شبکه عصبی مصنوعی

در ارزیابی مدل ارائه شده در این تحقیق برای پیش‌بینی داده‌های pIC<sub>50</sub> بازدارنده‌های تیروزین

کیناز به عنوان فاکتور رشد اپیدرمی، روش‌های زیر به کار گرفته شد:

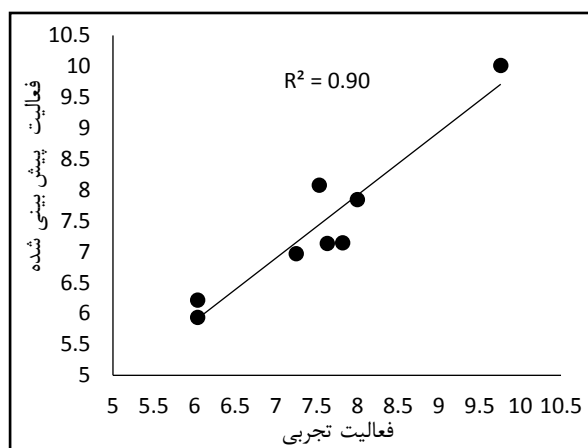
#### - ارزیابی مدل با استفاده از سری آزمون

بعد از انتخاب شبکه عصبی بهینه، قدرت پیش‌بینی شبکه به وسیله ۸ داده‌ی سری آزمون که در آموزش استفاده نشده‌اند، مورد بررسی قرار گرفت. نتایج حاصله در جدول ( ۴-۱۲) نشان داده شده است. درصد خطای پایین پیش‌بینی حاکی از قدرت پیش‌بینی بالای مدل توسعه یافته است. همچنین شکل ( ۴-۶) نشان دهنده تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون است. پراکندگی نتایج حول محور صفر، عدم وجود خطای سیستماتیک مدل توسعه یافته را اثبات می‌کند.



جدول (۴-۱۲) نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از داده‌های سری آزمون

شماره ترکیب	pIC <sub>50</sub>		درصد خطای نسبی
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۲	۶/۰۴	۶/۲۲	۲/۹۸
۳	۷/۶۳	۷/۱۳	-۶/۵۵
۷	۶/۰۴	۵/۹۴	-۱/۶۶
۱۰	۸/۰۰	۷/۸۴	-۲/۰۰
۱۳	۷/۲۵	۶/۹۷	-۳/۸۶
۲۲	۷/۵۳	۸/۰۸	۷/۳۰
۴۰	۷/۸۲	۷/۱۴	-۸/۷۰
۴۲	۹/۷۶	۱۰/۰۱	۲/۵۶



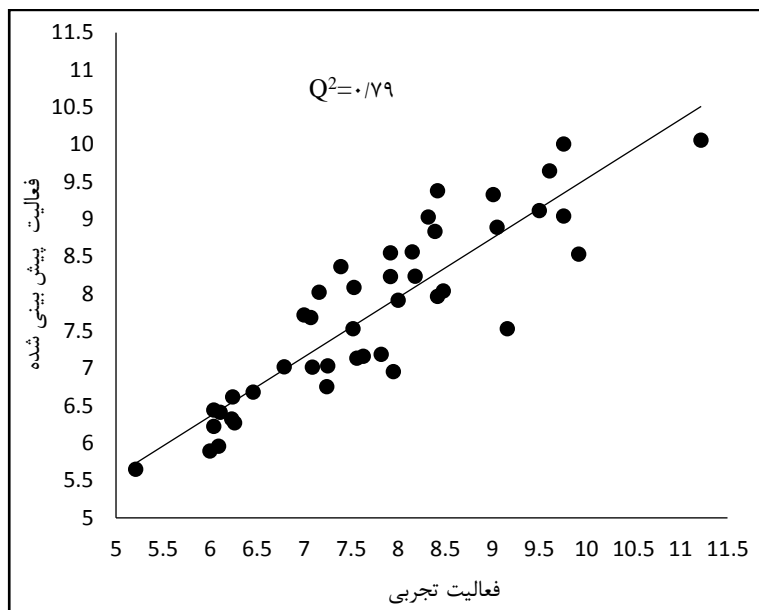
شکل (۴-۶) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون

#### - ارزیابی مدل به روش رد مرحله‌ای تک تک

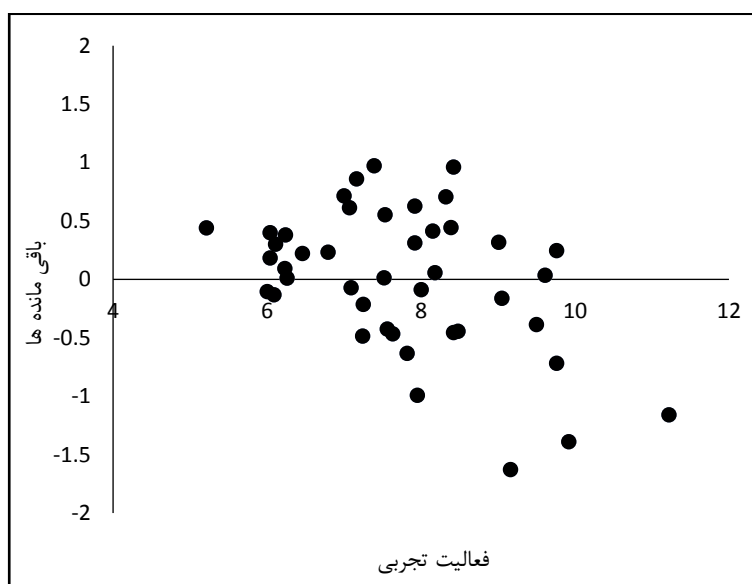
تکنیک رد مرحله‌ای تک تک (برای کل داده‌ها) روش دیگری برای ارزیابی بیشتر مدل برتر شبکه عصبی است که نتایج حاصل از این روش در جدول (۴-۱۳) ارائه شده است. این نتایج تعمیم‌پذیری بالای مدل توسعه یافته شبکه عصبی را بیان می‌کند. با توجه به شکل (۴-۷)، ضریب تعیین بین مقادیر پیش‌بینی شده (pIC<sub>50</sub>) و مقادیر تجربی بر نزدیکی مقادیر پیش‌بینی شده به مقادیر تجربی آن‌ها دلالت دارد. همچنین در شکل (۴-۸) نمودار باقی‌مانده بر حسب فعالیت تجربی ترکیبات مورد بحث، رسم شده است که عدم وجود خطای سیستماتیک را نشان می‌دهد. تمرکز نقاط به سمت محور افقی (مقدار باقی‌مانده صفر) نشان دهنده‌ی نزدیک بودن مقادیر پیش‌بینی شده به مقادیر تجربی pIC<sub>50</sub> است.

جدول (۴-۱۳) نتایج حاصل از ارزیابی مدل شبکه عصبی به روش رد مرحله‌ای تک تک برای کل داده‌ها

شماره ترکیب	PIC <sub>50</sub>		درصد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۱	۶/۴۶	۶/۶۸	۳/۴۱
۲	۶/۰۴	۶/۲۲	۲/۹۸
۳	۷/۶۳	۷/۱۶	-۶/۱۶
۴	۷/۵۶	۷/۱۴	-۵/۵۶
۵	۷/۰۹	۷/۰۲	-۰/۹۹
۶	۶/۲۳	۶/۳۲	۱/۴۴
۷	۶/۰۴	۶/۴۴	۶/۶۲
۸	۷/۵۲	۷/۵۳	۰/۱۳
۹	۶/۰۰	۵/۹۰	-۱/۶۷
۱۰	۸/۰۰	۷/۹۱	-۱/۱۳
۱۱	۹/۷۶	۹/۰۴	-۷/۳۸
۱۲	۹/۹۲	۸/۵۳	-۱۴/۰۱
۱۳	۷/۲۵	۷/۰۴	-۲/۹۰
۱۴	۷/۲۴	۶/۷۵	-۶/۷۷
۱۵	۶/۱۱	۶/۴۱	۴/۹۱
۱۶	۶/۲۴	۶/۶۲	۶/۰۹
۱۷	۷/۰۰	۷/۷۲	۱۰/۲۹
۱۸	۸/۴۸	۸/۰۴	-۵/۱۹
۱۹	۵/۲۱	۵/۶۵	۸/۴۵
۲۰	۶/۰۹	۵/۹۶	-۲/۱۳
۲۱	۶/۲۶	۶/۲۷	۰/۱۶
۲۲	۷/۵۳	۸/۰۸	۷/۳۰
۲۳	۸/۴۲	۷/۹۷	-۵/۳۴
۲۴	۹/۵	۹/۱۱	-۴/۱۱
۲۵	۹/۰۵	۸/۸۹	-۱/۷۷
۲۶	۹/۶۱	۹/۶۵	۰/۴۲
۲۷	۸/۳۹	۸/۸۳	۵/۲۴
۲۸	۷/۰۷	۷/۶۸	۸/۶۳
۲۹	۷/۹۲	۸/۵۵	۷/۹۵
۳۰	۸/۳۲	۹/۰۳	۸/۵۳
۳۱	۷/۳۹	۸/۳۶	۱۳/۱۳
۳۲	۸/۱۵	۸/۵۶	۵/۰۳
۳۳	۷/۹۲	۸/۲۳	۳/۹۱
۳۴	۷/۹۵	۶/۹۶	-۱۲/۴۵
۳۵	۹/۱۶	۷/۵۳	-۱۷/۷۹
۳۶	۶/۷۹	۷/۰۲	۳/۳۹
۳۷	۸/۴۲	۹/۳۸	۱۱/۴۰
۳۸	۸/۱۸	۸/۲۴	۰/۷۳
۳۹	۷/۱۶	۸/۰۲	۱۲/۰۱
۴۰	۷/۸۲	۷/۱۹	-۸/۰۶
۴۱	۱۱/۲۲	۱۰/۰۶	-۱۰/۳۴
۴۲	۹/۷۶	۱۰/۰۱	۲/۵۶
۴۳	۹/۰۱	۹/۳۳	۳/۵۵



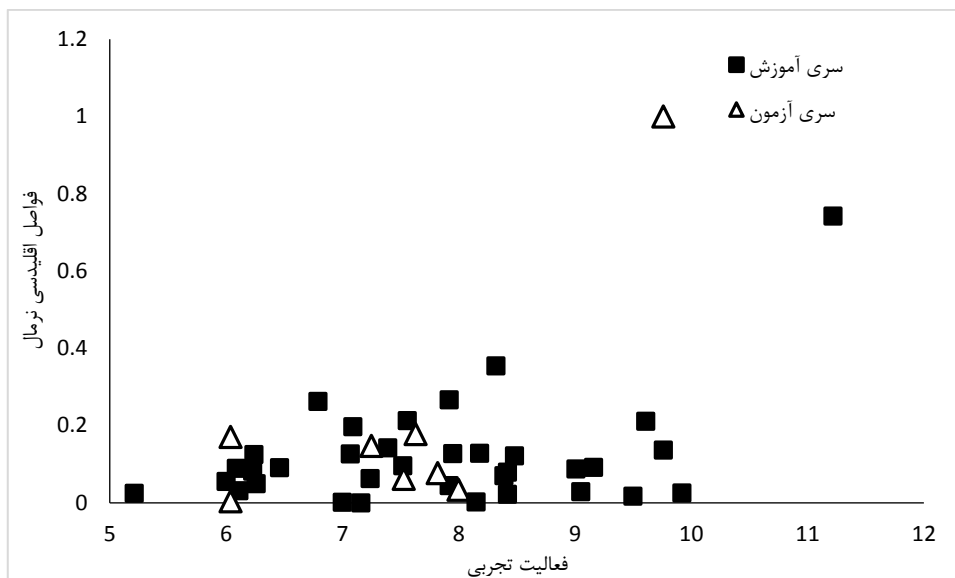
شکل ( ۷-۴ ) نمودار تغییرات فعالیت پیش بینی شده بر حسب فعالیت تجربی به روش رد مرحله ای تک تک



شکل ( ۸-۴ ) نمودار باقی مانده بر حسب فعالیت تجربی برای کل داده ها

### - ارزیابی مدل با استفاده از آزمون پراکندگی

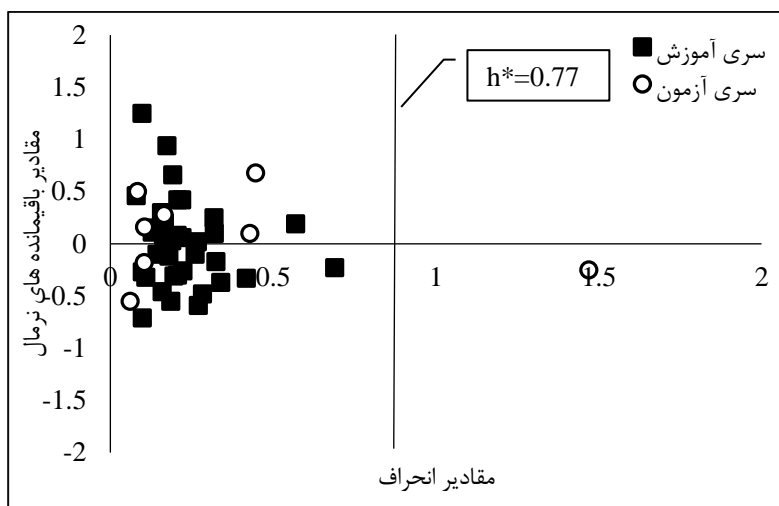
به منظور بررسی صحت تقسیم بندی داده های سری آموزش و آزمون نمودار آزمون پراکندگی از رسم مقادیر فواصل اقلیدسی بر حسب فعالیت های تجربی به دست آمد. با توجه به شکل ( ۹-۴ ) مشاهده می شود که سری آزمون نماینده مناسبی از سری آموزش می باشد.



شکل (۹-۴) آزمون پراکندگی برای تقسیم بندی داده‌های ۴- آنیلینو کوئینازولین

#### - ارزیابی مدل با استفاده از قلمرو کاربرد

یک مرحله بسیار مهم در توسعه مدل QSAR مشخص کردن قلمرو کاربرد (AD) مدل‌های منتخب است. همانطور که قبلاً توضیح داده شد، قلمرو کاربرد با استفاده از باقیمانده استاندارد شده فعالیت پیش‌بینی شده و مقدار حد انحراف مشخص شد. مقادیر حد انحراف برای ترکیبات سری آموزش و پیش‌بینی محاسبه شد. قلمرو کاربرد مدل ارائه شده در شکل (۴-۱۰) نشان می‌دهد که ۹۸٪ ترکیبات سری آموزش و آزمون در دامنه کاربرد مدل توسعه یافته قرار دارند و تنها ترکیب شماره ۴۲ از سری ترکیبات آزمون دارای حد انحراف بزرگتر از  $h^*$  می‌باشد. علاوه بر این، هیچ یک از ترکیبات دارای باقیمانده استاندارد شده بزرگتر از ۳ واحد انحراف استاندارد نیستند. بنابراین هیچ پاسخی در افتاده‌ای در این محدوده مشاهده نشد و این مدل را می‌توان با اطمینان بالایی در این قلمرو کاربرد استفاده کرد.



شکل (۴-۱۰) نمودار قلمرو کاربرد برای مدل شبکه عصبی ترکیبات ۴- آنیلینو کوئینازولین

#### ۴-۱-۵ نتیجه گیری مدل سازی با مجموع توصیف کننده های داکینگ مولکولی و ساختاری

در این کار تحقیقاتی، ترکیبی از توصیف کننده های ساختاری RDF و داکینگ مولکولی برای ساخت یک مدل QSAR برای پیش بینی فعالیت ضد سرطان ترکیبات ۴ آنیلینو کوئینازولین استفاده شد. از طرفی به منظور بررسی تاثیر توصیف کننده ها، مدل شبکه عصبی با توصیف کننده های داکینگ مولکولی نیز توسعه یافت. پارامترهای ضریب تعیین و MSE برای پیش بینی فعالیت های دارویی ترکیبات مورد مطالعه در سری آزمون برای هر دو مدل محاسبه شد که در جدول (۴-۱۴) خلاصه شده است. این نتایج نشان می دهد که با اضافه کردن توصیف کننده های RDF به توصیف کننده های داکینگ مولکولی که حاوی اطلاعات ارزشمند در مورد تعامل ترکیبات فعال دارویی با پروتئین گیرنده است، قدرت پیش بینی مدل QSAR افزایش می یابد. در مجموع، استفاده ترکیبی از توصیف کننده های داکینگ مولکولی و RDF، یک مدل پیش بینی کننده کارآمد را برای پیش بینی فعالیت ضد سرطان مشتقات ۴- آنیلینو کوئینازولین به عنوان مهار کننده های قوی EGFR فراهم می کند.

جدول (۴-۱۴) مقایسه مدل های شبکه عصبی توسعه یافته با توصیف کننده های متفاوت در ارزیابی با استفاده از سری آزمون

پارامتر آماری	مدل توسعه یافته با توصیف کننده های	مدل توسعه یافته با مجموع توصیف کننده های
	داکینگ مولکولی	RDF و داکینگ مولکولی
R <sup>2</sup>	۰/۵۲	۰/۹۰
MSE	۰/۶۶	۰/۱۵

#### ۴-۱-۶ ارزیابی مدل با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با

#### استفاده از پارامترهای آماری

برای ارزیابی کارایی شبکه عصبی مصنوعی ارائه شده برای پیش‌بینی فعالیت بازدارندگی ترکیبات ۴-آنیلینو کوئینازولین، برخی از پارامترهای آماری محاسبه شد و نتایج حاصل در جدول (۴-۱۵) آمده است. با توجه به توضیحات پارامترهای آماری در بخش (۲-۷-۱)، نتایج حاکی از قدرت پیش‌بینی بالای مدل غیرخطی شبکه‌ی عصبی می‌باشد.

جدول (۴-۱۵) پارامترهای آماری برای مدل ANN با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

ردیف	پارامتر آماری	سری آموزش	سری آزمون	محدوده قابل قبول
۱	PRESS	۶/۱۱	۱/۲۲	-
۲	SEP	۰/۴۲	۰/۳۹	-
۳	MAE	۰/۳۲	۰/۳۴	-
۴	REP(%)	۵/۳۶	۵/۲۰	-
۵	MSE	۰/۱۷	۰/۱۵	-
۶	MRE	۴/۱۱	۴/۴۵	-
۷	$R^2$	۰/۹۰	۰/۹۰	$> 0.6$
۸	$R_0^2$	۰/۸۸	۰/۹۰	نزدیک به $R^2$
۹	$R_{0,R}^2$	۰/۰۲	۰/۰۰	$< 0.1$
۱۰	$R_m^2$	۰/۷۷	۰/۹۰	$> 0.5$
۱۱	$R_0'^2$	۰/۹۰	۰/۸۸	نزدیک به $R^2$
۱۲	$R_{0,R}'^2$	۰/۰۰	۰/۰۲	$< 0.1$
۱۳	$R_m'^2$	۰/۹	۰/۷۷	$> 0.5$
۱۴	$\Delta R_m^2$	۰/۱۳	۰/۱۳	$< 0.2$
۱۵	k	۱/۰۰	۰/۹۹	$0.15 \leq k \leq 0.85$
۱۶	k'	۱/۰۰	۱/۰۱	$0.15 \leq k' \leq 0.85$

#### ۷-۴-۱-۴ آزمون Y-تصادفی

برای اثبات قدرت مدل ایجاد شده و اینکه نتایج به دست آمده تصادفی نبوده است از آزمون Y- تصادفی استفاده می‌شود. بدین منظور  $pIC_{50}$  (فعالیت دارویی به عنوان متغیر پاسخ) به صورت تصادفی در محدوده داده‌های متغیر پاسخ ایجاد شد. شبکه عصبی بهینه شده با استفاده از توصیف کننده‌ها و مقادیر تصادفی از فعالیت دارویی ۴۳ ترکیب مورد مطالعه توسعه یافت. نتایج حاصل از چندین بار اجرای این آزمون در جدول (۴-۱۶) آورده شده که مقادیر ضریب تعیین کوچک، نشان‌دهنده‌ی عدم وابستگی توصیف کننده‌ها، در مدل توسعه یافته به وسیله شبکه عصبی مصنوعی است.

جدول (۴-۱۶) مقادیر  $R^2$  برای سری آزمون با استفاده از آزمون Y-تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
$R^2$	۰/۰۴	۰/۱۱	۰/۰۳	۰/۰۰	۰/۲۱	۰/۰۴	۰/۰۱	۰/۰۱	۰/۰۰	۰/۱۰

## ۲-۴ مدل سازی فعالیت دارویی مشتقات با مشابهت مولکولی دی آریل پیریمیدین‌ها<sup>۱</sup>

### (DAPY-like) به‌عنوان بازدارنده‌های غیرنوکلئوزیدی ویروس HIV

در این پژوهش به بررسی و مطالعه رابطه کمی ساختار-فعالیت برای پیش‌بینی فعالیت دارویی بازدارنده‌های غیر نوکلئوزیدی نسخه بردار معکوس با مشابهت ساختاری دی آریل پیریمیدین‌ها (DAPY-like NNRTIs) پرداخته شد. ابتدا ۱۰ توصیف‌کننده‌ی داکینگ مولکولی از برهمکنش بازدارنده‌ها - پروتئین مربوطه با استفاده از داکینگ مولکولی استخراج شد و پس از انتخاب موثرترین توصیف‌کننده‌ها، مدل شبکه عصبی توسعه یافت. به منظور بهبود مدل QSAR، برای اولین بار از ترکیب توصیف‌کننده‌های داکینگ مولکولی و توصیف‌کننده‌های شمارش گروه‌های عاملی استفاده شد. این ترکیب به عنوان یک دسته جدید از توصیف‌کننده‌ها در مطالعه QSAR فعالیت ضد HIV گروهی از ترکیبات DAPY-Like به عنوان NNRTIs جدید استفاده شد. ۱۵۴ توصیف‌کننده‌ی شمارش گروه‌های عاملی از طریق نرم افزار DRAGON 5.5 محاسبه و به ۱۰ توصیف‌کننده‌ی داکینگ مولکولی افزوده شد. موثرترین توصیف‌کننده‌ها از بین ۱۶۴ توصیف‌کننده با استفاده از رگرسیون گام به گام انتخاب و به عنوان ورودی شبکه عصبی مصنوعی استفاده شدند. عملکرد مدل توسعه یافته با پیش‌بینی مقادیر  $pEC_{50}$  سری ارزیابی و سری آزمون بررسی شد. نتایج به دست آمده، نشان‌دهنده قدرت پیش‌بینی مناسب مدل توسعه یافته با استفاده از ترکیب توصیف‌کننده‌های شمارش گروه‌های عاملی به توصیف‌کننده‌های داکینگ مولکولی می‌باشد.

### ۱-۲-۴ مجموعه داده‌ها

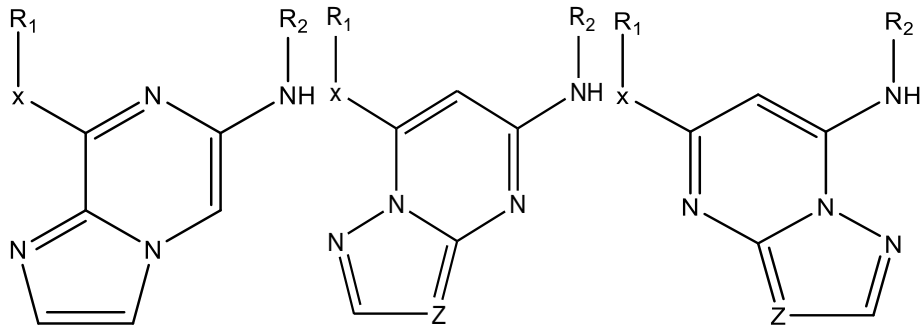
مجموعه داده‌های مورد استفاده در این تحقیق، فعالیت دارویی ۷۳ ترکیب از مشتقات DAPY-Like NNRTIs می‌باشد که از مطالعات هانگ<sup>۲</sup> و همکارانش استخراج شده است [۲۶-۲۴، ۹۹]. اسکلت

<sup>۱</sup> Diarylpyrimidine Like

<sup>۲</sup> Huang



اصلی ترکیبات (لیگاندها) در شکل (۱۱-۴) آورده شده است. ساختارهای شیمیایی و فعالیت دارویی (-)  $\log(EC_{50})$  ترکیبات مورد مطالعه در جدول (۱۷-۴) آورده شده است.  $EC_{50}$  کمترین غلظتی از دارو می باشد که تا ۵۰٪ اثر بازدارندگی داشته باشد. مقادیر  $EC_{50}$  در شرایط یکسان اندازه گیری شده است.



مولکول های ۱-۱۲

مولکول های ۱۳-۶۹

مولکول های ۷۰-۷۳

شکل (۱۱-۴) اسکلت های اصلی ترکیبات مورد مطالعه

جدول ( ٤-١٧) ساختار و مقادير pEC<sub>50</sub> تركيبات مورد مطالعه

شماره تركيب	R <sub>1</sub>	R <sub>2</sub>	Z	pEC <sub>50</sub>
١	2,6-Dimethyl-Ph	4-Cyano-Ph	C	٥/١٩
٢ <sup>t</sup>	2,4,6-Trimethyl-Ph	4-Cyano-Ph	C	٦/٥٩
٣ <sup>v</sup>	2,4,6-Trimethyl-Ph	4-Nitro-Ph	C	٥/٨٩
٤	2,4,6-Trimethyl-Ph	4-Cyano methylene-Ph	C	٥/٠٤
٥	2,6-Dimethyl-4-Cyano-Ph	4-Cyano-Ph	C	٦/٤٩
٦	2,6-Dimethyl-4-Cyano-Ph	4-Nitro-Ph	C	٥/٦٤
٧ <sup>v</sup>	2,6-Dimethyl-4-Cyano-Ph	4-Cyano methylene -Ph	C	٤/٧٢
٨	2,4,6-Trifluoro-Ph	4-Cyano-Ph	C	٥/١٤
٩	2,6-Dimethyl-4-Bromo-Ph	4-Cyano-Ph	C	٦/٠١
١٠	2,6-Dimethyl-4-Bromo-Ph	4-Nitro-Ph	C	٥/٤٠
١١	2,6-Dimethyl-4-Chloro-Ph	4-Cyano-Ph	C	٦/٠٦
١٢ <sup>t</sup>	2,6-Dimethyl-4-Chloro-Ph	4-Nitro-Ph	C	٥/٣٤
١٣	2,6-Dimethyl-4-Cyano-Ph	4-Cyano-Ph	C	٥/٨٢
١٤	2,6-Dimethyl-4-Cyano-Ph	4-Bromo-Ph	C	٥/٨٠
١٥	2,6-Dimethyl-4-Cyano-Ph	4-Chloro-Ph	C	٥/٩٥
١٦ <sup>t</sup>	2,6-Dimethyl-4-Cyano-Ph	4-Fluoro-Ph	C	٥/٥١
١٧	2,6-Dimethyl-4-Cyano-Ph	4-Methyl -Ph	C	٦/٣٤
١٨	2,6-Dimethyl-4-Cyano-Ph	4-Methoxy-Ph	C	٦/٠٤
١٩ <sup>v</sup>	2,6-Dimethyl-4-Cyano-Ph	4-Nitro-Ph	C	٥/٣٩
٢٠	2,4,6-Trimethyl-Ph	4-Cyano-Ph	C	٧/١٥
٢١	2,4,6-Trimethyl-Ph	4-Bromo-Ph	C	٦/٨٥
٢٢	2,4,6-Trimethyl-Ph	4-Chloro-Ph	C	٦/٧٧
٢٣	2,4,6-Trimethyl-Ph	4-Fluoro-Ph	C	٦/١٤
٢٤ <sup>t</sup>	2,4,6-Trimethyl-Ph	4-Methyl-Ph	C	٦/٧٧
٢٥	2,4,6-Trimethyl-Ph	4-Methoxy-Ph	C	٦/٥٥
٢٦	2,4,6-Trimethyl-Ph	4-Nitro-Ph	C	٦/١٦
٢٧	2,6-Dimethyl-4-Bromo-Ph	4-Cyano-Ph	C	٦/٣١
٢٨	2,4,6-Trifluoro-Ph	4-Cyano-Ph	C	٥/٢٢
٢٩ <sup>v</sup>	3,5-Dimethyl-Ph	4-Cyano-Ph	C	٥/٢٩
٣٠	2,6-Dimethyl-Ph	4-Cyano-Ph	N	٥/٨٩
٣١	2,6-Dichloro-Ph	4-Cyano-Ph	N	٥/٤٧
٣٢ <sup>v</sup>	2,4,6-Trimethyl-Ph	4-Cyano-Ph	N	٧/٣٠
٣٣	2,4,6-Trimethyl-Ph	4-Nitro-Ph	N	٦/٢٢
٣٤	2,4,6-Trimethyl-Ph	4-Chloro-Ph	N	٧/٣٠
٣٥	2,4,6-Trimethyl-Ph	4-Methyl-Ph	N	٧/١٥
٣٦ <sup>t</sup>	2,4,6-Trimethyl-Ph	4-Methoxy-Ph	N	٦/٧٠
٣٧	2,6-Dimethyl-4-Bromo-Ph	4-Cyano-Ph	N	٦/٦٤
٣٨	2,6-Dimethyl-4-Cyano-Ph	4-Cyano-Ph	N	٥/٩٢
٣٩	2,6-Dimethyl-4-Chloro-Ph	4-Cyano-Ph	N	٦/٤٠
٤٠	2,4,6-Trichloro-Ph	4-Cyano-Ph	N	٦/٢٢
٤١ <sup>v</sup>	2,4,6-Tribromo-Ph	4-Cyano-Ph	N	٦/٥٢
٤٢	2,6-Dibromo-4-Methyl-Ph	4-Cyano-Ph	N	٧/٣٠
٤٣	2,4,6-Trimethyl-Ph	4-Cyano-Ph	N	٧/٧٠
٤٤	2,6-Dimethyl-Ph	4-Cyano-Ph	N	٥/٦٨
٤٥ <sup>t</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -Cyclohexyl	N	٦/٦١
٤٦	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -Pyridin-4-yl	N	٧/٧٧
٤٧ <sup>t</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -Ph	N	٧/٤٤
٤٨	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-SO <sub>2</sub> Me-Ph	N	٨/٠٩
٤٩	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-SO <sub>2</sub> NH <sub>2</sub> -Ph	N	٧/٥٧
٥٠	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-COOMe-Ph	N	٧/٨٢
٥١	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-NO <sub>2</sub> -Ph	N	٧/٤٦
٥٢	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-CONH <sub>2</sub> -Ph	N	٧/٦٨
٥٣	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-COCH <sub>3</sub> -Ph	N	٨/٠٦
٥٤ <sup>v</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-Bromo-Ph	N	٧/٠٤
٥٥	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-Chloro-Ph	N	٧/٩٢

ادامه جدول (۴-۱۷)

شماره ترکیب	R <sub>1</sub>	R <sub>2</sub>	z	pEC <sub>50</sub>
۵۶	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-Cyano-Ph	N	۷/۷۰
۵۷	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -3-Cyano-Ph	N	۷/۷۲
۵۸ <sup>t</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -2-Cyano-Ph	N	۷/۱۱
۵۹	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-Fluoro-Ph	N	۷/۲۵
۶۰	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -3-Fluoro-Ph	N	۶/۹۸
۶۱ <sup>v</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -2-Fluoro-Ph	N	۷/۳۲
۶۲	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -4-CH <sub>3</sub> -Ph	N	۷/۱۴
۶۳ <sup>t</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -2-CH <sub>3</sub> -Ph	N	۷/۳۸
۶۴ <sup>v</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -2,4-Difluoro-Ph	N	۷/۰۵
۶۵	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -3,4-Difluoro-Ph	N	۷/۱۷
۶۶	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -2,5-Difluoro-Ph	N	۷/۰۱
۶۷	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> -2,6-Difluoro-Ph	N	۶/۳۶
۶۸ <sup>t</sup>	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> CONH-2-Chloro-4-COOMe-Ph	N	۶/۵۶
۶۹	2,4,6-Trimethyl-Ph	PPR-CH <sub>2</sub> CONH-2-Bromo-4-COOH-Ph	N	۵/۶۴
۷۰	2,6-Dimethyl-Ph	4-Cyano-Ph	N	۵/۸۲
۷۱	2,4,6-Trimethyl-Ph	4-Cyano-Ph	N	۷/۱۵
۷۲ <sup>v</sup>	2,6-Dimethyl-4-Bromo-Ph	4-Cyano-Ph	N	۶/۵۲
۷۳	2,4,6-Trichloro-Ph	4-Cyano-Ph	N	۶/۵۲

\* v داده‌های سری ارزیابی، t داده‌های سری آزمون، Ph: مخفف Phenyl، PPR: مخفف Piperidine و Me: مخفف Methyl. X در همه ترکیبات نشان‌دهنده O بوده و در ترکیبات ۴۳ و ۴۴ نشان‌دهنده NH است.

#### ۲-۲-۴ رسم و بهینه سازی ساختار ترکیبات مورد مطالعه

به منظور محاسبه و استخراج توصیف کننده‌ها، رسم و بهینه سازی ساختارهای مولکولی ترکیبات الزامی است. از این رو کلیه ساختارهای مورد مطالعه مطابق با بخش (۲-۳) در نرم افزار Hyperchem رسم و بهینه سازی شدند. سپس ساختارهای بهینه شده با دو فرمت (\*.hin) و (\*.mol) ذخیره شدند.

#### ۳-۲-۴ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه با استفاده از توصیف کننده‌های حاصل

##### از داکینگ مولکولی

به منظور ساخت مدل QSAR با استفاده از الگوریتم شبکه عصبی مصنوعی، توصیف کننده‌های حاصل از داکینگ مولکولی برای ترکیبات مورد مطالعه استخراج گردید و در نهایت از بین مدل‌های طراحی شده، مدل برتر QSAR معرفی و ارزیابی شد. در ادامه به چگونگی استخراج توصیف کننده‌های حاصل از داکینگ مولکولی، انتخاب موثرترین آن‌ها و ارزیابی مدل اشاره خواهد شد.

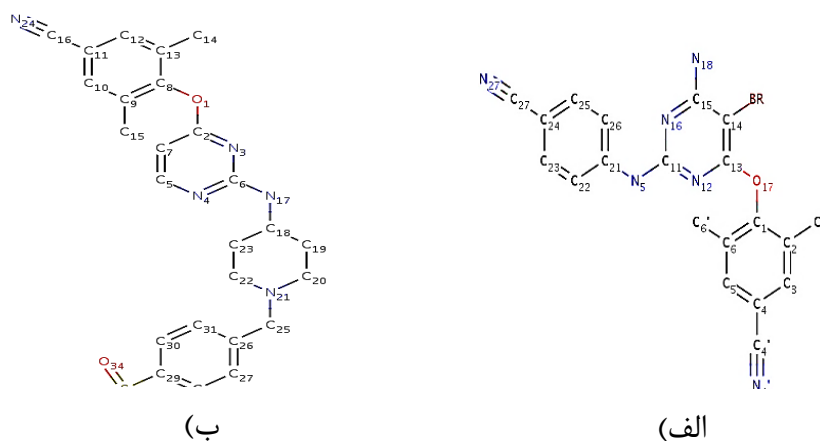
#### ۱-۲-۲-۴ استخراج توصیف کننده‌های حاصل از داکینگ مولکولی

به منظور استخراج توصیف کننده‌های داکینگ مولکولی، کلیه لیگاندها در شرایط بهینه داکینگ به طور جداگانه در جایگاه فعال پروتئین مربوطه (مولکول‌های ۱-۴۴ و ۷۳-۷۰ در پروتئین 3MEC و

مولکول‌های ۴۵-۶۹ در پروتئین 3M8Q) داک شدند. در نهایت توصیف کننده‌های داکینگ مولکولی با استفاده از اطلاعات برهمکنش پایدارترین کمپلکس لیگاند-گیرنده استخراج شدند. در ادامه روش کار و نتایج حاصل از آن آورده خواهد شد.

### - استخراج و آماده سازی لیگاند موجود در پروتئین

لازمه انجام فرایند اعتبارسنجی، آماده سازی لیگاند کریستالوگرافی و پروتئین می‌باشد. با استناد به گزارش‌های موجود [۲۶-۲۴، ۹۹] ساختارهای کریستالوگرافی 3M8Q و 3MEC و لیگاندهای هدف این دو پروتئین با فرمت pdb از بانک اطلاعاتی پروتئین دریافت شدند [۶۲] و مطابق با توضیحات بخش (۳-۱-۳) استخراج و آماده سازی لیگاند کریستالوگرافی انجام شد. ساختارهای لیگاندهای کریستالوگرافی در هر دو پروتئین مورد مطالعه در شکل (۴-۱۲) نمایش داده شده است.

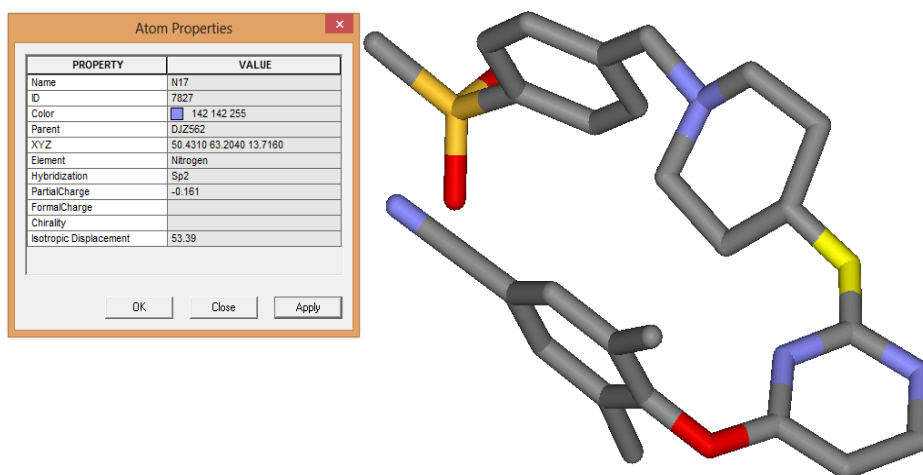


شکل (۴-۱۲) ساختار لیگاندهای موجود در پروتئین الف) 3MEC ب) 3M8Q

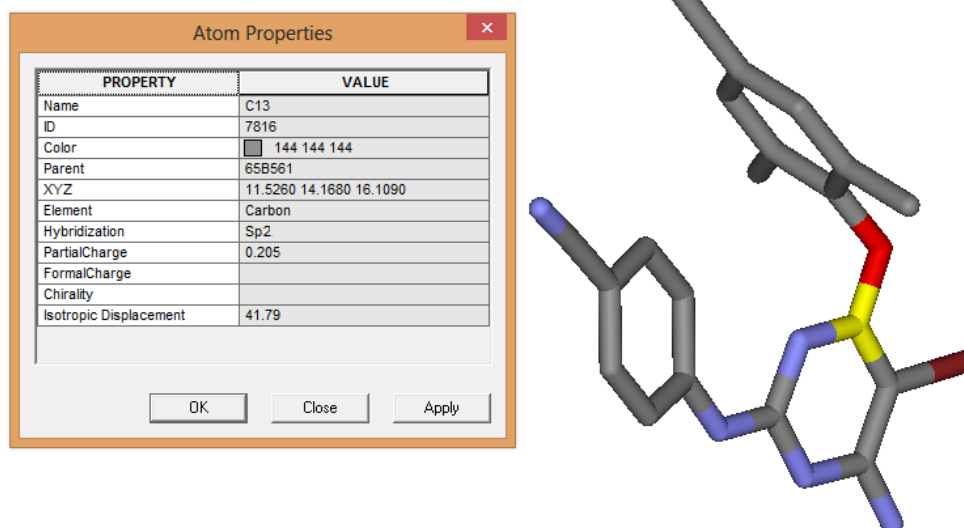
مختصات مرکز ثقل هر لیگاند به‌عنوان مختصات جایگاه فعال پروتئین مربوطه در نظر گرفته شد و از آن در مراحل بعدی داکینگ مولکولی نیز استفاده شد. استخراج مختصات مرکز ثقل لیگاندها در نرم افزار ViewerLite5.0 طبق مراحل نشان داده شده در شکل (۴-۱۳) و شکل (۴-۱۴) انجام گردید. نتایج به‌دست آمده مختصات ۵۰/۴۳۱ ، ۶۳/۲۰۴ ، ۱۳/۷۱۶ برای مرکز ثقل لیگاند موجود در پروتئین

3M8Q و مختصات ۱۱/۵۲۸، ۱۴/۱۶۸ و ۱۶/۱۰۹ برای مرکز ثقل لیگاند موجود در پروتئین 3MEC را

نشان داد.



شکل (۴-۱۳) مختصات جایگاه فعال به دست آمده از جایگاه لیگاند در پروتئین 3M8Q



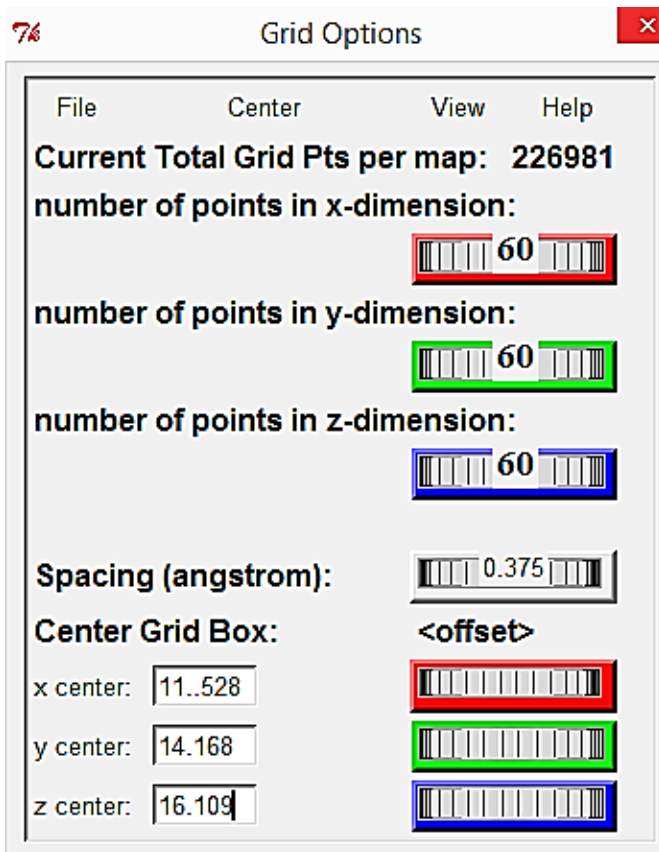
شکل (۴-۱۴) مختصات جایگاه فعال به دست آمده از جایگاه لیگاند در پروتئین 3MEC

## - آماده سازی پروتئین

به منظور آماده سازی پروتئین جهت استفاده در فرایند اعتبارسنجی و داکینگ لیگاندها، دو ساختار کریستالوگرافی 3M8Q و 3MEC با فرمت pdb از بانک اطلاعاتی پروتئین [۶۲] دریافت شدند. ارزش تفکیک مربوط به ساختارهای کریستالوگرافی برای پروتئین با شناسه کریستالوگرافی 3M8Q برابر با  $2/7A^{\circ}$  و برای پروتئین با شناسه کریستالوگرافی 3MEC برابر با  $2/3A^{\circ}$  می باشد. پس از فراخوانی مجزای دو ساختار کریستالوگرافی 3M8Q و 3MEC در نرم افزار ViewerLite5.0، آماده سازی پروتئین مطابق بخش (۲-۳-۳) انجام شد و با فرمت pdb ذخیره شد. آماده سازی بیشتر پروتئین ها در نرم افزار Autodock4.2 انجام شد.

## - فرایند اعتبارسنجی (داک-ریداک)

به منظور یافتن شرایط بهینه، لیگاند کریستالوگرافی و پروتئین مربوطه در نرم افزار Autodock4.2 فراخوانی و مطابق با بخش (۱-۳-۳-۳) آماده سازی های بیشتر انجام شد. سپس مختصات مرکز و ابعاد شبکه در برگیرنده جایگاه فعال مطابق با بخش (۲-۳-۳-۳) اعمال و ذخیره شد. شکل (۱۵-۴) روند تعریف ابعاد شبکه و مختصات جایگاه فعال را نشان می دهد. برای اجرای داکینگ مولکولی از الگوریتم ژنتیک لامارکین در سه حالت با تعداد اجرای الگوریتم ۱۰۰، ۱۵۰ و ۲۰۰ استفاده شد و مطابق با بخش های (۴-۳-۳-۳ و ۵-۳-۳-۳)، سایر پارامترهای داکینگ مولکولی نیز در مقادیر پیش فرض قرار داده شدند. پس از اعمال تمامی گزینه های مربوط به انجام فرایند داکینگ مولکولی شرایط بهینه با آنالیز فایل خروجی (dlg) به دست آمد. با توجه به معیارهای اشاره شده در بخش (۳-۳-۳) تعداد اجرای الگوریتم ژنتیک برابر با ۱۵۰ دارای RMSD کمتر از  $2A^{\circ}$ ، انرژی اتصال منفی تر، تعداد خوشه کمتر، حداکثر تعداد کنفورماسیون در خوشه برتر و زمان اجرای کمتر می باشد. نتایج حاصل در جدول (۱۸-۴) و جدول (۱۹-۴) آورده شده است.



شکل ( ۴-۱۵) تعیین مختصات شبکه در بخش اتوگرید نرم افزار

جدول ( ۴-۱۸) نتایج اجراهای الگوریتم ژنتیک متفاوت برای پروتئین با شناسه کریستالوگرافی 3MQ8

زمان اجرا (دقیقه)	تعداد کنفورماسیون‌های موجود در خوشه اول	تعداد خوشه‌های چند کنفورماسیونه	میانگین انرژی اتصال (kcal/mol)	RMSD (Å)	تعداد اجرا الگوریتم ژنتیک
۷۴	۲۸	۱۳	-۱۱/۶۹	۰/۴۴	۱۰۰
۱۱۶	۷۵	۷	-۱۲/۲۰	۰/۶۹	۱۵۰
۱۴۲	۵۴	۱۷	-۱۱/۷۶	۰/۵۱	۲۰۰

جدول ( ۴-۱۹) نتایج اجراهای الگوریتم ژنتیک متفاوت برای پروتئین با شناسه کریستالوگرافی 3MEC

زمان اجرا (دقیقه)	تعداد کنفورماسیون‌های موجود در خوشه اول	تعداد خوشه‌های چند کنفورماسیونه	میانگین انرژی اتصال (kcal/mol)	RMSD (Å)	تعداد اجرا الگوریتم ژنتیک
۳۴	۴۳	۱۰	-۱۱/۶۴	۰/۶۲	۱۰۰
۵۵	۱۳۴	۵	-۱۱/۹۳	۰/۶۱	۱۵۰
۱۴۲	۱۰۴	۱۲	-۱۱/۸۳	۰/۶۳	۲۰۰

## ❖ داکینگ لیگاند های مورد مطالعه در پروتئین های مربوطه

بعد از آماده سازی لیگاندها و پروتئین ها، عملیات داکینگ مولکولی لیگاندها با پروتئین مربوطه در شرایط بهینه حاصل از روش اعتبارسنجی انجام شد و توصیف کننده های مولکولی با توجه به اطلاعات بهترین اتصال لیگاند-گیرنده استخراج گردید. در ادامه توصیف کننده های حاصل از داکینگ مولکولی معرفی و پس از انتخاب بهترین توصیف کننده ها، مدل های QSAR و نتایج حاصل از آن ها ارائه شد.

### - آماده سازی لیگاندهای مورد مطالعه

ساختار ترکیبات مورد مطالعه مطابق بخش (۲-۳)۰ رسم و بهینه سازی شدند. سپس ساختارهای بهینه شده با فرمت mol\* به نرم افزار ViewerLite5.0 فراخوانده و با پسوند pdb ذخیره شدند و در نهایت به عنوان ورودی های نرم افزار Autodock4.2 مورد استفاده قرار گرفتند.

### - اجرای عملیات داکینگ مولکولی لیگاندها و پروتئین

پس از آماده سازی لیگاندهای مورد مطالعه و پروتئین (پروتئین 3MEC برای لیگاندهای ۴۴-۱ و ۷۳-۷۰ و پروتئین 3M8Q برای لیگاندهای ۶۹-۴۵)، پروتئین در نرم افزار Autodock4.2 فراخوانی شد و آماده سازی بیشتر مطابق بخش (۳-۳-۳-۱) برای پروتئین انجام گرفت. سپس تمامی لیگاندها به طور جداگانه با توجه به پروتئین متناظر در نرم افزار Autodock4.2 فراخوانی شدند. در ادامه مطابق بخش (۳-۳-۳-۲) مختصات جایگاه فعال و شبکه بر روی پروتئین اعمال گردید. پس از انتخاب الگوریتم ژنتیک لامارکین و تنظیم تمامی گزینه های داکینگ و اجرای الگوریتم ژنتیک لامارکین در شرایط بهینه حاصل از مرحله اعتبارسنجی، لیگاندها در پروتئین مربوطه داک شدند. پس از اتمام اجرای داکینگ مولکولی، نتایج خروجی با فرمت dlg ذخیره و در نهایت توصیف کننده هایی با اطلاعات ارزشمند حاصل از برهم کنش لیگاند-پروتئین در خوشه ی برتر با بیشترین صورتبندی و کمترین انرژی اتصال استخراج و آنالیز شدند.



## - استخراج توصیف کننده‌های حاصل از داکینگ مولکولی

یکی از مهم‌ترین مراحل مدل‌سازی، محاسبه توصیف کننده‌ها می‌باشد. به این منظور در تحقیق حاضر برهمکنش‌های بین مهار کننده‌ها و پروتئین مربوطه با استفاده از داکینگ مولکولی آنالیز شدند و ۱۰ توصیف کننده عددی ذکر شده در بخش (۷-۵-۲) با استفاده از نرم افزار ++Notpad استخراج گردید. این ۱۰ توصیف کننده که حاوی اطلاعات مفیدی از برهمکنش لیگاند-پروتئین هستند برای تمام ۷۳ لیگاند مورد مطالعه استخراج گردید و به‌عنوان متغیرهای مستقل برای ساخت مدل QSAR استفاده شدند.

### ۲-۳-۲-۴ انتخاب موثرترین توصیف کننده‌های داکینگ مولکولی

در مطالعات QSAR، هر چه تعداد توصیف کننده‌های به کار رفته در ساخت مدل کمتر باشد، از پیچیدگی مدل کاسته می‌شود. لذا در این مطالعه، ۱۰ توصیف کننده استخراج شده از داکینگ مولکولی به‌عنوان توصیف کننده و منفی لگاریتم فعالیت دارویی ( $pEC_{50}$ ) به‌عنوان پاسخ در نظر گرفته شد و سپس برای انتخاب موثرترین توصیف کننده‌ها مطابق بخش (۴-۳) عمل شد. در نهایت ۲ توصیف کننده به‌عنوان موثرترین توصیف کننده‌ها انتخاب شدند که نام و نوع آن‌ها در جدول (۴-۲۰) نشان داده شده است. جدول (۴-۲۱) ماتریس همبستگی بین توصیف کننده‌ها را نشان می‌دهد، نتایج حاصل عدم وجود همبستگی معنادار بین این توصیف کننده‌ها را تایید می‌نماید.

جدول (۴-۲۰) توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

No	Symbol	Meaning
۱	VE	(vdW + Hbond + desolv) Energy
۲	pK <sub>i</sub>	-Log( Estimated Inhibition Constant, K <sub>i</sub> )

جدول (۴-۲۱) ماتریس همبستگی توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

	VE	pK <sub>i</sub>
VE	۱	
pK <sub>i</sub>	-۰/۶۹	۱

به منظور بررسی هم‌خطی بین توصیف‌کننده‌های منتخب، کمیت عامل تورم واریانس (VIF) برای هر توصیف‌کننده مطابق با رابطه (۲-۱۶) محاسبه شد که نتایج آن در جدول (۴-۲۲) آورده شده است. با توجه به مقادیر VIF کمتر از ۱۰، عدم وجود هم‌خطی در بین توصیف‌کننده‌ها اثبات می‌گردد.

جدول (۴-۲۲) مقادیر عامل تورم واریانس توصیف‌کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

توصیف‌کننده	VE	pK <sub>i</sub>
VIF	۴/۶۱	۴/۶۱

### ۳-۳-۲-۴ مدل شبکه عصبی مصنوعی

نظر به این‌که پدیده‌های بیولوژیکی اغلب دارای ویژگی‌های غیر خطی هستند، در این پژوهش از شبکه عصبی مصنوعی به‌عنوان روش غیر خطی برای کشف رابطه ممکن بین توصیف‌کننده‌های ورودی و فعالیت دارویی (EC<sub>50</sub>) ترکیبات جدید مورد مطالعه استفاده شد. شبکه طراحی شده در این مطالعه، شبکه پیشرو با الگوریتم پس انتشار می‌باشد.

به منظور ساخت مدل، ۴ شبکه عصبی با توجه به دو تابع آموزش لونیگ-مارکوات و تنظیم بیزی و دو تابع انتقال تانژانت سیگموئید و لگاریتم سیگموئید طراحی شد. مجموعه داده‌ها به‌طور تصادفی به سه دسته سری آموزش (۵۳ ترکیب)، سری ارزیابی (۱۰ ترکیب) و سری آزمون (۱۰ ترکیب) تقسیم شد. برای بررسی توانمندی مدل از سری آزمون استفاده شد. سری ارزیابی در طول توسعه مدل به منظور جلوگیری از بیش‌برازش شبکه در اثر آموزش بیش از حد استفاده شد تا به محض بروز بیش‌برازش و افزایش خطای سری ارزیابی، آموزش شبکه متوقف گردد [۱۰۰]. در هر یک از این شبکه‌ها مدل با تعداد ورودی ۲، تعداد گره از ۲ تا ۱۰ و تعداد دور آموزشی از ۲ تا ۵۰ با گام ۲ به صورت همزمان بهینه شد. ملاک بهینه‌سازی، کمینه‌سازی MSE سری ارزیابی، است. نتایج حاصل از بهینه‌سازی شبکه‌های طراحی شده، در جدول (۴-۲۳) جمع‌آوری شد. از بین مدل‌های ساخته شده، مدل شبکه عصبی با تعداد ۲ توصیف‌کننده در لایه اول، ۷ گره در لایه میانی، ۵۰ دور آموزش، تابع آموزشی لونیگ-مارکوات و تابع انتقال لگاریتم سیگموئیدی دارای حداقل مقدار MSE است.

جدول ( ۴-۲۳ ) توابع و پارامترهای شبکه‌های بهینه شده با توصیف کننده‌های انتخاب شده داکینگ مولکولی

MSE	تعداد دور آموزش	تعداد گره	تابع انتقال	تابع آموزش	تعداد توصیف کننده
۰/۴۹	۶	۳	لگاریتم-سیگموئید	تنظیم بیزی	۲
۰/۲۶	۵۰	۷	لگاریتم-سیگموئید	لونبرگ-مارکوات	۲
۰/۴۹	۶	۲	تانژانت-سیگموئید	تنظیم بیزی	۲
۰/۳۲	۱۴	۹	تانژانت-سیگموئید	لونبرگ-مارکوات	۲

#### ۴-۲-۳-۴ ارزیابی مدل شبکه عصبی مصنوعی

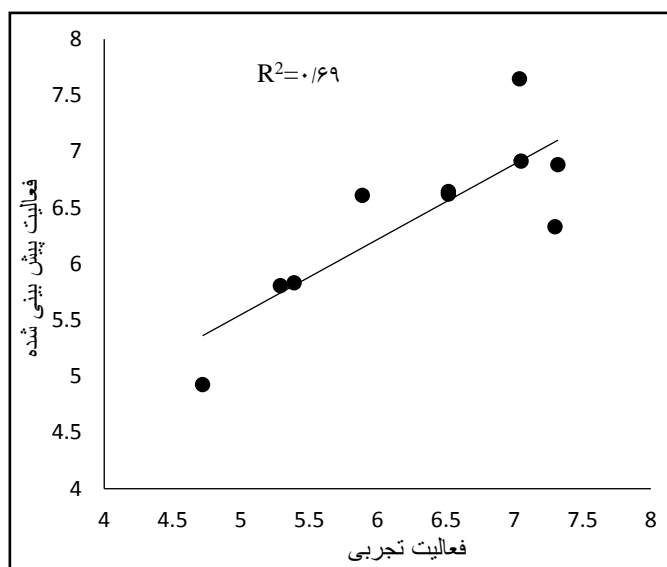
بررسی اعتبار و اهمیت مدل در مطالعات رابطه کمی ساختار-فعالیت الزامی است. قدرت پیش‌بینی مدل وابسته به توانایی آن در ارائه خروجی‌های رضایت‌بخش برای مولکول‌هایی است که در سری آموزش وجود نداشته‌اند. در این تحقیق برای ارزیابی مدل ارائه شده روش‌های زیر به کار گرفته شد.

#### - ارزیابی مدل با استفاده از سری ارزیابی

به این منظور، شبکه عصبی با ۵۳ ترکیب سری آموزش، آموزش داده شد. معیار بهینه سازی شبکه، حداقل بودن MSE سری ارزیابی (۱۰ ترکیب) است. نتایج حاصل در جدول ( ۴-۲۴ ) آورده شده است. ترسیم از فعالیت پیش‌بینی شده در برابر فعالیت تجربی در شکل ( ۴-۱۶ ) نشان داده شده است.

جدول ( ۴-۲۴ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری ارزیابی

شماره ترکیب	pEC <sub>50</sub>		درصد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۳	۵/۸۹	۶/۶۱	۱۲/۲۲
۷	۴/۷۲	۴/۹۳	۴/۴۵
۱۹	۵/۳۹	۵/۸۳	۸/۱۶
۲۹	۵/۲۹	۵/۸۱	۹/۸۳
۷۲	۶/۵۲	۶/۶۴	۱/۸۴
۳۲	۷/۳	۶/۳۳	-۱۳/۲۹
۴۱	۶/۵۲	۶/۶۲	۱/۵۳
۵۴	۷/۰۴	۷/۶۵	۸/۶۶
۶۱	۷/۳۲	۶/۸۸	-۶/۰۱
۶۴	۷/۰۵	۶/۹۲	-۱/۸۴



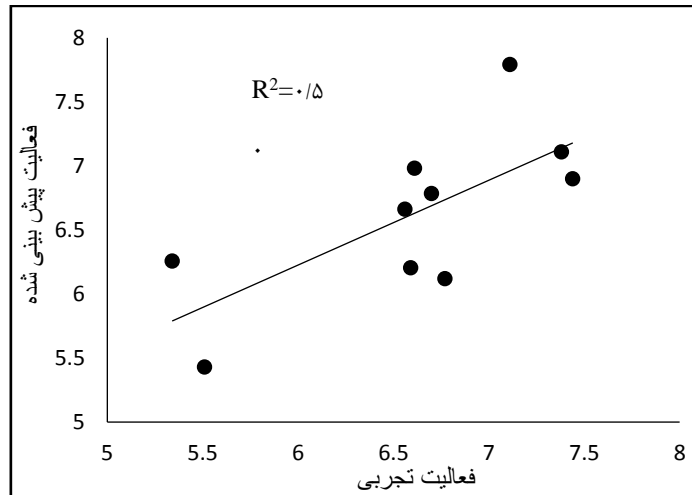
شکل (۴-۱۶) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی

#### - ارزیابی مدل با استفاده از سری آزمون

ارزیابی مدل با پیش‌بینی فعالیت ۱۰ ترکیب سری آزمون که در هیچ کدام از مراحل آموزش و ارزیابی شبکه شرکت نداشته‌اند، نیز انجام شد. بدین منظور، فعالیت این ترکیبات با استفاده از مدل بهینه پیش‌بینی شد. نتایج حاصل در جدول (۴-۲۵) آمده است. شکل (۴-۱۷) نمودار تغییرات فعالیت پیش‌بینی شده بر حسب فعالیت تجربی سری آزمون را نشان می‌دهد. مقدار ضریب تعیین ( $R^2$ ) به دست آمده برای این سری ۰/۵۰ می‌باشد.

جدول (۴-۲۵) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون

شماره ترکیب	pEC <sub>50</sub>		در صد خطا
	فعالیّت تجربی	فعالیّت پیش‌بینی شده	
۲	۶/۵۹	۶/۲۰	-۵/۹۲
۱۲	۵/۳۴	۶/۲۶	۱۷/۲۳
۱۶	۵/۵۱	۵/۴۳	-۱/۴۵
۲۴	۶/۷۷	۶/۱۲	-۹/۶۰
۳۶	۶/۷۰	۶/۷۹	۱/۳۴
۴۵	۶/۶۱	۶/۹۸	۵/۶۰
۴۷	۷/۴۴	۶/۹۰	-۷/۲۶
۵۸	۷/۱۱	۷/۷۹	۹/۵۶
۶۳	۷/۳۸	۷/۱۱	-۳/۶۶
۶۸	۶/۵۶	۶/۶۶	۱/۵۲



شکل (۴-۱۷) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون

با توجه به نتایج مدل سازی و ارزیابی‌های انجام شده، اگرچه مقدار ضریب تعیین ( $R^2$ ) به دست آمده برای سری ارزیابی دارای مقدار قابل قبول ( $R^2 > 0.6$ ) [۷۶] است، اما  $R^2$  به دست آمده برای سری آزمون، کوچک‌تر از مقدار قابل قبول است. بنابراین می‌توان نتیجه گرفت که مدل غیرخطی شبکه عصبی مصنوعی با توصیف کننده‌های داکینگ منتخب از صحت پیش‌بینی پایینی برخوردار است. به منظور بهبود مدل غیر خطی شبکه عصبی مصنوعی، توصیف کننده‌های شمارش گروه‌های عاملی با استفاده از نرم افزار DRAGON محاسبه و به توصیف کننده‌های داکینگ مولکولی افزوده شدند. با توجه به این که توصیف کننده‌های ساختاری بیان بهتری از ویژگی‌های شیمیایی ساختار را دارند، انتظار می‌رود ارتباط توصیف کننده‌ها با فعالیت دارویی بهبود یابد.

#### ۴-۲-۴ مدل سازی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

در راستای بهبود مدل شبکه عصبی مصنوعی توسعه یافته به وسیله توصیف کننده‌های داکینگ مولکولی (۱۰ توصیف کننده)، توصیف کننده‌های شمارش گروه عاملی با استفاده از نرم افزار DRAGON محاسبه و به مجموعه توصیف کننده‌های داکینگ مولکولی افزوده شدند. سپس با استفاده از روش انتخاب متغیر رگرسیون گام به گام، موثرترین توصیف کننده‌ها انتخاب و به عنوان ورودی شبکه عصبی مصنوعی استفاده شدند. سپس کارایی بهترین مدل با استفاده از ارزیابی سری آزمون و تکنیک رد مرحله‌ای تک تک برای همه ترکیبات مجموعه داده‌ها بررسی شد.

## ۱-۴-۲-۴ توصیف کننده‌های شمارش گروه‌های عاملی

محاسبه توصیف کننده‌هایی که در عین سادگی و قابل تفسیر بودن، بتوانند کل ویژگی‌های ترکیبات مورد بررسی را نشان دهند و به علاوه به راحتی قابل محاسبه باشند، بسیار قابل توجه است. در این مطالعه، مطابق بخش (۲-۳)، پس از رسم و بهینه سازی ساختار ۷۳ ترکیب مورد مطالعه، ۱۵۴ توصیف کننده از دسته‌ی شمارش گروه‌های عاملی به وسیله‌ی نرم افزار DRAGON محاسبه شد. این گروه مستقل از صورت‌بندی مولکولی هستند و از فرمول شیمیایی مولکول مشتق می‌شوند. این توصیف کننده‌ها به توصیف کننده‌های ساختاری نیز معروفند که تعداد و نوع اتم‌ها و هر عاملی از خصوصیات اتمی را تعیین می‌کنند. همچنین شامل توصیف کننده‌های مربوط به نوع پیوندها، تعداد پیوندها (ساده، دوگانه، سه‌گانه، آروماتیک) و تعداد کل حلقه‌ها بر اساس تعداد اتم‌های آن (حلقه‌های شش‌تایی، پنج‌تایی و ...)، تعداد گروه‌های عاملی خاص و تعداد یک عنصر شیمیایی خاص (C, N, O, H, F, ...) در مولکول نیز می‌شوند. در این گروه ممکن است مقادیر عددی، مربوط به حضور و عدم حضور یک ویژگی باشد که به ترتیب با مقادیر ۱ و ۰ نشان داده می‌شود. برخی توصیف کننده‌ها نیز بر اساس تعداد ویژگی مربوطه محاسبه می‌شوند و مقادیر عددی شمارشی را به خود اختصاص می‌دهند.

## ۲-۴-۲-۴ انتخاب موثرترین توصیف کننده‌ها از مجموع توصیف کننده‌های داکینگ مولکولی و

### ساختاری

همانطور که قبلاً در بخش (۴-۳) اشاره شد مهم‌ترین مسئله در توسعه مدل‌های QSAR، انتخاب توصیف کننده‌هایی است که دارای ارتباط بسیار زیادی با فعالیت دارویی مورد نظر باشند. بنابراین باید با استفاده از روش‌های موجود، تعداد توصیف کننده‌ها را کاهش داد. در این پژوهش برای کاهش اندازه ماتریس داده‌ها، موثرترین توصیف کننده‌ها مطابق بخش (۴-۳) با استفاده از نرم افزار SPSS و با اجرای روش رگرسیون گام به گام (SR) بر روی ۱۶۴ توصیف کننده به تعداد ۷ توصیف کننده انتخاب شدند که نام و نوع آن‌ها در جدول (۴-۲۶) نشان داده شده است.

جدول (۴-۲۶) توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR

No	Symbol	Meaning	Class
۱	VE	vdW + Hbond + desolv Energy	Docking derived
۲	nCP	Number of terminal primary C (SP3)	Molecular properties descriptors
۳	nRCN	Number of nitriles (aliphatic)	Molecular properties descriptors
۴	EE	Electrostatic Energy	Docking derived descriptors
۵	K <sub>i</sub>	Estimated Inhibition Constant, Ki	Docking derived descriptors
۶	nARNO2	Number of nitro groups (aromatic)	Molecular properties descriptors
۷	nPyridines	Number of pyridines	Molecular properties descriptors

جدول (۴-۲۷) ماتریس همبستگی بین توصیف کننده‌های منتخب را نشان می‌دهد، این ماتریس

عدم وجود همبستگی معنادار بین این توصیف کننده‌ها را تایید می‌نماید.

جدول (۴-۲۷) ماتریس همبستگی توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با

روش SR

	VE	nCP	nRCN	EE	K <sub>i</sub>	nARNO2	nPyridines
VE	۱						
nCP	۰/۲۹	۱					
nRCN	۰/۰۴	۰/۰۰	۱				
EE	-۰/۳۷	-۰/۰۶	-۰/۱۵	۱			
K <sub>i</sub>	۰/۴۰	-۰/۰۱	-۰/۰۵	-۰/۰۱	۱		
nARNO2	۰/۰۹	۰/۰۱	-۰/۰۶	-۰/۴۰	-۰/۰۷	۱	
nPyridines	-۰/۰۲	۰/۰۶	-۰/۰۲	۰/۰۶	-۰/۰۲	-۰/۰۴	۱

به منظور بررسی وجود یا عدم وجود هم خطی در بین توصیف کننده‌های انتخاب شده، عامل

تورم واریانس مربوط به هر توصیف کننده، محاسبه شد. مقدار این کمیت برای توصیف کننده‌ها در جدول

(۴-۲۸) گزارش شده است و با توجه به این که برای کلیه توصیف کننده‌ها، عامل تورم واریانس از

۱۰ کمتر است، عدم وجود هم خطی در بین توصیف کننده‌ها اثبات می‌گردد.

جدول (۴-۲۸) مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های داکینگ مولکولی و

توصیف کننده‌های ساختاری با روش SR

توصیف کننده	VE	nCP	nRCN	EE	k <sub>i</sub>	nARNO2	nPyridines
VIF	۱/۶	۱/۱۶	۱/۰۵	۱/۵۱	۱/۲۵	۱/۲۲	۱/۰۱

## ۳-۴-۲-۴ مدل سازی شبکه عصبی مصنوعی با استفاده از مجموع توصیف کننده‌های داکینگ

### مولکولی و ساختاری

در این مرحله نیز به منظور مدل سازی فعالیت دارویی ترکیبات ضد HIV شبکه عصبی پیشرو با الگوریتم آموزشی پس انتشار انتخاب شد و به منظور فرآیند بهینه سازی پارامترهای شبکه، مجموعه داده‌ها به سه بخش سری آزمون (۱۰ ترکیب) و سری ارزیابی (۱۰ ترکیب) و همچنین سری آموزش (۵۳ ترکیب) تقسیم شدند. مقادیر توصیف کننده‌ها به‌عنوان ورودی به شبکه داده شد. مقدار تجربی فعالیت دارویی ضد HIV (سری آموزش) نیز به‌عنوان هدف به دنبال ورودی‌ها وارد شبکه شدند. برای آموزش شبکه دو الگوریتم آموزشی لونیبرگ-مارکوات و تنظیم بیزی انتخاب شد. دو تابع انتقال تانژانت سیگموئید و لگاریتم سیگموئید به‌عنوان تابع انتقال در لایه پنهان در نظر گرفته شد. در هر یک از شبکه‌های طراحی شده بهینه سازی همزمان تعداد ورودی‌ها، تعداد نرون‌های لایه پنهان (گره) و تعداد دور آموزشی انجام شد. در طی فرایند آموزش، MSE سری ارزیابی محاسبه گردید و معیار بهینه سازی، کمینه سازی مقدار MSE سری ارزیابی در نظر گرفته شد. برای این منظور هر شبکه با تعداد ورودی توصیف کننده‌ها و تعداد گره از ۲ تا ۱۰ با گام ۱ و تعداد دور آموزشی از ۲ تا ۵۰ با گام ۲ به‌طور همزمان آموزش داده شد. نتایج حاصل در جدول (۴-۲۹) خلاصه شد.

جدول (۴-۲۹) توابع و پارامترهای شبکه‌های بهینه توصیف کننده‌های منتخب از مجموع توصیف کننده‌های ساختاری و داکینگ

#### مولکولی با روش SR

MSE	تعداد دور آموزش	تعداد گره	تابع انتقال	تابع آموزش	تعداد توصیف کننده
۰/۰۷	۱۴	۳	لگاریتم-سیگموئید	تنظیم بیزی	۴
۰/۰۵	۱۰	۴	لگاریتم-سیگموئید	لونیبرگ-مارکوات	۶
۰/۰۷	۲۴	۹	تانژانت-سیگموئید	تنظیم بیزی	۶
۰/۰۶	۶	۶	تانژانت-سیگموئید	لونیبرگ-مارکوات	۶

از میان شبکه‌های ایجاد شده، شبکه با تعداد ۶ توصیف کننده در لایه اول، ۴ گره در لایه میانی، ۱۰ دور آموزشی، تابع آموزش لونیبرگ-مارکوات و تابع انتقال لگاریتم سیگموئیدی به‌عنوان شبکه برتر انتخاب شد.



#### ۴-۲-۴-۴ ارزیابی مدل شبکه عصبی مصنوعی

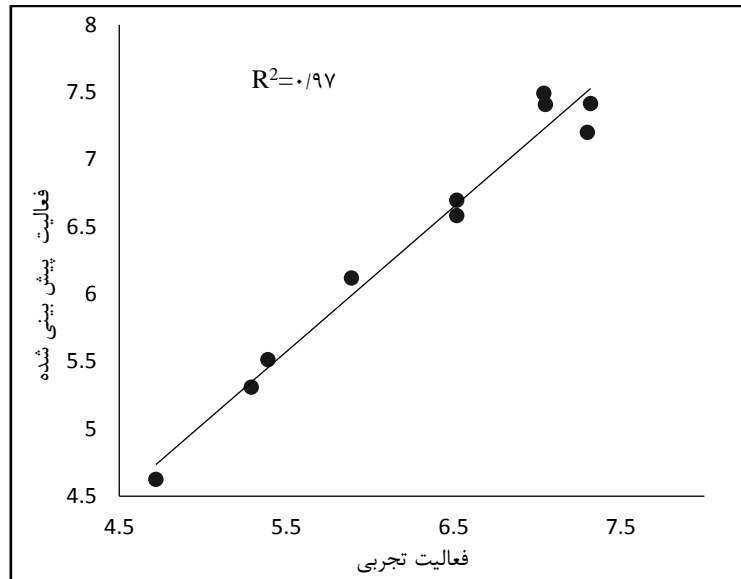
اساسی‌ترین مرحله در مطالعات رابطه کمی ساختار-فعالیت ارزیابی قدرت پیش‌بینی مدل می‌باشد. قدرت پیش‌بینی مدل شبکه عصبی با بررسی توانایی آن در ارائه پیش‌بینی‌های رضایت‌بخش برای مولکول‌هایی که در سری آموزش وجود نداشته‌اند، تعریف می‌شود. در این تحقیق ارزیابی مدل با استفاده از چندین روش مورد بررسی قرار گرفت و فعالیت ضد HIV ترکیبات مورد مطالعه پیش‌بینی شد.

#### - ارزیابی مدل با استفاده از سری ارزیابی

به این منظور، شبکه عصبی با ۵۳ ترکیب سری آموزش، آموزش داده شد. معیار برای بهینه‌سازی شبکه، حداقل MSE سری ارزیابی (۱۰ ترکیب) می‌باشد. نتایج پیش‌بینی مدل بهینه در جدول (۴-۳۰) آورده شده است. ترسیمی از فعالیت پیش‌بینی شده بر حسب فعالیت تجربی نیز در شکل (۴-۱۸) نشان داده شده است.

جدول (۴-۳۰) نتایج حاصل از ارزیابی مدل شبکه عصبی بهینه با استفاده از داده‌های سری ارزیابی

شماره ترکیب	pEC <sub>50</sub>		در صد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۳	۵/۸۹	۶/۱۲	۳/۹۰
۷	۴/۷۲	۴/۶۲	-۲/۱۲
۱۹	۵/۳۹	۵/۵۱	۲/۲۳
۲۹	۵/۲۹	۵/۳۱	۰/۳۸
۷۲	۶/۵۲	۶/۷۰	۲/۷۶
۳۲	۷/۳۰	۷/۲۰	-۱/۳۷
۴۱	۶/۵۲	۶/۵۸	۰/۹۲
۵۴	۷/۰۴	۷/۴۹	۶/۳۹
۶۱	۷/۳۲	۷/۴۱	۱/۲۳
۶۴	۷/۰۵	۷/۴۱	۵/۱۱



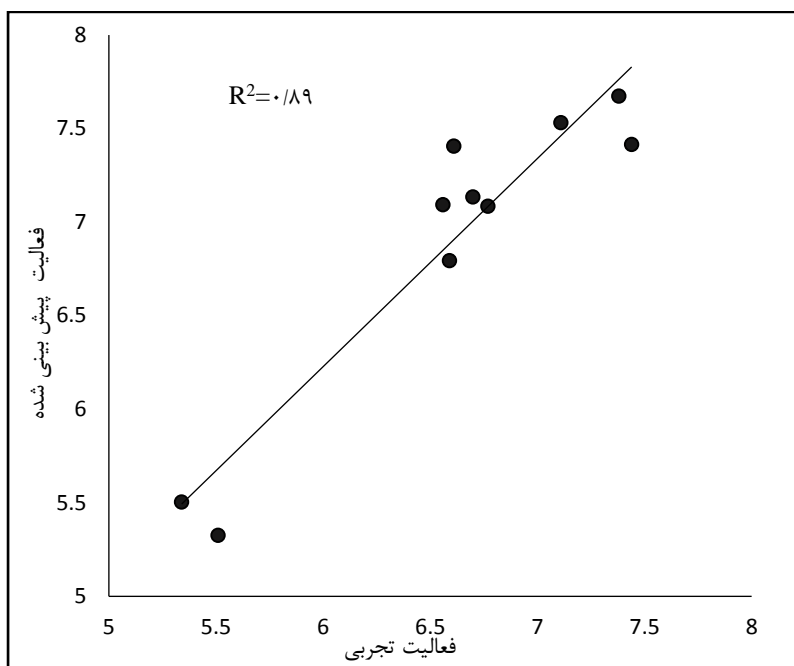
شکل ( ۴-۱۸ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی

### - ارزیابی مدل با استفاده از سری آزمون

بعد از آموزش شبکه و انتخاب شبکه عصبی بهینه با استفاده از سری ارزیابی قدرت پیش‌بینی مدل، به وسیله ۱۰ ترکیب که در مرحله آموزش و ارزیابی استفاده نشده‌اند، مورد بررسی قرار گرفت. نتایج حاصل در جدول ( ۴-۳۱ ) جمع آوری شده است. همچنین شکل ( ۴-۱۹ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی داده‌های سری آزمون را نشان می‌دهد.

جدول ( ۴-۳۱ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون

شماره ترکیب	pEC <sub>50</sub>		در صد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۲	۶/۵۹	۶/۷۹	۳/۰۳
۱۲	۵/۳۴	۵/۵۰	۳/۰۰
۱۶	۵/۵۱	۵/۳۳	-۳/۲۷
۲۴	۶/۷۷	۷/۰۹	۴/۷۳
۳۶	۶/۷۰	۷/۱۳	۶/۴۲
۴۵	۶/۶۱	۷/۴۱	۱۲/۱۰
۴۷	۷/۴۴	۷/۴۱	-۰/۴۰
۵۸	۷/۱۱	۷/۵۳	۵/۹۱
۶۳	۷/۳۸	۷/۶۷	۳/۹۳
۶۸	۶/۵۶	۷/۰۹	۸/۰۸



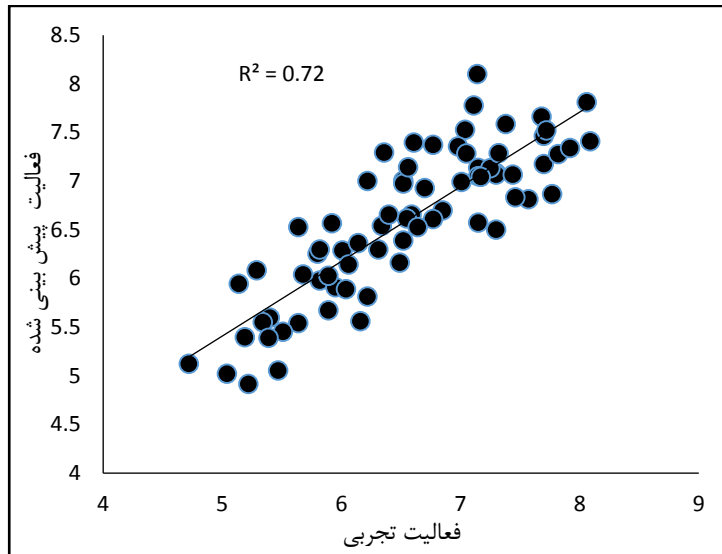
شکل ( ۴-۱۹) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری آزمون

#### - ارزیابی مدل به روش رد مرحله‌ای تک تک

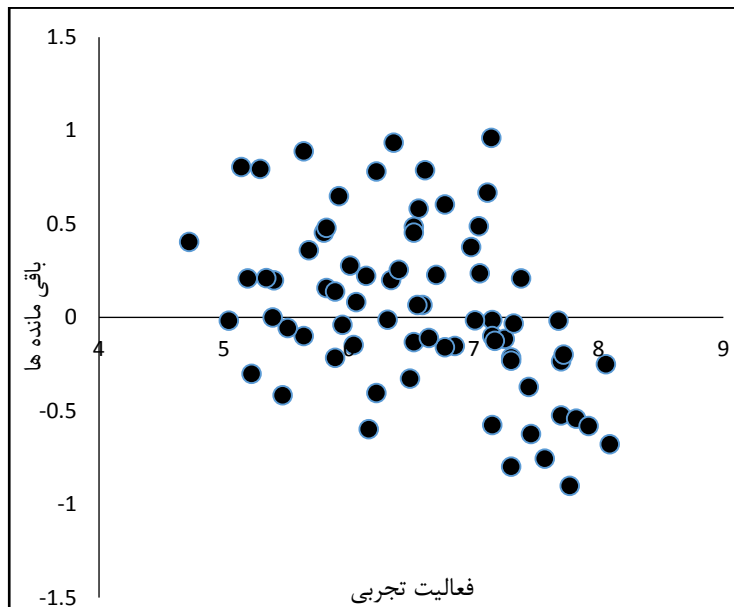
برای ارزیابی بیشتر شبکه عصبی بهینه، روش رد مرحله‌ای تک تک مورد استفاده قرار گرفت. در این روش هر بار یک ترکیب به‌عنوان سری آزمون کنار گذاشته شد و در شرایط بهینه به‌دست آمده از بخش (۳-۴-۲-۴) با ۷۲ ترکیب باقی‌مانده مدل سازی صورت گرفت سپس مدل به‌دست آمده برای پیش‌بینی  $pEC_{50}$  ترکیب کنار گذاشته شده به کار گرفته شد و این فرآیند برای تک تک ترکیبات تکرار گردید. نتایج حاصل از این روش در جدول ( ۴-۳۲) آورده شده است. شکل ( ۴-۲۰) فعالیت پیش‌بینی شده بر حسب فعالیت تجربی برای کل داده‌ها را نشان می‌دهد. شکل ( ۴-۲۱) از رسم باقی‌مانده فعالیت پیش‌بینی شده از فعالیت تجربی بر حسب فعالیت تجربی برای کل مولکول‌ها به‌دست آمد. نتایج بیانگر خطای پیش‌بینی کم و قدرت تعمیم‌پذیری بالای مدل می‌باشد. لذا با این مدل می‌توان فعالیت ضد HIV را برای هر ترکیب جدید با خطای قابل قبولی پیش‌بینی کرد. با توجه به نمودار باقی‌مانده‌ها، توزیع نسبتاً یکنواخت و اتفاقی مقادیر در دو طرف خط صفر نشان دهنده‌ی عدم وجود خطای سیستماتیک مدل می‌باشد.

جدول ( ۴-۳۲ ) نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از روش رد مرحله‌ای تک تک برای کل داده‌ها

شماره ترکیب	pEC <sub>50</sub>			شماره ترکیب	pEC <sub>50</sub>		
	فعالیت تجربی	فعالیت پیش‌بینی شده	درصد خطا		فعالیت تجربی	فعالیت پیش‌بینی شده	درصد خطا
۱	۵/۱۹	۵/۴۰	۴/۰۵	۳۸	۵/۹۲	۶/۵۷	۱۰/۹۸
۲	۶/۵۹	۶/۶۶	۱/۰۶	۳۹	۶/۴۰	۶/۶۶	۴/۰۶
۳	۵/۸۹	۵/۶۷	-۳/۷۴	۴۰	۶/۲۲	۵/۸۱	-۶/۵۹
۴	۵/۰۴	۵/۰۲	-۰/۴۰	۴۱	۶/۵۲	۶/۹۷	۶/۹۰
۵	۶/۴۹	۶/۱۶	-۵/۰۸	۴۲	۷/۳۰	۶/۵۰	-۱۰/۹۶
۶	۵/۶۴	۵/۵۴	-۱/۷۷	۴۳	۷/۷۰	۷/۱۷	-۶/۸۸
۷	۴/۷۲	۵/۱۲	۸/۴۷	۴۴	۵/۶۸	۶/۰۴	۶/۳۴
۸	۵/۱۴	۵/۹۴	۱۵/۵۶	۴۵	۶/۶۱	۷/۴۰	۱۱/۹۵
۹	۶/۰۱	۶/۲۹	۴/۶۶	۴۶	۷/۷۷	۶/۸۷	-۱۱/۵۸
۱۰	۵/۴۰	۵/۶۰	۳/۷۰	۴۷	۷/۴۴	۷/۰۷	-۴/۹۷
۱۱	۶/۰۶	۶/۱۴	۱/۳۲	۴۸	۸/۰۹	۷/۴۱	-۸/۴۱
۱۲	۵/۳۴	۵/۵۵	۳/۹۳	۴۹	۷/۵۷	۶/۸۱	-۱۰/۰۴
۱۳	۵/۸۲	۵/۹۸	۲/۷۵	۵۰	۷/۸۲	۷/۲۸	-۶/۹۱
۱۴	۵/۸۰	۶/۲۵	۷/۷۶	۵۱	۷/۴۶	۶/۸۳	-۸/۴۵
۱۵	۵/۹۵	۵/۹۱	-۰/۶۷	۵۲	۷/۶۸	۷/۶۶	-۰/۲۶
۱۶	۵/۵۱	۵/۴۵	-۱/۰۹	۵۳	۸/۰۶	۷/۸۱	-۳/۱۰
۱۷	۶/۳۴	۶/۵۴	۳/۱۵	۵۴	۷/۰۴	۷/۵۳	۶/۹۶
۱۸	۶/۰۴	۵/۸۹	-۲/۴۸	۵۵	۷/۹۲	۷/۳۴	-۷/۳۲
۱۹	۵/۳۹	۵/۳۹	-۰/۰۰	۵۶	۷/۷۰	۷/۴۷	-۲/۹۹
۲۰	۷/۱۵	۶/۵۷	-۸/۱۱	۵۷	۷/۷۲	۷/۵۲	-۲/۵۹
۲۱	۶/۸۵	۶/۷۰	-۲/۱۹	۵۸	۷/۱۱	۷/۷۸	۹/۴۲
۲۲	۶/۷۷	۶/۶۱	-۲/۳۶	۵۹	۷/۲۵	۷/۱۳	-۱/۶۶
۲۳	۶/۱۴	۶/۳۶	۳/۵۸	۶۰	۶/۹۸	۷/۳۶	۵/۴۴
۲۴	۶/۷۷	۷/۳۷	۸/۸۶	۶۱	۷/۳۲	۷/۲۹	-۰/۴۱
۲۵	۶/۵۵	۶/۶۲	۱/۰۷	۶۲	۷/۱۴	۸/۱۰	۱۳/۴۵
۲۶	۶/۱۶	۵/۵۶	-۹/۷۴	۶۳	۷/۳۸	۷/۵۹	۲/۸۵
۲۷	۶/۳۱	۶/۳۰	-۰/۱۶	۶۴	۷/۰۵	۷/۲۹	۳/۴۰
۲۸	۵/۲۲	۴/۹۲	-۵/۷۵	۶۵	۷/۱۷	۷/۰۴	-۱/۸۱
۲۹	۵/۲۹	۶/۰۸	۱۴/۹۳	۶۶	۷/۰۱	۶/۹۹	-۰/۲۹
۳۰	۵/۸۹	۶/۰۳	۲/۳۸	۶۷	۶/۳۶	۷/۲۹	۱۴/۶۲
۳۱	۵/۴۷	۵/۰۵	-۷/۶۸	۶۸	۶/۵۶	۷/۱۴	۸/۸۴
۳۲	۷/۳۰	۷/۰۸	-۳/۰۱	۶۹	۵/۶۴	۶/۵۳	۱۵/۷۸
۳۳	۶/۲۲	۷/۰۰	۱۲/۵۴	۷۰	۵/۸۲	۶/۳۰	۸/۲۵
۳۴	۷/۳۰	۷/۰۷	-۳/۱۵	۷۱	۷/۱۵	۷/۱۴	-۰/۱۴
۳۵	۷/۱۵	۷/۰۵	-۱/۴۰	۷۲	۶/۵۲	۷/۰۰	۷/۳۶
۳۶	۶/۷۰	۶/۹۳	۳/۴۳	۷۳	۶/۵۲	۶/۳۹	-۱/۹۹
۳۷	۶/۶۴	۶/۵۳	-۱/۶۶				



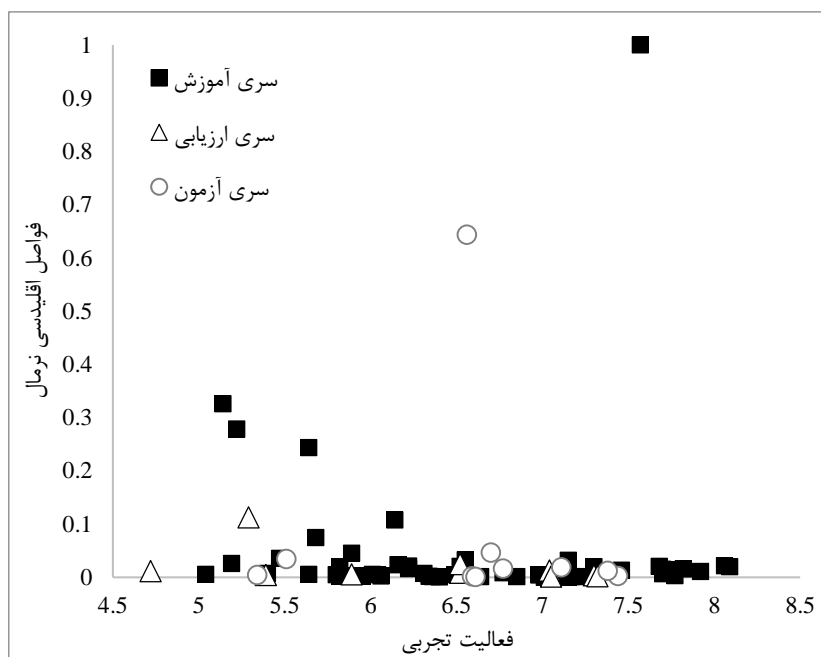
شکل ( ۴-۲۰ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی به روش رد مرحله‌ای تک تک برای کل داده‌ها



شکل ( ۴-۲۱ ) نمودار باقی مانده بر حسب فعالیت تجربی برای کل داده‌ها

### - ارزیابی مدل با استفاده از آزمون پراکندگی

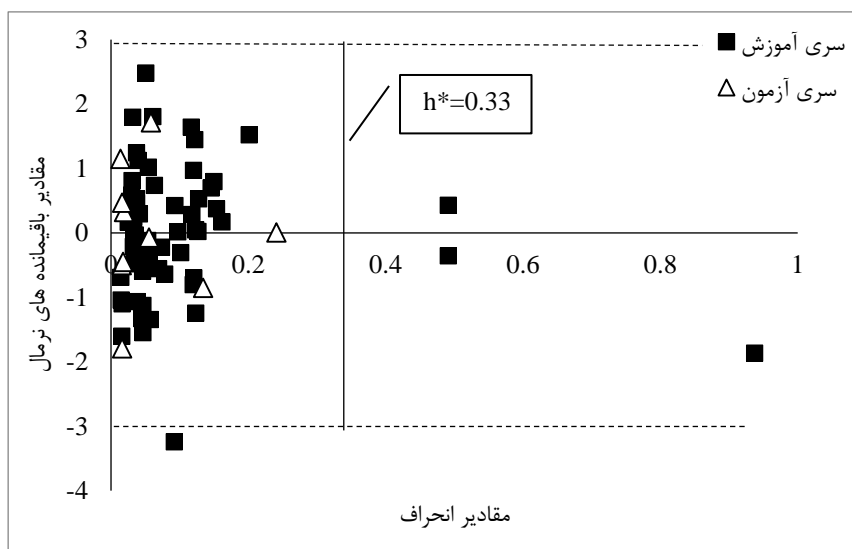
برای ایجاد بهترین تقسیم بندی در داده‌های سری آموزش، ارزیابی و آزمون، فواصل اقلیدسی همه ترکیبات با استفاده از توصیف کننده‌های مدل برتر، مطابق رابطه ( ۲-۱۷ ) محاسبه شد. نمودار پراکندگی (شکل ( ۴-۲۲ )) از رسم فواصل اقلیدسی نرمال شده بین ۰ و ۱ بر حسب فعالیت تجربی به دست آمد. شکل ( ۴-۲۲ ) توزیع مناسب داده‌های سری آزمون و ارزیابی را در محدوده داده‌های سری آموزش نشان می‌دهد.



شکل (۴-۲۲) نمودار پراکندگی برای تقسیم بندی ترکیبات DAPY-like

#### - ارزیابی مدل با استفاده از قلمرو کاربرد

قلمرو کاربرد مدل برتر این مطالعه، برای تعیین وجود داده‌های دور افتاده به کار گرفته شد. نمودار ویلیامز از رسم مقادیر باقیمانده استاندارد شده فعالیت پیش‌بینی شده ترکیبات DAPY-like بر حسب مقادیر حد انحراف به دست آمد. مطابق شکل (۴-۲۳)، از بین همه ترکیبات، فقط ۳ ترکیب از سری آموزش دارای مقادیر انحراف بزرگتر از حد انحراف آستانه ( $h^* = 0.33$ ) هستند و ۱ ترکیب نیز از ۳- برابر انحراف استاندارد باقیمانده‌ها کوچکتر شده است. با توجه به اینکه حدود ۹۵٪ از کل داده‌ها در قلمرو کاربرد مدل برتر حضور دارند، بنابراین اعتبار و قدرت پیش‌بینی مدل اثبات می‌شود.



شکل ( ۴-۲۳ ) نمودار قلمرو کاربرد برای مدل شبکه عصبی با ترکیبات DAPY-like

#### ۴-۲-۵ نتیجه‌گیری مدل سازی با مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

در این مطالعه برای پیش‌بینی فعالیت بازدارندگی ۷۳ مشتق DAPY-like، از مخلوط توصیف کننده‌های داکینگ مولکولی و شمارش گروه عاملی در مدل سازی QSAR به روش شبکه عصبی مصنوعی استفاده شد. مقادیر  $R^2$  و MSE حاصل از پیش‌بینی فعالیت ضد ایدز سری آزمون با استفاده از مدل توسعه یافته با توصیف کننده‌های داکینگ مولکولی و مخلوط توصیف کننده‌های داکینگ مولکولی و ساختاری در جدول ( ۴-۳۰ ) خلاصه شده است. مقایسه نتایج نشان می‌دهد که با افزودن توصیف کننده‌های شمارش گروه عاملی ترکیبات مورد مطالعه به توصیف کننده‌های داکینگ مولکولی، قدرت مدل QSAR افزایش می‌یابد.

جدول ( ۴-۳۳ ) مقایسه مدل های شبکه عصبی توسعه یافته با توصیف کننده‌های متفاوت در ارزیابی مدل با استفاده از سری آزمون

پارامتر آماری	مدل توسعه یافته با توصیف کننده های	
	مدل توسعه یافته با توصیف کننده های داکینگ مولکولی	شمارش گروه عاملی و داکینگ مولکولی
$R^2$	۰/۵۰	۰/۸۹
MSE	۰/۲۴	۰/۱۶

## ۴-۲-۴ ارزیابی مدل با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با

### استفاده از پارامترهای آماری

برای ارزیابی کارایی شبکه عصبی مصنوعی ارائه شده برای پیش‌بینی فعالیت بازدارندگی ترکیبات مورد مطالعه، برخی از پارامترهای آماری محاسبه شد و قدرت پیش‌بینی مدل با استفاده از محاسبه این پارامترها برای سری ارزیابی و سری آزمون مورد بررسی قرار گرفت و نتایج پارامترهای آماری به همراه محدوده قابل قبول برای هر پارامتر در جدول (۴-۳۴) آورده شده است. این نتایج حاکی از قدرت پیش‌بینی بالای مدل غیرخطی شبکه‌ی عصبی با مجموع توصیف کننده‌های داکینگ مولکولی-ساختاری می‌باشد.

جدول (۴-۳۴) پارامترهای آماری برای مدل ANN با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

ردیف	پارامتر آماری	سری ارزیابی	سری آزمون	محدوده قابل قبول
۱	PRESS	۰/۴۶	۱/۵۷	-
۲	SEP	۰/۲۱	۰/۴۰	-
۳	MAE	۰/۱۷	۰/۳۴	-
۴	REP(%)	۳/۴۱	۶/۰۰	-
۵	MSE	۰/۰۵	۰/۱۶	-
۶	MRE	۲/۶۴	۵/۱۱	-
۷	R <sup>2</sup>	۰/۹۷	۰/۸۹	>۰/۶
۸	R <sub>0</sub> <sup>2</sup>	۰/۹۷	۰/۸۹	نزدیک به R <sup>2</sup>
۹	R <sub>0,R</sub> <sup>2</sup>	۰/۰۱	۰/۰۰	<۰/۱
۱۰	R <sub>m</sub> <sup>2</sup>	۰/۹۳	۰/۸۴	>۰/۵
۱۱	R <sub>0</sub> ' <sup>2</sup>	۰/۹۷	۰/۸۶	نزدیک به R <sup>2</sup>
۱۲	R <sub>0,R</sub> ' <sup>2</sup>	۰/۰۱	۰/۰۴	<۰/۱
۱۳	R <sub>m</sub> ' <sup>2</sup>	۰/۹۰	۰/۷۳	>۰/۵
۱۴	ΔR <sub>m</sub> <sup>2</sup>	۰/۰۳	۰/۱۱	<۰/۲
۱۵	k	۱/۰۲	۱/۰۵	۰/۸۵ ≤ k ≤ ۱/۱۵
۱۶	k'	۰/۹۸	۰/۹۶	۰/۸۵ ≤ k ≤ ۱/۱۵

### ۴-۲-۴-۷ آزمون Y - تصادفی

برای حصول اطمینان از تصادفی نبودن نتایج، ۱۰ مجموعه از ۷۳ عدد تصادفی در دامنه‌ی تغییرات مقادیر فعالیت دارویی تعریف شد. مدل QSAR جدید با استفاده از ماتریس توصیف کننده‌های



اصلی و مقادیر تصادفی از  $pEC_{50}$  توسعه یافت. نتایج حاصل از مدل سازی شبکه عصبی با استفاده از ده دسته توصیف کننده تصادفی در جدول (۴-۳۵) آورده شده است. مقادیر کوچک ضریب تعیین ( $R^2$ ) بیانگر عدم وجود ارتباط تصادفی یا وابستگی ساختاری به سری آموزش در مدل توسعه یافته شبکه عصبی مصنوعی می باشد.

جدول (۴-۳۵) مقادیر  $R^2$  برای سری آزمون با استفاده از آزمون  $Y$ -تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
$R^2$	۰/۱۴	۰/۰۷	۰/۰۲	۰/۰۰	۰/۰۰	۰/۰۱	۰/۰۳	۰/۰۴	۰/۰۹	۰/۰۹

### ۳-۴ مدل سازی فعالیت دارویی مشتقات آزینی به عنوان بازدارنده های غیرنوکلئوزیدی

#### ویروس HIV

در این مطالعه، ارتباط کمی ساختار- فعالیت ضد ایدز ۷۳ ترکیب از مشتقات آزینی به عنوان بازدارنده های غیرنوکلئوزیدی نسخه بردار معکوس مورد بررسی قرار گرفت. ترکیبات مورد مطالعه در جایگاه فعال دو پروتئین با کد 3M8Q و 3MEC (به ترتیب برای مولکول های ۱-۴۳ و ۴۴-۷۳) داک شدند و اطلاعات برهم کنش لیگاند - پروتئین با استفاده از بهترین پیکربندی ترکیبات در جایگاه فعال پروتئین، به دست آمد و به عنوان توصیف کننده های داکینگ مولکولی ذخیره شد. موثرترین توصیف کننده ها با استفاده از روش رگرسیون گام به گام انتخاب و به عنوان ورودی شبکه عصبی مصنوعی برای ساخت مدل QSAR استفاده شدند. سپس به منظور بهبود مدل شبکه عصبی مصنوعی، توصیف کننده های شاخص های شبه دارویی<sup>۱</sup> از زیر مجموعه توصیف کننده های ویژگی مولکولی<sup>۲</sup>، با استفاده از نرم افزار DRAGON 5.5 محاسبه شد. موثرترین توصیف کننده ها با استفاده از روش رگرسیون گام به گام به دست آمد. مدل شبکه عصبی با توصیف کننده های منتخب توسعه یافت و کارایی مدل توسعه یافته مورد ارزیابی قرار گرفت. در نهایت، نتایج به دست آمده، پیش بینی و تعمیم پذیری بالای مدل

<sup>۱</sup> Drug like indexes

<sup>۲</sup> Molecular properties

توسعه یافته را نشان می‌دهد.

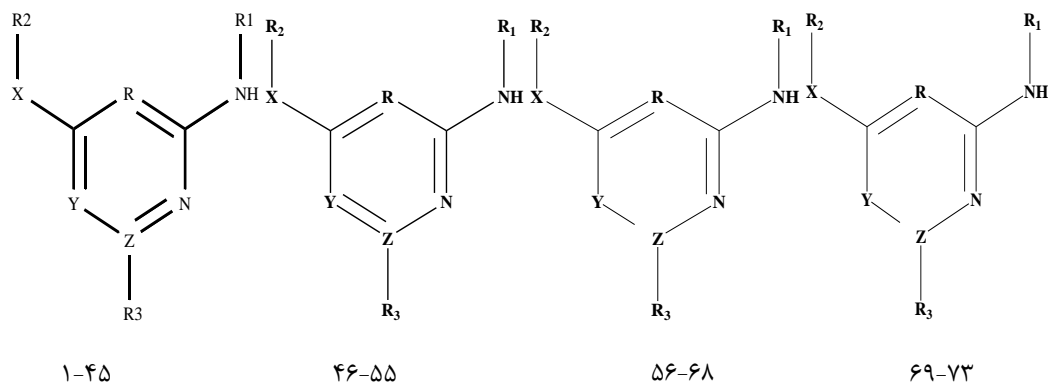
### ۱-۳-۴ مجموعه داده‌ها

مجموعه داده‌ها شامل فعالیت دارویی ۷۳ بازدارنده HIV-1 است [۱۰۵-۱۰۱]. ساختار پایه

ترکیبات در شکل (۴-۲۴) و مقادیر عددی  $pEC_{50}$  متناظر با هر ترکیب و جزئیات استخلاف‌ها در جدول

(۴-۳۶) آورده شده است.  $EC_{50}$  غلظتی از بازدارنده (دارو) می‌باشد که ۵۰٪ از تکثیر ویروس HIV-1

جلوگیری می‌کند.



شکل (۴-۲۴) اسکلت‌های اصلی ترکیبات مورد مطالعه

جدول ( ٤-٣٦ ) ساختار و مقادير pEC<sub>50</sub> تركيبات مورد مطالعه

شماره تركيب	X	Y	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pE <sub>50</sub>
١	O	N	piperidine	2,4,6-TriMe-Ph	-NH-4-CN-Ph	٦/٨٥
٢ <sup>١</sup>	O	N	piperidine	2,6-DiMe-4-CN-Ph	-NH-4-CN-Ph	٧/٤٢
٣ <sup>٧</sup>	O	N	p-1-yl- 4-SO <sub>2</sub> NH <sub>2</sub> -Ph	2,4,6-TriMe-Ph	-NH-4-CN-Ph	٦/٧٢
٤	O	N	p-1-yl -4-SO <sub>2</sub> CH <sub>3</sub> -Ph	2,4,6-TriMe-Ph	-NH-4-CN-Ph	٦/٨٩
٥	O	N	p-1-yl - 4-Pyridyl	2,4,6-TriMe-Ph	-NH-4-CN-Ph	٦/٦٢
٦	O	N	p-1-yl - 4-SO <sub>2</sub> NH <sub>2</sub> -Ph	2,6-DiMe-4-CN-Ph	-NH-4-CN-Ph	٦/٨٩
٧ <sup>١</sup>	O	N	p-1-yl - 4-SO <sub>2</sub> CH <sub>3</sub> -Ph	2,6-DiMe-4-CN-Ph	-NH-4-CN-Ph	٧/٢٤
٨ <sup>٧</sup>	O	N	4-Pyridyl-p-1-yl	2,6-DiMe-4-CN-Ph	-NH-4-CN-Ph	٧/٣٣
٩	NH	N	4-SO <sub>2</sub> NH <sub>2</sub> -Ph -p-1-yl	2,4,6-TriMe-Ph	-NH-4-CN-Ph	٧/١١
١٠	NH	N	p-1-yl - Pyridin-4-yl	2,4,6-TriMe-Ph	NH <sub>2</sub>	٨/٣١
١١ <sup>١</sup>	NH	N	p-1-yl - 4-COOEt-Ph	2,4,6-TriMe-Ph	NH <sub>2</sub>	٧/٩٤
١٢	NH	N	p-1-yl - 4-CONH <sub>2</sub> -Ph	2,4,6-TriMe-Ph	NH <sub>2</sub>	٧/٦٩
١٣	NH	N	p-1-yl - 4-SO <sub>2</sub> NH <sub>2</sub> -Ph	2,4,6-TriMe-Ph	NH <sub>2</sub>	٨/٢٤
١٤	NH	N	p-1-yl- 4-SO <sub>2</sub> Me-Ph	2,4,6-TriMe-Ph	NH <sub>2</sub>	٨/٢٦
١٥	NH	N	p-1-yl- Phenyl	2,4,6-TriMe-Ph	NH <sub>2</sub>	٧/٩٩
١٦	NH	N	p-1-yl- 4-NO <sub>2</sub> -Ph	2,4,6-TriMe-Ph	NH <sub>2</sub>	٨/١٠
١٧	NH	N	p-1-yl- 4-CN-Ph	2,4,6-TriMe-Ph	NH <sub>2</sub>	٨/٢٥
١٨ <sup>٧</sup>	NH	N	p-1-yl - Pyridin-4-yl	2,4,6-TriMe-Ph	NHMe	٨/١٤
١٩	NH	N	p-1-yl - 4-COOEt-Ph	2,4,6-TriMe-Ph	NHMe	٧/٧٨
٢٠	NH	N	p-1-yl - 4-CONH <sub>2</sub> -Ph	2,4,6-TriMe-Ph	NHMe	٨/٣٤
٢١ <sup>١</sup>	NH	N	p-1-yl - 4-SO <sub>2</sub> NH <sub>2</sub> -Ph	2,4,6-TriMe-Ph	NHMe	٨/١٦
٢٢ <sup>١</sup>	NH	N	p-1-yl -4-SO <sub>2</sub> Me-Ph	2,4,6-TriMe-Ph	NHMe	٨/٢٢
٢٣	NH	N	p-1-yl - Pyridin-4-yl	2,4,6-TriMe-Ph	OMe	٧/٨٨
٢٤ <sup>٧</sup>	NH	N	p-1-yl - 4-COOEt-Ph	2,4,6-TriMe-Ph	OMe	٧/٨٣
٢٥ <sup>٧</sup>	NH	N	p-1-yl - 4-CONH <sub>2</sub> -Ph	2,4,6-TriMe-Ph	OMe	٨/٠٦
٢٦	NH	N	p-1-yl - 4-SO <sub>2</sub> NH <sub>2</sub> -Ph	2,4,6-TriMe-Ph	OMe	٧/٩٧
٢٧	NH	N	p-1-yl - 4-SO <sub>2</sub> Me-Ph	2,4,6-TriMe-Ph	OMe	٧/٩٢
٢٨	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiMe-Ph	H	٦/٣٤
٢٩	O	-C-NO <sub>2</sub>	4-CN-Ph	2,4,6-TriMe-Ph	H	٧/٢٥
٣٠	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiMe-4-CN-Ph	H	٦/٩٦
٣١ <sup>٧</sup>	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiMe-4-Br-Ph	H	٦/٥٧
٣٢	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiMe-4-Cl-Ph	H	٦/٨٠
٣٣	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-Di-OMe-Ph	H	٥/٣٧
٣٤ <sup>١</sup>	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiCl-Ph	H	٦/٢١
٣٥ <sup>٧</sup>	O	-C-NO <sub>2</sub>	4-CN-Ph	2,4,6-TriCl-Ph	H	٦/٧٧
٣٦	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiCl-4-NO <sub>2</sub> -Ph	H	٤/٧٧
٣٧	O	-C-NO <sub>2</sub>	4-CN-Ph	2,4,6-TriBr-Ph	H	٦/٩٦
٣٨	O	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiBr-4-Me-Ph	H	٧/٤٧
٣٩	O	-C-NO <sub>2</sub>	4-CN-Ph	2,4,6-TriF-Ph	H	٦/١٤
٤٠	NH	-C-NO <sub>2</sub>	4-CN-Ph	2,4,6-TriMe-Ph	H	٤/٥٣
٤١	NH	-C-NO <sub>2</sub>	4-CN-Ph	2,6-DiBr-4-Me-Ph	H	٥/٤١
٤٢	NH	-C-NO <sub>2</sub>	4-CH <sub>3</sub> -Ph	2,4-DiBr-6-F-Ph	H	٥/٥٢
٤٣	O	-C-NO <sub>2</sub>	4-NO <sub>2</sub> -Ph	2,6-DiMe-4-CN-Ph	H	٦/١٤
٤٤	O	C	4-CN-Ph	2,4,6-TriMe-Ph	-OMe	٦/٠٨
٤٥	O	C	4-CN-Ph	2,6-DiMe-4-CN	-OMe	٦/١٥
٤٦ <sup>٧</sup>	O	-C-Cl	p-Cl-Ph	2,4,6-TriMe-Ph	-	٦/٨٦
٤٧ <sup>١</sup>	O	-C-Cl	p-Me-Ph	2,4,6-TriMe-Ph	-	٦/٧٠
٤٨	O	-C-Cl	p-CN-Ph	2,4,6-TriMe-Ph	-	٧/١٠

ادامه جدول ( ۴-۳۶ )

شماره ترکیب	X	Y	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pE <sub>50</sub>
۴۹	O	-C-Cl	p-CN-Ph	2,4,6-TriCl-Ph	-	۶/۸۱
۵۰	O	-C-Cl	p-CN-Ph	2,4,6-TriBr-Ph	-	۶/۷۴
۵۱	O	-C-Cl	p-CN-Ph	2,6-DiBr-4-Me-Ph	-	۷/۴۷
۵۲	O	-C-Cl	p-CN-Ph	2,6-DiMe-4-Br-Ph	-	۶/۴۰
۵۳	O	-C-Cl	p-CN-Ph	2,6-DiCl-Ph	-	۶/۶۶
۵۴	O	-C-Cl	p-CN-Ph	4-CN-2,6-DiMe-Ph	-	۶/۸۴
۵۵	O	-C-Cl	p-CN-Ph	2,6-DiMe-Ph	-	۶/۳۷
۵۶	O	-C=O	p-Cl-Ph	2,4,6-TriMe-Ph	H	۶/۱۱
۵۷ <sup>t</sup>	O	-C=O	p-Me-Ph	2,4,6-TriMe-Ph	H	۵/۸۱
۵۸	O	-C=O	p-NO <sub>2</sub> -Ph	2,4,6-TriMe-Ph	H	۵/۳۹
۵۹	O	-C=O	p-OMe-Ph	2,4,6-TriMe-Ph	H	۵/۸۰
۶۰ <sup>v</sup>	O	-C=O	p-CN-Ph	2,4,6-TriMe-Ph	H	۶/۵۹
۶۱ <sup>t</sup>	O	-C=O	p-CN-Ph	2,4,6-TriBr-Ph	H	۶/۲۰
۶۲	O	-C=O	p-CN-Ph	2,6-DiBr-4-Me-Ph	H	۶/۶۸
۶۳	O	-C=O	p-CN-Ph	2,6-DiMe-4-Br-Ph	H	۵/۹۲
۶۴	O	-C=O	m-Cl-Ph	2,4,6-TriMe-Ph	H	۴/۸۴
۶۵	O	-C=O	p-CN-Ph	2,6-DiMe-Ph	H	۵/۷۲
۶۶	O	-C=O	p-CN-Ph	4-CN-2,6-DiMe-Ph	H	۵/۵۵
۶۷	O	-C=O	p-CN-Ph	2,4,6-TriMe-Ph	Me	۵/۹۴
۶۸*	O	-C=O	p-CN-Ph	2,4,6-TriMe-Ph	Me	۵/۴۷
۶۹ <sup>v</sup>	O	C	4NO <sub>2</sub> -ph	2,4,6-TriMe-Ph	=O	۵/۸۵
۷۰	O	C	4-Me-Ph	2,4,6-TriMe-Ph	=O	۵/۸۵
۷۱	O	C	4-CN-Ph	2,4,6-TriMe-Ph	=O	۶/۴۳
۷۲	O	C	4-CN-Ph	2,6-DiMe-4-CN-Ph	=O	۶/۸۲
۷۳ <sup>t</sup>	O	C	4-CN-Ph	2,6-DiMe-Ph	=O	۵/۸۵

\* لیگاند ۶۸ در اسکلت اصلی به جای NH، NMe دارد. R در همه ساختارها نشان‌دهنده کربن است به جز ساختارهای ۲۷-۱۰ که نیتروژن دارند. Z در همه ساختارها نماینده کربن است به جز ساختارهای ۴۶-۶۸ که نماینده نیتروژن می‌باشد. (p-1-yl) بیانگر piperidine-1-yl- methyl می‌باشد. (-) به معنای نبود استخلاف در R<sub>3</sub> می‌باشد. t نمایانگر داده‌های سری آزمون، v نمایانگر داده‌های سری ارزیابی می‌باشند و سایر داده‌ها در سری آموزش قرار دارند.

## ۲-۳-۴ رسم و بهینه سازی ساختار ترکیبات مورد مطالعه

پس از جمع آوری مجموعه داده‌ها، ساختار ترکیبات مورد مطالعه مطابق بخش (۲-۳) رسم، بهینه سازی و با فرمت \*.mol و \*.hin ذخیره شدند.

## ۳-۳-۴ مدل سازی فعالیت دارویی ترکیبات مورد مطالعه با استفاده از توصیف کننده‌های حاصل

### از داکینگ مولکولی

در راستای ساخت مدل QSAR برای مشتقات آزینی ضد ایدز، توصیف کننده‌های حاصل از اطلاعات برهمکنش لیگاندها-پروتئین با استفاده از نرم افزار Autodock4.2 محاسبه و مدل QSAR با استفاده از این توصیف کننده‌ها ساخته شد. در ادامه، مراحل استخراج توصیف کننده‌های داکینگ مولکولی و انتخاب مهم‌ترین آن‌ها شرح داده می‌شود.

### ۱-۲-۳-۴ استخراج توصیف کننده‌های حاصل از داکینگ مولکولی

به منظور استخراج توصیف کننده‌های داکینگ مولکولی، دو پروتئین 3M8Q و 3MEC از بانک اطلاعاتی پروتئین دریافت و با فرمت pdb ذخیره شدند. سپس استخراج، آماده سازی لیگاند کریستالوگرافی و تعیین مختصات مرکز ثقل لیگاند کریستالوگرافی، آماده سازی پروتئین و تعیین بهترین شرایط داکینگ مولکولی به روش اعتبارسنجی دقیقاً مطابق مراحل توضیح داده شده در بخش (۱-۳-۲-۴) انجام شد. همه لیگاندها در شرایط بهینه داکینگ به‌طور جداگانه در جایگاه فعال پروتئین مربوطه (مولکول‌های ۱-۴۳ در پروتئین 3M8Q و مولکول‌های ۷۳-۴۴ در پروتئین 3MEC) داک شدند. در بخش‌های زیر، مراحل استخراج توصیف کننده‌های داکینگ مولکولی، مورد بحث واقع شده است.

### - داکینگ لیگاندهای مورد مطالعه در پروتئین‌های مربوطه

جهت استخراج توصیف کننده‌های داکینگ مولکولی، پس از آماده سازی لیگاندهای مورد مطالعه عملیات داکینگ مولکولی لیگاندها در پروتئین‌های مربوطه انجام شد و ۱۰ توصیف کننده داکینگ مولکولی برای این ترکیبات استخراج گردید.

## ❖ آماده سازی لیگاندهای مورد مطالعه

ساختار ترکیبات مورد مطالعه، رسم، بهینه سازی و ذخیره سازی شد و سپس در نرم افزار ViewerLite5.0 فراخوانی و با پسوند pdb ذخیره شدند.

## ❖ اجرای عملیات داکینگ مولکولی لیگاندها و پروتئین

به منظور اجرای عملیات داکینگ لیگاندها در پروتئین، ساختار اصلاح شده‌ی پروتئین (پروتئین 3M8Q برای لیگاندهای ۴۳-۱ و پروتئین 3MEC برای لیگاندهای ۷۳-۴۴) در نرم افزار Autodock4.2 فراخوانی و آماده سازی بیشتر مطابق بخش‌های (۲-۳-۳-۳ و ۱-۳-۳-۳) بر روی پروتئین انجام گرفت. مختصات جایگاه فعال و شبکه مطابق بخش (۲-۳-۳-۳) روی پروتئین اعمال شد. سپس تمامی لیگاندها به صورت جداگانه در نرم افزار Autodock4.2 فراخوانی شدند. تعداد اجرای الگوریتم ژنتیک لامارکین در ۱۵۰ (مقدار بهینه مشخص شده در مرحله اعتبارسنجی) تنظیم و سایر پارامترهای الگوریتم ژنتیک لامارکین و گزینه‌های اجرای داکینگ در شرایط پیش فرض برنامه تنظیم شدند. در نهایت همه لیگاندها در پروتئین مربوطه داک و نتایج خروجی آن‌ها با فرمت dlg ذخیره شد.

## - استخراج توصیف‌کننده‌های حاصل از داکینگ مولکولی

برای ایجاد ارتباط ریاضی بین ساختار-فعالیت، به مقادیر عددی نیازمند است، به این منظور برهمکنش‌های بین لیگاندها با پروتئین مربوطه، با استفاده از داکینگ مولکولی آنالیز شد و مقادیر عددی ۱۰ توصیف‌کننده داکینگ مولکولی از اطلاعات خوشه با بیشترین تعداد صورتبندی و کمترین انرژی اتصال برای هر لیگاند مورد مطالعه، با استفاده از نرم افزار Notepad++ استخراج شد.

## ۲-۳-۳-۴ انتخاب مؤثرترین توصیف‌کننده‌های داکینگ مولکولی

مهم‌ترین مرحله در مطالعات QSAR انتخاب توصیف‌کننده‌هایی است که دارای بیشترین ارتباط با فعالیت دارویی ترکیبات مورد نظر هستند و اطلاعات صحیحی از عوامل مؤثر بر نحوه اتصال دارو به پروتئین را ارائه می‌دهند. بنابراین مطابق بخش (۴-۳) با به‌کارگیری نرم افزار SPSS و اجرای روش

رگرسیون گام به گام، ۳ توصیف کننده موثر انتخاب شد. نام توصیف کننده‌ها در جدول (۴-۳۷) خلاصه شده است. نتایج حاصل در جدول (۴-۳۸) عدم وجود همبستگی معنادار بین توصیف کننده‌ها را نشان می‌دهد.

جدول (۴-۳۷) توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

No	Symbol	Meaning
۱	EE	Electrostatic Energy
۲	TI	Final Total Internal Energy
۳	pK <sub>i</sub>	-Log(Estimated Inhibition Constant, K <sub>i</sub> )

جدول (۴-۳۸) ماتریس همبستگی توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

	EE	TI	pK <sub>i</sub>
EE	۱		
TI	۰/۷۸	۱	
pK <sub>i</sub>	-۰/۰۲	-۰/۳۵	۱

به منظور بررسی وجود پدیده‌ی هم‌خطی در بین توصیف کننده‌های انتخاب شده، عامل تورم واریانس (VIF) مربوط به هر توصیف کننده طبق رابطه (۲-۱۶) محاسبه شد که نتایج آن در جدول (۴-۳۹) خلاصه شده است. با توجه به مقادیر VIF کمتر از ۱۰، عدم وجود هم‌خطی در بین توصیف کننده‌ها اثبات می‌گردد.

جدول (۴-۳۹) مقادیر عامل تورم واریانس توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

توصیف کننده	EE	TI	pK <sub>i</sub>
VIF	۱/۳۳۱	۱/۳۲۳	۱/۰۱۵

### ۴-۳-۳-۴ مدل شبکه عصبی مصنوعی

در این مرحله برای یافتن رابطه غیر خطی بین توصیف کننده‌ها و فعالیت دارویی، از شبکه عصبی مصنوعی پیشرو با الگوریتم آموزشی پس انتشار استفاده شد. به منظور بهینه سازی پارامترهای شبکه، مجموعه داده‌ها به سه دسته آموزش (۵۳ ترکیب)، ارزیابی (۱۰ ترکیب) و آزمون (۱۰ ترکیب) تقسیم شد. شبکه با استفاده از سری آموزش، آموزش داده شد و معیار بهینه سازی آموزش شبکه، حداقل

MSE سری ارزیابی در نظر گرفته شد. به منظور آموزش شبکه و بهترین برازش بین داده‌ها از دو تابع آموزش لونیبرگ-مارکوات و تنظیم بیزی و دو تابع انتقال غیر خطی لگاریتم سیگموئیدی و تانژانت سیگموئیدی در لایه پنهان استفاده شد. برای بهینه سازی همزمان پارامترهای موثر، هر شبکه با تعداد ورودی برابر با ۳، تعداد گره بین ۲ تا ۱۰ با گام ۱ در لایه پنهان و تعداد دور آموزش بین ۲ تا ۵۰ با گام ۲ آموزش داده شد. براساس تعداد مختلف گره، دور آموزشی و توابع متفاوت آموزش و انتقال، بهترین شبکه‌ها به دست آمد که نتایج آن در جدول (۴-۴۰) مشاهده می‌شود. در نهایت شبکه با تابع آموزش لونیبرگ-مارکوات، تابع انتقال لگاریتم سیگموئیدی، تعداد ۳ ورودی، تعداد ۵ گره در لایه پنهان و ۵۰ دور آموزش به عنوان شبکه برتر انتخاب شد.

جدول (۴-۴۰) توابع و پارامترهای شبکه‌های بهینه با توصیف کننده‌های انتخاب شده از داکینگ مولکولی با روش SR

MSE	تعداد دور آموزش	تعداد گره	تابع انتقال	تابع آموزش	تعداد توصیف کننده
۰/۱۲	۲	۹	لگاریتم-سیگموئید	تنظیم بیزی	۳
۰/۱۱	۵۰	۵	لگاریتم-سیگموئید	لونیبرگ-مارکوات	۳
۰/۱۲	۱۸	۴	تانژانت-سیگموئید	تنظیم بیزی	۳
۰/۱۳	۴	۶	تانژانت-سیگموئید	لونیبرگ-مارکوات	۳

#### ۴-۳-۳-۴ ارزیابی مدل شبکه عصبی مصنوعی

اعتبار و اهمیت مدل پیشگو وقتی مشخص می‌گردد که فعالیت دارویی مولکول‌های مجهولی که در دسته‌ی آموزش موجود نیستند را پیشگویی کند. در ادامه به روش‌های به کار گرفته شده برای ارزیابی مدل منتخب اشاره می‌شود.

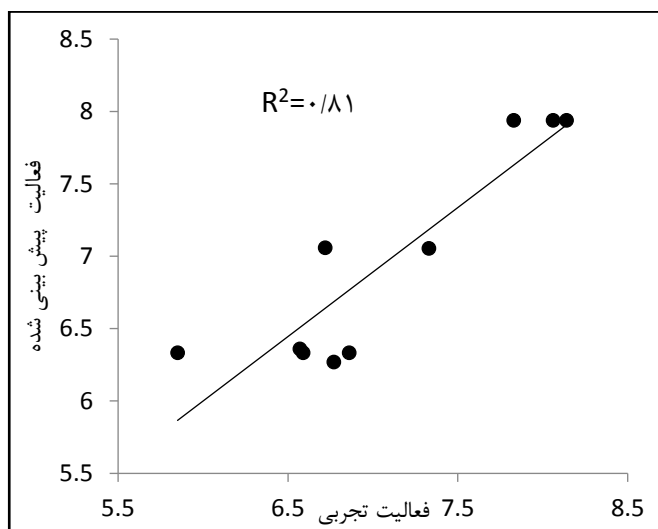
#### - ارزیابی مدل با استفاده از سری ارزیابی

به این منظور، شبکه عصبی با استفاده از ۵۳ مولکول سری آموزش، آموزش داده شد. معیار برای بهینه سازی شبکه کمترین مقدار MSE سری ارزیابی است. نتایج حاصل در جدول (۴-۴۱) و ترسیمی از مقادیر پیش بینی شده بر حسب مقادیر واقعی نیز در شکل (۴-۲۵) نشان داده شده است.



جدول (۴-۴) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری ارزیابی

شماره	pEC <sub>50</sub>		در صد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۳	۶/۷۲	۷/۰۶	۵/۰۶
۸	۷/۳۳	۷/۰۵	-۳/۸۲
۱۸	۸/۱۴	۷/۹۴	-۲/۴۶
۲۴	۷/۸۳	۷/۹۴	۱/۴۰
۲۵	۸/۰۶	۷/۹۴	-۱/۴۹
۳۱	۶/۵۷	۶/۳۶	-۳/۲۰
۳۵	۶/۷۷	۶/۲۷	-۷/۳۹
۴۶	۶/۸۶	۶/۳۳	-۷/۷۳
۶۰	۶/۵۹	۶/۳۳	-۳/۹۵
۶۹	۵/۸۵	۶/۳۳	۸/۲۱



شکل (۴-۲۵) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی

#### - ارزیابی مدل با استفاده از سری آزمون

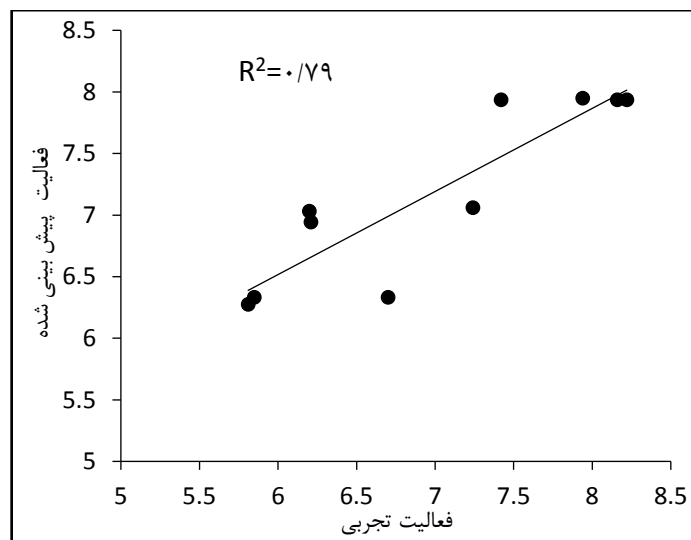
بعد از انتخاب شبکه عصبی بهینه با استفاده از سری ارزیابی، قدرت پیش‌بینی شبکه به وسیله

۱۰ مولکول سری آزمون که در آموزش استفاده نشده‌اند، مورد بررسی قرار گرفت. نتایج حاصل در جدول

(۴-۴۲) و همچنین شکل (۴-۲۶) نشان داده شده است

جدول ( ۴-۴۲ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون

شماره ترکیب	pEC <sub>50</sub>		درصد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۲	۷/۴۲	۷/۹۴	۷/۰۱
۷	۷/۲۴	۷/۰۶	-۲/۴۹
۱۱	۷/۹۴	۷/۹۵	۰/۱۳
۲۱	۸/۱۶	۷/۹۴	-۲/۷۰
۲۲	۸/۲۲	۷/۹۴	-۳/۴۱
۳۴	۶/۲۱	۶/۹۵	۱۱/۹۲
۴۷	۶/۷۰	۶/۳۳	-۵/۵۲
۵۷	۵/۸۱	۶/۲۸	۸/۰۹
۶۱	۶/۲۰	۷/۰۳	۱۳/۳۹
۷۳	۵/۸۵	۶/۳۳	۸/۲۱



شکل ( ۴-۲۶ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی

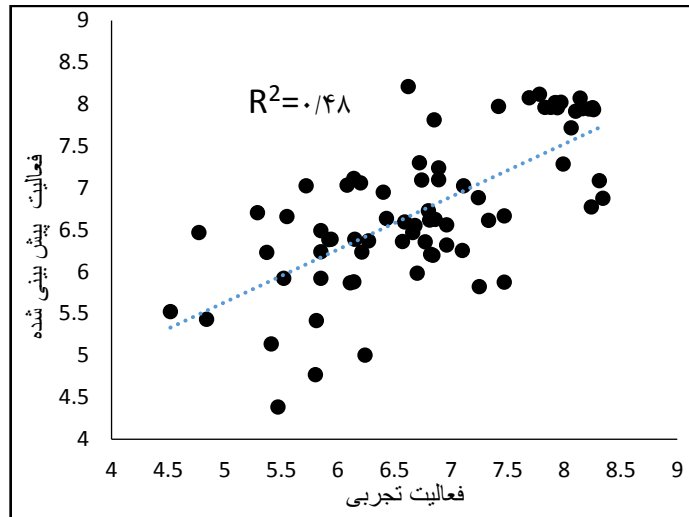
#### - ارزیابی مدل به روش رد مرحله‌ای تک تک

این ارزیابی مطابق با توضیحات بخش (۴-۲-۴) انجام شد و نتایج حاصل از ارزیابی شبکه با این روش در جدول (۴-۴۳) نشان داده شده است. شکل (۴-۲۷) هم فعالیت پیش‌بینی شده بر حسب فعالیت تجربی را نشان می‌دهد. همچنین در شکل (۴-۲۸) نمودار باقی‌مانده بر حسب فعالیت تجربی نمایش داده شده است. با توجه به نتایج به دست آمده از سه مرحله ارزیابی می‌توان نتیجه گرفت که اگر چه سری ارزیابی و آزمون دارای ضریب تعیین قابل قبول می‌باشند اما چون ضریب تعیین به دست آمده

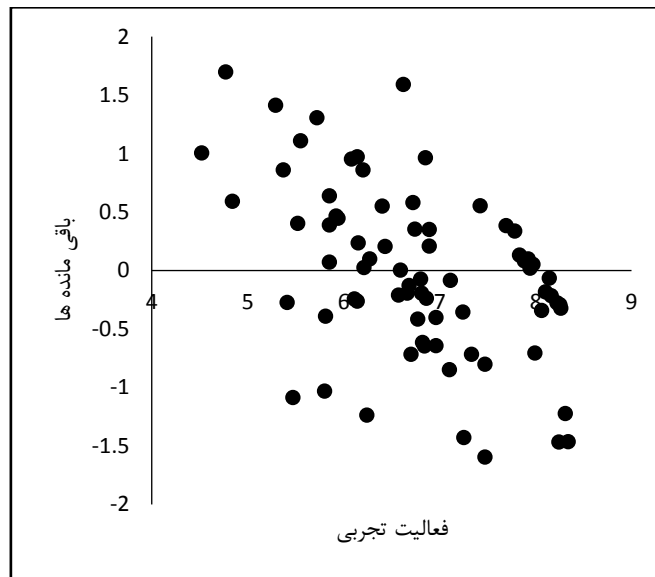
در ارزیابی به روش رد مرحله‌ای تک تک در محدوده قابل قبول نمی‌باشد [۳۴] و با این مدل نمی‌توان مقادیر pEC<sub>50</sub> را برای هر ترکیب جدید با خطای کم و قابل قبول پیش‌بینی کرد. در مراحل بعدی این پژوهش با افزودن گروهی از توصیف‌کننده‌های ویژگی مولکولی به توصیف‌کننده‌های داکینگ مولکولی شاهد بهبود کارایی مدل خواهیم بود.

جدول (۴-۴) نتایج حاصل از ارزیابی مدل با استفاده از روش رد مرحله‌ای تک تک

شماره ترکیب	pEC <sub>50</sub>		درصد خطا	شماره ترکیب	pEC <sub>50</sub>		درصد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده			فعالیت تجربی	فعالیت پیش‌بینی شده	
۱	۶/۸۵	۷/۸۱	۱۴/۰۱	۳۸	۷/۴۷	۵/۸۸	-۲۱/۲۹
۲	۷/۴۲	۷/۹۷	۷/۴۱	۳۹	۶/۱۴	۵/۸۸	-۴/۲۳
۳	۶/۷۲	۷/۳۰	۸/۶۳	۴۰	۴/۵۲	۵/۵۲	۲۲/۱۲
۴	۶/۸۹	۷/۲۴	۵/۰۸	۴۱	۵/۴۱	۵/۱۴	-۴/۹۹
۵	۶/۶۲	۸/۲۱	۲۴/۰۲	۴۲	۵/۵۲	۵/۹۲	۷/۲۵
۶	۶/۸۹	۷/۱۰	۳/۰۵	۴۳	۶/۱۴	۷/۱۱	۱۵/۸۰
۷	۷/۲۴	۶/۸۸	-۴/۹۷	۴۴	۶/۰۸	۷/۰۳	۱۵/۶۳
۸	۷/۳۳	۶/۶۱	-۹/۸۲	۴۵	۶/۱۵	۶/۳۹	۳/۹۰
۹	۷/۱۱	۷/۰۳	-۱/۱۳	۴۶	۶/۸۶	۶/۶۲	-۳/۵۰
۱۰	۸/۳۱	۷/۰۹	-۱۴/۶۸	۴۷	۶/۷۰	۵/۹۸	-۱۰/۷۵
۱۱	۷/۹۴	۷/۹۶	۰/۲۵	۴۸	۷/۱۰	۶/۲۵	-۱۱/۹۷
۱۲	۷/۶۹	۸/۰۸	۵/۰۷	۴۹	۶/۸۱	۶/۶۲	-۲/۷۹
۱۳	۸/۲۴	۶/۷۷	-۱۷/۸۴	۵۰	۶/۷۴	۷/۰۹	۵/۱۹
۱۴	۸/۲۶	۷/۹۴	-۳/۸۷	۵۱	۷/۴۷	۶/۶۷	-۱۰/۷۱
۱۵	۷/۹۹	۷/۲۸	-۸/۸۹	۵۲	۶/۴۰	۶/۹۵	۸/۵۹
۱۶	۸/۱۰	۷/۹۲	-۲/۲۲	۵۳	۶/۶۶	۶/۴۷	-۲/۸۵
۱۷	۸/۲۵	۷/۹۶	-۳/۵۲	۵۴	۶/۸۴	۶/۱۹	-۹/۵۰
۱۸	۸/۱۴	۸/۰۷	-۰/۸۶	۵۵	۶/۳۷	۶/۳۷	۱/۵۹
۱۹	۷/۷۸	۸/۱۲	۴/۳۷	۵۶	۶/۱۱	۵/۸۷	-۳/۹۳
۲۰	۸/۳۴	۶/۸۸	-۱۷/۵۱	۵۷	۵/۸۱	۵/۴۲	-۶/۷۱
۲۱	۸/۱۶	۷/۹۵	-۲/۵۷	۵۸	۵/۲۹	۶/۷۰	۲۶/۶۵
۲۲	۸/۲۲	۷/۹۴	-۳/۴۱	۵۹	۵/۸۰	۴/۷۷	-۱۷/۷۶
۲۳	۷/۸۸	۷/۹۶	۱/۰۲	۶۰	۶/۵۹	۶/۵۹	۰/۰۰
۲۴	۷/۸۳	۷/۹۶	۱/۶۶	۶۱	۶/۲۰	۷/۰۶	۱۳/۸۷
۲۵	۸/۰۶	۷/۷۲	-۴/۲۲	۶۲	۶/۶۸	۶/۵۵	-۱/۹۵
۲۶	۷/۹۷	۸/۰۲	۰/۶۳	۶۳	۵/۹۲	۶/۳۸	۷/۷۷
۲۷	۷/۹۲	۸/۰۲	۱/۲۶	۶۴	۴/۸۴	۵/۴۳	۱۲/۱۹
۲۸	۶/۲۴	۵/۰۰	-۱۹/۸۷	۶۵	۵/۷۲	۷/۰۳	۲۲/۹۰
۲۹	۷/۲۵	۵/۸۲	-۱۹/۷۲	۶۶	۵/۵۵	۶/۶۶	۲۰/۰۰
۳۰	۶/۹۶	۶/۵۶	-۵/۷۵	۶۷	۵/۹۴	۶/۳۹	۷/۵۸
۳۱	۶/۵۷	۶/۳۶	-۳/۲۰	۶۸	۵/۴۷	۴/۳۸	-۱۹/۹۳
۳۲	۶/۸۰	۶/۷۳	-۱/۰۳	۶۹	۵/۸۵	۶/۲۴	۶/۶۷
۳۳	۵/۳۷	۶/۲۳	۱۶/۰۱	۷۰	۵/۸۵	۵/۹۲	۱/۲۰
۳۴	۶/۲۱	۶/۲۳	۰/۳۲	۷۱	۶/۴۳	۶/۶۳	۳/۱۱
۳۵	۶/۷۷	۶/۳۶	-۶/۰۶	۷۲	۶/۸۲	۶/۲۱	-۸/۹۴
۳۶	۴/۷۷	۶/۷۲	۳۵/۶۴	۷۳	۵/۸۵	۶/۴۹	۱۰/۹۴
۳۷	۶/۹۶	۶/۳۲	-۹/۲۰				



شکل (۴-۲۷) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی به روش رد مرحله‌ای تک تک برای کل داده‌ها



شکل (۴-۲۸) نمودار باقی مانده بر حسب فعالیت تجربی برای کل داده‌ها

## ۴-۳-۴ مدل سازی با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

در مطالعه حاضر به منظور بهبود مدل شبکه عصبی مصنوعی برای پیش‌بینی فعالیت ضد ایدز ترکیبات مورد مطالعه، ۱۶ شاخص شبه دارویی برای ۷۳ مشتق آزینی مورد مطالعه با استفاده از نرم افزار DRAGON محاسبه شدند و به همراه ۱۰ توصیف کننده‌ی داکینگ مولکولی برای ساخت مدل شبکه عصبی مصنوعی مورد استفاده قرار گرفتند. برای انتخاب بهترین توصیف کننده‌ها از رگرسیون خطی استفاده شد و در نهایت موثرترین توصیف کننده‌ها به‌عنوان ورودی شبکه عصبی مصنوعی برای ساخت مدل غیر خطی QSAR استفاده شد. در ادامه کارایی بهترین مدل با استفاده از ارزیابی سری آزمون و روش رد مرحله‌ای تک تک برای همه مولکول‌های مجموعه داده‌ها بررسی شد.

## ۴-۳-۴ شاخص‌های شبه دارویی

هر شاخص شبه دارویی از تجزیه و تحلیل توزیع خواص فیزیکوشیمیایی همچون لگاریتم ضریب توزیع آب-اکتانول (چربی دوستی)<sup>۱</sup> محاسبه شده با روش قوز-کرپین-ویزوانادان<sup>۲</sup> (ALOGP)، ضریب شکست مولار اتمی<sup>۳</sup> (AMR)، وزن مولکولی<sup>۴</sup> (MW) و تعداد کل اتم‌ها<sup>۵</sup> (nAT) و اجزای شیمیایی مولکول‌های دارویی شناخته شده موجود در پایگاه جامع شیمی دارویی<sup>۶</sup> (CMC) و هفت کلاس دارویی دیگر معرفی شده توسط مرکز بیماری‌ها<sup>۷</sup> به دست می‌آیند [۱۰۶].

ALOGP شامل معادله رگرسیون بر اساس سهم چربی دوستی ۱۱۵ نوع اتم است و AMR براساس مدل Ghose-Crippon و با روش مشارکت گروهی محاسبه می‌شود. هر اتم در هر سازه به یکی از ۱۱۵ نوع اتم تقسیم می‌شود. مقادیر کمی ALOGP و AMR برای هر ترکیب با دسته‌بندی اتم‌ها به اتم‌های متصل برای لحاظ کردن برهمکنش‌های درون مولکولی با استفاده از رابطه زیر محاسبه

<sup>۱</sup> Hydrophobisity

<sup>۲</sup> Ghose-Crippen-Viswanadhan

<sup>۳</sup> Atomic Molar refractivity

<sup>۴</sup> Molecural weight

<sup>۵</sup> Number of atoms

<sup>۶</sup> Comprehensive Medicial Chemistry

<sup>۷</sup> Desease state

می‌گردد [۱۰۹-۱۰۶].

$$P_{cal} = \sum_i n_i a_i \quad \text{رابطه (۲-۴)}$$

در این عبارت  $P_{cal}$  خصوصیت محاسبه شده و  $n_i$  تعداد اتم‌های نوع  $i$  در مولکول و  $a_i$  سهم مشارکت اتم‌های نوع  $i$  می‌باشد.

توجه داشته باشید که برآوردهای ALOGP و AMR فقط برای ترکیبات دارای اتم‌های H, C, O, N, S, Se, P, B, Si و هالوژن‌ها ارائه می‌شود. اگر مقدار عددی ALOGP, AMR, MW, nAT و ترکیب مورد نظر در محدوده مشخص شده در DRAGON باشد و علاوه بر این شامل تعدادی از گروه‌های عاملی از قبیل حلقه بنزنی، گروه کربوکسامید، یک گروه هیدروکسیل الکلی، یک حلقه هتروسیکل (آلیفاتیک و آروماتیک)، یک آمین آلیفاتیک، یک کربوکسی استر و یک گروه کتونی باشد، مقدار کمی شاخص‌های شبه دارویی برابر با یک و در غیر این صورت برابر با صفر خواهد شد. علاوه بر این دو دامنه متفاوت برای خواص فیزیکوشیمیایی پیشنهاد شده است که عبارتند از: محدوده واجد شرایط<sup>۱</sup> که تقریباً ۸۰٪ داروهای مورد مطالعه را پوشش می‌دهد و محدوده ترجیحی<sup>۲</sup> که تقریباً ۵۰٪ از داروها را پوشش می‌دهد [۱۱۰]. که در این مطالعه از ۱۶ شاخص شبه دارویی محاسبه شده به وسیله نرم افزار DRAGON در مدل‌سازی QSAR ۷۳ مشتق آزینی به عنوان بازدارنده HIV استفاده شد.

۲-۴-۳-۴ انتخاب موثرترین توصیف کننده‌ها از مجموع توصیف کننده‌های داکینگ مولکولی و

### ساختاری

انتخاب موثرترین توصیف کننده‌ها برای ساخت یک مدل QSAR کارآمد بسیار ضروری می‌باشد. به این منظور مطابق بخش (۳-۴) از میان ۲۶ توصیف کننده‌ی داکینگ مولکولی و شاخص‌های شبه دارویی مربوط به ساختار شیمیایی ترکیبات مورد مطالعه ۴ توصیف کننده ارائه شده در جدول (۴-۴۴)

<sup>۱</sup> Qualifying range

<sup>۲</sup> Preferred range

شامل ۳ توصیف کننده داکینگ مولکولی و ۱ توصیف کننده شاخص شبه دارویی به عنوان موثرترین توصیف کننده‌ها انتخاب شدند. همچنین جدول (۴-۴۵) عدم همبستگی معنادار بین این توصیف کننده‌ها را نشان می‌دهد. با توجه به جدول (۴-۴۶) مقادیر VIF محاسبه شده برای توصیف کننده‌ها کوچکتر از ۱۰ بوده و در نتیجه عدم وجود هم‌خطی در بین توصیف کننده‌ها اثبات می‌شود.

جدول (۴-۴۴) توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR

No	Symbol	Meaning	Class
۱	EE	Electrostatic Energy	Docking derived descriptor
۲	Neolastic 80	Ghose-Viswanadhan-Wendoloskiantineoplastic-like index at 80%	DLIs
۳	TI	Final Total Internal Energy	Docking derived descriptor
۴	pK <sub>i</sub>	(-Log) of Estimated Inhibition Constant, K <sub>i</sub>	Docking derived descriptor

جدول (۴-۴۵) ماتریس همبستگی توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR

	EE	Neolastic 80	TI	pK <sub>i</sub>
EE	۱			
Neolastic 80	-۰/۲۴	۱		
TI	۰/۷۸	-۰/۲۶	۱	
pK <sub>i</sub>	-۰/۰۲	۰/۰۳	-۰/۳۵	۱

جدول (۴-۴۶) مقادیر تورم واریانس توصیف کننده‌های انتخاب شده از مجموع توصیف کننده‌های ساختاری و داکینگ مولکولی با روش SR

توصیف کننده	EE	Neolastic 80	TI	pK <sub>i</sub>
VIF	۳/۱۶	۱/۰۸	۳/۶۷	۱/۴۲

#### ۳-۴-۳-۴ مدل سازی شبکه عصبی مصنوعی با استفاده از مجموع توصیف کننده‌های داکینگ

##### مولکولی و ساختاری

به منظور ارائه یک مدل QSAR معتبر برای پیش‌بینی فعالیت ضد ایدز مشتقات آزینی مورد مطالعه، مجموعه داده‌ها به سه دسته سری آموزش (۵۳ ترکیب)، سری ارزیابی (۱۰ ترکیب) و سری آزمون (۱۰ مولکول ترکیب) تقسیم شد. سپس برای یافتن مقادیر بهینه پارامترهای مؤثر، هر شبکه با تعداد ورودی ۴، با دو الگوریتم آموزشی لوبز-مارکوات و تنظیم بیزی، دو تابع انتقال لگاریتم سیگموئید

و تانژانت سیگموئید، تعداد گره در لایه پنهان بین ۲ تا ۱۰ با گام ۱ و تعداد دور آموزش بین ۲ تا ۵۰ با گام ۲ آموزش داده شد. در روند بهینه سازی پارامترهای فوق، کمینه سازی MSE سری ارزیابی به عنوان معیار انتخاب شد. نتایج حاصل از انتخاب بهترین شبکه‌ها بر اساس کمترین مقدار MSE، در جدول (۴-۴۷) گزارش شده است. با توجه به بررسی و مقایسه نتایج شبکه‌های طراحی شده، شبکه با تابع آموزش لونیبرگ-مارکوات، تابع انتقال غیر خطی تانژانت سیگموئیدی، تعداد ۴ ورودی، تعداد دو گره در لایه پنهان و ۱۰ دور آموزش دارای ساختار ساده‌تری می‌باشد، از این رو در این بخش به عنوان شبکه بهینه انتخاب شد.

جدول (۴-۴۷) توابع و پارامترهای شبکه‌های بهینه توصیف کننده‌های منتخب از مجموع توصیف کننده های ساختاری و داکینگ مولکولی با روش SR

MSE	تعداد دور آموزش	تعداد گره	تابع انتقال	تابع آموزش	تعداد توصیف کننده
۰/۱۲	۱۲	۷	لگاریتم سیگموئید	تنظیم بیزی	۴
۰/۱۰	۲۴	۴	لگاریتم سیگموئید	لونیبرگ مارکوات	۴
۰/۱۳	۸	۴	تانژانت سیگموئید	تنظیم بیزی	۴
۰/۱۰	۱۰	۲	تانژانت سیگموئید	لونیبرگ مارکوات	۴

#### ۴-۳-۴ ارزیابی مدل شبکه عصبی مصنوعی

مدل برتر شبکه عصبی به منظور بررسی قابلیت تعمیم‌پذیری و جلوگیری از خطای بیش برآزش مورد ارزیابی قرار می‌گیرد. تعمیم‌پذیری مدل با استفاده از پیش‌بینی مقادیر فعالیت ترکیباتی که در ساخت مدل حضور نداشته‌اند بررسی می‌شود. در این پژوهش برای ارزیابی مدل انتخاب شده از روش‌های ارزیابی با استفاده از سری ارزیابی، سری آزمون و تکنیک رد مرحله‌ای تک تک استفاده شد.

#### - ارزیابی مدل با استفاده از سری ارزیابی

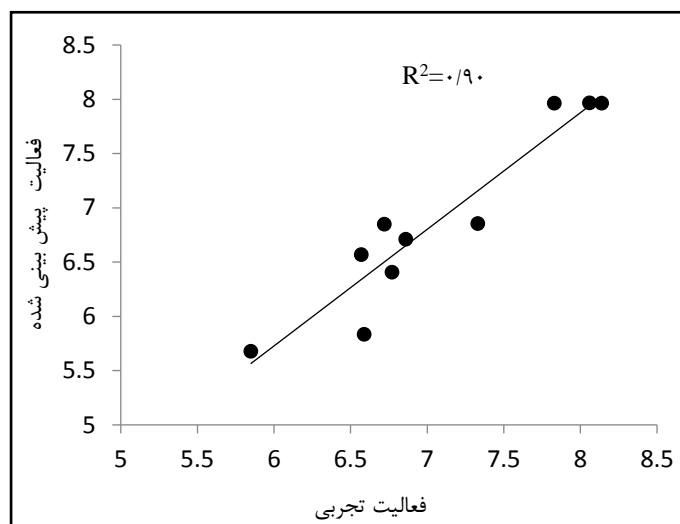
مدل منتخب شبکه عصبی مصنوعی جهت پیش‌بینی  $pEC_{50}$  ۱۰ مولکول سری ارزیابی، به کار گرفته شد. مقادیر فعالیت دارویی پیش‌بینی شده داده‌های سری ارزیابی در جدول (۴-۴۸) آورده شده است. همچنین شکل (۴-۲۹) ضریب تعیین فعالیت پیش‌بینی شده بر حسب فعالیت تجربی این سری



را نشان می‌دهد.

جدول ( ۴-۴۸ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری ارزیابی

شماره ترکیب	pEC <sub>50</sub>		درصد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۳	۶/۷۲	۶/۸۵	۱/۹۱
۸	۷/۳۳	۶/۸۶	-۶/۴۷
۱۸	۸/۱۴	۷/۹۶	-۲/۱۵
۲۴	۷/۸۳	۷/۹۶	۱/۷۲
۲۵	۸/۰۶	۷/۹۷	-۱/۱۷
۳۱	۶/۵۷	۶/۵۷	۰/۰۰
۳۵	۶/۷۷	۶/۴۱	-۵/۳۷
۴۶	۶/۸۶	۶/۷۱	-۲/۲۰
۶۰	۶/۵۹	۵/۸۳	-۱۱/۴۷
۶۹	۵/۸۵	۵/۶۸	-۲/۹۵



شکل ( ۴-۲۹ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی

### - ارزیابی مدل با استفاده از سری آزمون

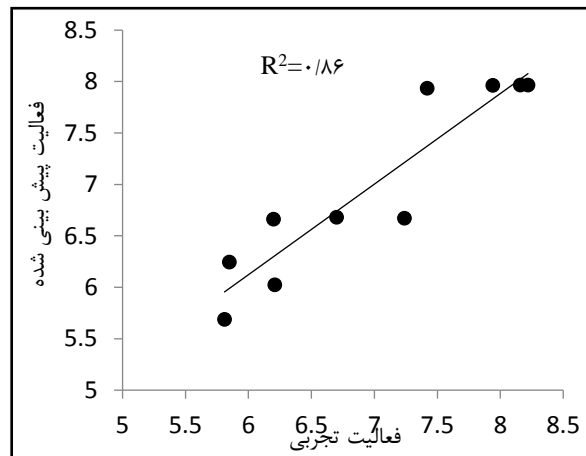
قدرت پیش‌بینی شبکه به وسیله ۱۰ ترکیب سری آزمون مورد بررسی قرار گرفت. نتایج

پیش‌بینی‌های فعالیت دارویی در جدول ( ۴-۴۹ ) آورده شده است. همچنین شکل ( ۴-۳۰ ) ضریب

تعیین فعالیت پیش‌بینی شده بر حسب فعالیت تجربی سری آزمون را نشان می‌دهد.

جدول ( ۴-۴۹ ) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون

شماره ترکیب	pEC <sub>50</sub>		درصد خطا
	فعالیت تجربی	فعالیت پیش‌بینی شده	
۲	۷/۴۲	۷/۹۴	۷/۰۱
۷	۷/۲۴	۶/۶۷	-۷/۸۷
۱۱	۷/۹۴	۷/۹۶	۰/۲۵
۲۱	۸/۱۶	۷/۹۷	-۲/۳۳
۲۲	۸/۲۲	۷/۹۷	-۳/۰۴
۳۴	۶/۲۱	۶/۰۳	-۲/۹۰
۴۷	۶/۷۰	۶/۶۸	-۰/۳۰
۵۷	۵/۸۱	۵/۶۹	-۲/۰۷
۶۱	۶/۲۰	۶/۶۶	۷/۴۲
۷۳	۵/۸۵	۶/۲۵	۶/۸۴



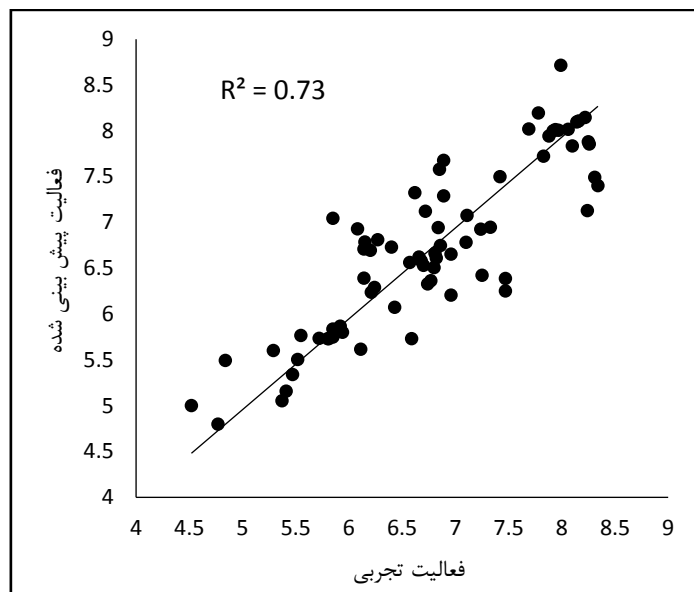
شکل ( ۴-۳۰ ) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی برای داده‌های سری ارزیابی

#### - ارزیابی مدل به روش رد مرحله‌ای تک تک

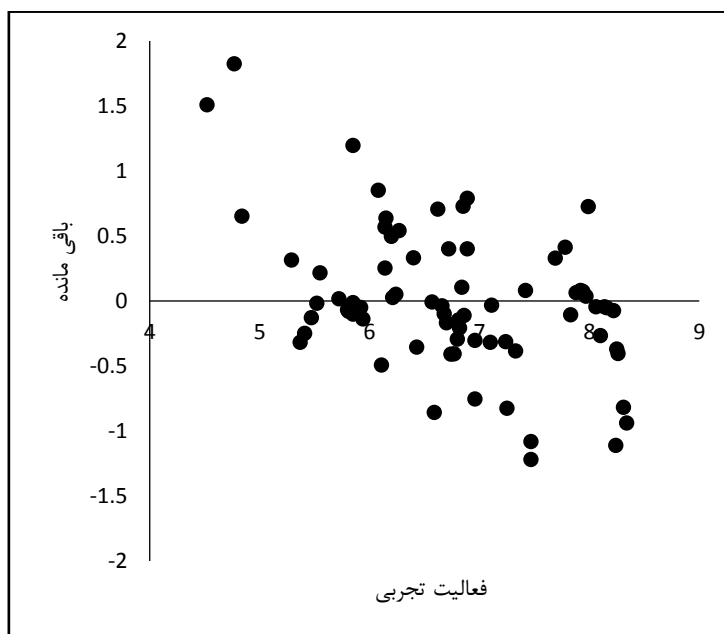
به منظور بررسی قدرت پیش‌بینی مدل شبکه عصبی مصنوعی، روش رد مرحله‌ای تک تک مورد استفاده قرار گرفت و نتایج خلاصه شده در جدول ( ۴-۵۰ )، تعمیم‌پذیری بالای مدل را نشان می‌دهد. ضریب تعیین ۰/۷۳ (شکل ( ۴-۳۱ )) نشان‌دهنده نزدیکی فعالیت پیش‌بینی شده به فعالیت تجربی است. شکل ( ۴-۳۲ ) نمودار تفاضل فعالیت تجربی از فعالیت پیش‌بینی شده (باقی‌مانده‌ها) بر حسب فعالیت تجربی را نشان می‌دهد. تقارن پراکندگی نقاط در دو طرف محور افقی عدم وجود خطای سیستماتیک را نشان می‌دهد.

جدول ( ۴-۵۰ ) نتایج حاصل از ارزیابی مدل با استفاده از روش رد مرحله‌ای تک تک

pEC <sub>50</sub>				pEC <sub>50</sub>			
شماره ترکیب	فعالیت تجربی	فعالیت پیش‌بینی شده	درصد خطا	شماره ترکیب	فعالیت تجربی	فعالیت پیش‌بینی شده	درصد خطا
۱	۶/۸۵	۷/۵۸	۱۰/۶۶	۳۸	۷/۴۷	۶/۲۵	-۱۶/۳۳
۲	۷/۴۲	۷/۵۰	۱/۰۸	۳۹	۶/۱۴	۶/۳۹	۴/۰۷
۳	۶/۷۲	۷/۱۲	۵/۹۵	۴۰	۴/۵۲	۵/۰۰	۱۰/۶۱
۴	۶/۸۹	۷/۲۹	۵/۸۱	۴۱	۵/۴۱	۵/۱۶	-۴/۶۲
۵	۶/۶۲	۷/۳۲	۱۰/۵۷	۴۲	۵/۵۲	۵/۵۰	-۰/۳۶
۶	۶/۸۹	۷/۶۸	۱۱/۴۷	۴۳	۶/۱۴	۶/۷۱	۹/۲۸
۷	۷/۲۴	۶/۹۳	-۴/۲۸	۴۴	۶/۰۸	۶/۹۳	۱۳/۹۸
۸	۷/۳۳	۶/۹۵	-۵/۱۸	۴۵	۶/۱۵	۶/۷۹	۱۰/۴۱
۹	۷/۱۱	۷/۰۸	-۰/۴۲	۴۶	۶/۸۶	۶/۷۵	-۱/۶۰
۱۰	۸/۳۱	۷/۴۹	-۹/۸۷	۴۷	۶/۷۰	۶/۵۳	-۲/۵۴
۱۱	۷/۹۴	۸/۰۱	۰/۸۸	۴۸	۷/۱۰	۶/۷۸	-۴/۵۱
۱۲	۷/۶۹	۸/۰۲	۴/۲۹	۴۹	۶/۸۱	۶/۶۶	-۲/۲۰
۱۳	۸/۲۴	۷/۱۳	-۱۳/۴۷	۵۰	۶/۷۴	۶/۳۳	-۶/۰۸
۱۴	۸/۲۶	۷/۸۶	-۴/۸۴	۵۱	۷/۴۷	۶/۳۹	-۱۴/۴۶
۱۵	۷/۹۹	۸/۷۲	۹/۱۴	۵۲	۶/۴۰	۶/۷۳	۵/۱۶
۱۶	۸/۱۰	۷/۸۳	-۳/۳۳	۵۳	۶/۶۶	۶/۶۲	-۰/۶۰
۱۷	۸/۲۵	۷/۸۸	-۴/۴۸	۵۴	۶/۸۴	۶/۹۴	۱/۴۶
۱۸	۸/۱۴	۸/۱۰	-۰/۴۹	۵۵	۶/۲۷	۶/۸۱	۸/۶۱
۱۹	۷/۷۸	۸/۱۹	۵/۲۷	۵۶	۶/۱۱	۵/۶۲	-۸/۰۲
۲۰	۸/۳۴	۷/۴۰	-۱۱/۲۷	۵۷	۵/۸۱	۵/۷۳	-۱/۳۸
۲۱	۸/۱۶	۸/۱۱	-۰/۶۱	۵۸	۵/۲۹	۵/۶۰	۵/۸۶
۲۲	۸/۲۲	۸/۱۴	-۰/۹۷	۵۹	۵/۸۰	۵/۷۳	-۱/۲۱
۲۳	۷/۸۸	۷/۹۴	۰/۷۶	۶۰	۶/۵۹	۵/۷۳	-۱۳/۰۵
۲۴	۷/۸۳	۷/۷۲	-۱/۴۰	۶۱	۶/۲۰	۶/۷۰	۸/۰۶
۲۵	۸/۰۶	۸/۰۱	-۰/۶۲	۶۲	۶/۶۸	۶/۵۸	-۱/۵۰
۲۶	۷/۹۷	۸/۰۱	۰/۵۰	۶۳	۵/۹۲	۵/۸۷	-۰/۸۴
۲۷	۷/۹۲	۸/۰۰	۱/۰۱	۶۴	۴/۸۴	۵/۴۹	۱۳/۴۳
۲۸	۶/۲۴	۶/۲۹	۰/۸۰	۶۵	۵/۷۲	۵/۷۴	۰/۳۵
۲۹	۷/۲۵	۶/۴۲	-۱۱/۴۵	۶۶	۵/۵۵	۵/۷۷	۳/۹۶
۳۰	۶/۹۶	۶/۶۶	-۴/۳۱	۶۷	۵/۹۴	۵/۸۰	-۲/۳۶
۳۱	۶/۵۷	۶/۵۶	-۰/۱۵	۶۸	۵/۴۷	۵/۳۴	-۲/۳۸
۳۲	۶/۸۰	۶/۵۱	-۴/۲۶	۶۹	۵/۸۵	۵/۸۴	-۰/۱۷
۳۳	۵/۳۷	۵/۰۵	-۵/۹۶	۷۰	۵/۸۵	۷/۰۵	۲۰/۵۱
۳۴	۶/۲۱	۶/۲۴	۰/۴۸	۷۱	۶/۴۳	۶/۰۷	-۵/۶۰
۳۵	۶/۷۷	۶/۳۶	-۶/۰۶	۷۲	۶/۸۲	۶/۶۱	-۳/۰۸
۳۶	۴/۷۷	۴/۸۰	۰/۶۳	۷۳	۵/۸۵	۵/۷۵	-۱/۷۱
۳۷	۶/۹۶	۶/۲۱	-۱۰/۷۸				



شکل ( ۴-۳۱) نمودار تغییرات فعالیت پیش‌بینی شده در مقابل فعالیت تجربی به روش رد مرحله‌ای تک تک برای کل داده‌ها

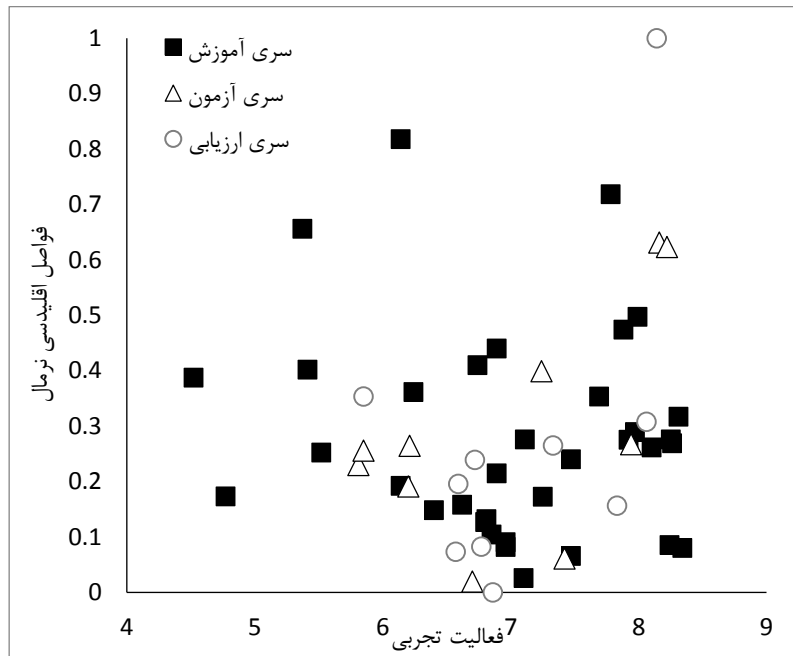


شکل ( ۴-۳۲) نمودار باقی مانده بر حسب فعالیت تجربی برای کل داده‌ها

### - ارزیابی مدل با استفاده از آزمون پراکندگی

فواصل اقلیدسی داده‌های سری آموزش، ارزیابی و آزمون، با استفاده از توصیف کننده‌های مدل برتر و مطابق با رابطه ( ۲-۱۷) به دست آمد. فواصل اقلیدسی نرمال شده بین ۰ و ۱ بر حسب فعالیت

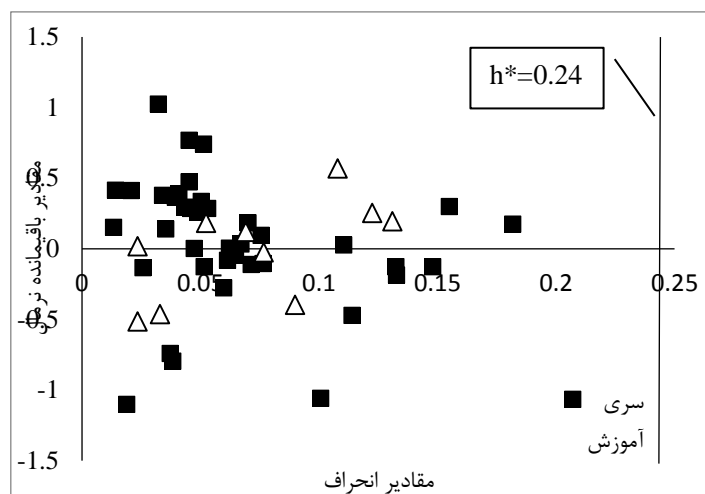
تجربی رسم شد. با توجه به شکل ( ۴-۳۳ ) توزیع داده‌های سری ارزیابی و آزمون در محدوده داده‌های سری آموزش نشان‌دهنده صحت تقسیم بندی داده‌ها می‌باشد.



شکل ( ۴-۳۳ ) نمودار پراکنندگی برای تقسیم بندی ترکیبات آزمینی

#### - ارزیابی مدل با استفاده از قلمرو کاربرد

قلمرو کاربرد مدل شبکه عصبی برتر این مطالعه، از رسم مقادیر باقیمانده استاندارد شده فعالیت پیش‌بینی شده بر حسب مقادیر حد انحراف به‌دست آمد. مطابق شکل ( ۴-۳۴ ) همه داده‌های سری آموزش و آزمون در محدوده قابل قبول قرار دارند. به این معنا که هیچ کدام از ترکیبات دارای مقدار  $h$  بزرگتر از  $h^*$  نشده و از طرفی همه داده‌ها در محدوده  $\pm 3$  برابر انحراف استاندارد نیز قرار دارند. در نتیجه مدل شبکه عصبی توسعه یافته از اعتبار و اطمینان بالایی برخوردار است.



شکل (۴-۳۴) نمودار قلمرو کاربرد برای مدل شبکه عصبی با ترکیبات آزمینی

### ۵-۴-۳-۴ نتیجه‌گیری مدل سازی با مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

در این مطالعه، ترکیبی از توصیف کننده‌های داکینگ مولکولی و شاخص‌های شبه دارویی در مدل سازی QSAR به روش شبکه عصبی مصنوعی استفاده شد. از طرفی به منظور بررسی تاثیر توصیف کننده‌ها، مدل شبکه عصبی با توصیف کننده‌های داکینگ مولکولی نیز توسعه یافت. پارامترهای ضریب تعیین و MSE برای پیش‌بینی فعالیت دارویی ترکیبات در سری آزمون برای هر دو مدل محاسبه و در جدول (۴-۵۱) خلاصه شد. مقایسه نتایج نشان می‌دهد که با افزودن شاخص‌های شبه دارویی ترکیبات مورد مطالعه به توصیف کننده‌های داکینگ مولکولی، قدرت مدل QSAR افزایش می‌یابد. دلیل بهبود کارایی مدل شبکه عصبی مصنوعی این است که این دسته از توصیف کننده‌ها حاوی اطلاعات برخی از خواص فیزیکی شیمیایی مهم مانند  $nAT$ ،  $MW$ ،  $MAR$  و  $ALOGP$  می‌باشند که در برهمکنش با پروتئین، جذب سلولی و دسترسی بیولوژیکی و در نتیجه فعالیت دارویی مولکول‌ها موثر می‌باشند.

جدول (۴-۵۱) مقایسه مدل های شبکه عصبی توسعه یافته با توصیف کننده‌های متفاوت در ارزیابی مدل با استفاده از سری آزمون

پارامتر آماری	مدل توسعه یافته با توصیف کننده های داکینگ مولکولی	مدل توسعه یافته با مجموع توصیف کننده های شاخص شبه دارویی و داکینگ مولکولی
$R^2$	۰/۷۹	۰/۸۶
MSE	۰/۲۲	۰/۱۱

#### ۴-۳-۶ ارزیابی مدل با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری با

#### استفاده از پارامترهای آماری

مطابق جدول (۴-۵۲) برخی از پارامترهای آماری، جهت ارزیابی توانایی پیشگویی مدل‌های ساخته شده به روش غیرخطی شبکه عصبی (ANN) به کار گرفته شد. نتایج رضایت‌بخش مدل غیرخطی نشان دهنده‌ی قدرت بالای پیش‌بینی مدل توسعه یافته می‌باشد

جدول (۴-۵۲) پارامترهای آماری برای مدل ANN با استفاده از مجموع توصیف کننده‌های داکینگ مولکولی و ساختاری

ردیف	پارامتر آماری	سری ارزیابی	سری آزمون	محدوده قابل قبول
۱	PRESS	۱/۰۵	۱/۱۱	-
۲	SEP	۰/۳۲	۰/۳۳	-
۳	MAE	۰/۲۵	۰/۲۷	-
۴	REP(%)	۴/۵۹	۴/۷۸	-
۵	MSE	۰/۱۰۵	۰/۱۱	-
۶	MRE	۳/۵۳	۴/۰۰	-
۷	R <sup>2</sup>	۰/۹۰	۰/۸۶	>۰/۶
۸	R <sub>0</sub> <sup>2</sup>	۰/۸۹	۰/۸۵	نزدیک به R <sup>2</sup>
۹	R <sub>0,R</sub> <sup>2</sup>	۰/۰۱	۰/۰۲	<۰/۱
۱۰	R <sub>m</sub> <sup>2</sup>	۰/۸۲	۰/۷۵	نزدیک به R <sup>2</sup>
۱۱	R <sub>0</sub> ' <sup>2</sup>	۰/۸۵	۰/۸۶	>۰/۵
۱۲	R <sub>0,R</sub> ' <sup>2</sup>	۰/۰۵	۰/۰۰	<۰/۱
۱۳	R <sub>m</sub> ' <sup>2</sup>	۰/۷۰	۰/۸۵	نزدیک به R <sup>2</sup>
۱۴	ΔR <sub>m</sub> <sup>2</sup>	۰/۰۴	۰/۱۲	<۰/۲
۱۵	k	۰/۹۷	۱/۰۰	۰/۸۵ ≤ k ≤ ۱/۱۵
۱۶	k'	۱/۰۳	۱/۰۰	۰/۸۵ ≤ k ≤ ۱/۱۵

#### ۴-۳-۷ آزمون Y-تصادفی

به منظور اجتناب از همبستگی‌های تصادفی و تضمین قدرت پیش‌بینی شبکه، آزمون Y-تصادفی انجام شد. لذا مقادیر pEC<sub>50</sub> مربوط به ۷۳ ترکیب در محدوده تغییرات خودشان به‌طور تصادفی ایجاد شدند. مدل QSAR جدید با استفاده از ماتریس توصیف کننده‌های اصلی و مقادیر تصادفی از فعالیت ترکیبات توسعه یافت. نتایج مدل سازی شبکه عصبی با استفاده از چندین مجموعه از فعالیت دارویی تصادفی در جدول (۴-۵۳) نشان داده شده است. با توجه به نتایج، هیچ وابستگی قابل ملاحظه‌ای بین

توصیف کننده‌ها و فعالیت بازدارندگی ترکیبات وجود ندارد، که نشان‌دهنده‌ی عدم همبستگی تصادفی در مدل شبکه عصبی مصنوعی ایجاد شده می‌باشد.

جدول (۴-۵۳) مقادیر  $R^2$  برای سری آزمون با استفاده از آزمون Y-تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
$R^2$	۰/۰۴	۰/۰۷	۰/۰۹	۰/۲۳	۰/۱۶	۰/۰۳	۰/۰۵	۰/۰۷	۰/۰۱	۰/۰۰



## فصل پنجم

### نتیجه گیری و آینده نگری

## ۱-۵ نتیجه‌گیری نهایی

در این مطالعه، برای پیش بینی فعالیت دارویی سه دسته از بازدارنده های سرطان و ایدز، توصیف کننده‌های داکینگ مولکولی با استفاده از برهم کنش لیگاند- پروتئین استخراج و در مدل شبکه عصبی مصنوعی استفاده شد. با توجه به فقدان اطلاعات ساختاری مناسب در توصیف کننده‌های داکینگ مولکولی، ارتباط مناسبی بین فعالیت دارویی و این توصیف کننده‌ها یافت نشد.

از این رو تلاش شد برای پیش‌بینی فعالیت دارویی مدل‌های شبکه عصبی مصنوعی ارائه شده، دسته ویژه‌ای از توصیف کننده‌های ساختاری به توصیف کننده‌های داکینگ مولکولی، افزوده گردد. بنابراین از میان ۲۲ طبقه از توصیف کننده‌های نرم افزار DRAGON، سه دسته از توصیف کننده‌ها (تابع توزیع شعاعی، شمارش گروه عاملی و ویژگی‌های مولکولی) به دلیل دارا بودن ویژگی‌های ساختاری منحصر به فرد محاسبه شد. سپس این توصیف کننده‌ها به توصیف کننده‌های داکینگ مولکولی افزوده شد. موثرترین توصیف کننده‌ها با به‌کارگیری روش رگرسیون گام به گام انتخاب شدند و شبکه عصبی مصنوعی با استفاده از این توصیف کننده‌های آموزش داده و بهینه شد. مدل شبکه عصبی بهینه با توجه به حداقل مقدار MSE سری ارزیابی انتخاب شد و برای پیش بینی فعالیت دارویی سری آزمون مورد استفاده قرار گرفت. نتایج جدول (۴-۱۵)، جدول (۴-۳۳) و جدول (۴-۵۲) نشان می‌دهند که توسعه مدل QSAR با استفاده از توصیف کننده‌های ساختاری ذکر شده و توصیف کننده‌های داکینگ مولکولی، علاوه بر حفظ اطلاعات برهم کنش لیگاند- پروتئین، قدرت پیش‌بینی و تعمیم پذیری مدل را در پیش بینی فعالیت دارویی مشتقات مورد استفاده، افزایش می‌دهد. در ادامه این بخش، ارتباط بین موثرترین توصیف کننده‌ها و فعالیت دارویی در سه مدل QSAR توسعه یافته مورد بررسی قرار گرفته است.

۵-۱-۱ بررسی ارتباط توصیف کننده‌های وارد شده در مدل با فعالیت بازدارندگی مشتقات ۴-

### آنیلینو کوئینازولین به عنوان بازدارنده‌های سرطان

در مطالعات QSAR، ایجاد ارتباط بین فعالیت بازدارندگی و توصیف کننده‌های ساختاری-داینک مولکولی بسیار حائز اهمیت است. پس از انتخاب توصیف کننده‌های مناسب، در نهایت ۸ توصیف کننده از میان کل توصیف کننده‌ها در مدل‌سازی شبکه‌ی عصبی مورد استفاده قرار گرفت که لیست کامل آن‌ها در جدول (۴-۸) آمده است. در ادامه به بررسی اهمیت این توصیف کننده‌ها و رابطه‌ی احتمالی آن‌ها با فعالیت بازدارندگی ترکیبات مورد مطالعه، پرداخته می‌شود. مدل MLR برای یافتن ارتباط بین مؤثرترین توصیف کننده‌های وارد شده در مدل ANN و مقادیر  $pIC_{50}$  به شرح زیر به دست آمد:

$$pIC_{50} = 17.499 - 1.302 VE + 2.235 U + 0.466 RDF115p - 0.326 RDF025u \\ + 0.220 RDF070e + 0.713 RDF105v - 0.263 RDF095u + 0.363 RDF065v \quad \text{رابطه (۵-۱)}$$

توصیف کننده‌های RDF مبتنی بر توزیع فواصل در نمایش هندسی مولکول و تشکیل یک کد تابع توزیع شعاعی است.

ضریب منفی دو توصیف کننده  $RDF025u$  و  $RDF095u$  در معادله، نشان‌دهنده ارتباط معکوس  $pIC_{50}$  با این دو توصیف کننده می‌باشد و با افزایش مقادیر این دو توصیف کننده فعالیت بازدارندگی ترکیبات کاهش می‌یابد. با مقایسه فعالیت ترکیب ۱۹ ( $pIC_{50}=5/21$ ) با  $RDF025u=25/68$  و  $RDF095u=4/48$  و ترکیب شماره ۱۱ ( $pIC_{50}=9/76$ ) با  $RDF025u=23/43$  و  $RDF095u=1/77$ ، اثر منفی این دو توصیف کننده بر روی فعالیت مشاهده می‌شود.

دو توصیف کننده  $RDF105v$  و  $RDF065v$  توسط حجم‌های واندروالس وزن‌دار شده‌اند. ضریب مثبت این دو توصیف کننده در معادله نشان‌دهنده‌ی اثر مثبت بر روی  $pIC_{50}$  می‌باشد و با افزایش مقادیر این دو توصیف کننده مقدار  $pIC_{50}$  ترکیبات افزایش می‌یابد. مقایسه فعالیت ترکیب شماره ۴۲ ( $pIC_{50}=9/76$ ) با  $RDF105v=2/41$  و  $RDF065v=12/75$  و فعالیت ترکیب شماره ۳۰ ( $pIC_{50}=8/32$ )

با  $RDF105v=0/05$   $RDF065v=3/31$  تاثیر مثبت این دو توصیف کننده بر روی فعالیت را نشان می دهند.

$RDF070e$  توصیف کننده تابع توزیع شعاعی است که توسط الکترونگاتیویته ساندerson<sup>۱</sup> وزن دار شده است. ضریب مثبت این توصیف کننده در معادله QSAR نشان دهنده تاثیر مثبت آن بر  $pIC_{50}$  می باشد. با توجه به رابطه مستقیم  $pIC_{50}$  و توصیف کننده  $RDF070e$ ، با افزایش مقادیر این توصیف کننده،  $pIC_{50}$  ترکیبات افزایش می یابد. مقایسه فعالیت ترکیب شماره ۴۲ ( $pIC_{50}=9/76$ ) با  $RDF070e=19/82$  و فعالیت ترکیب شماره ۷ ( $pIC_{50}=6/04$ ) با  $RDF070e=2/94$  نشان دهنده تاثیر مثبت این توصیف کننده تابع توزیع شعاعی بر فعالیت می باشد.

توصیف کننده  $RDF115p$  با استفاده از قطبش پذیری توزین شده است. ضریب مثبت این توصیف کننده نشان دهنده اثر مثبت بر روی  $pIC_{50}$  و در نتیجه اثر مثبت بر روی  $pIC_{50}$  دارو است. مقایسه فعالیت ترکیب شماره ۴۱ ( $pIC_{50}=11/22$ ) با  $RDF115p=4/06$  و فعالیت ترکیب شماره ۱۳ ( $pIC_{50}=7/25$ ) با  $RDF115p=0/135$  تاثیر مثبت این توصیف کننده تابع توزیع شعاعی بر فعالیت دارو را نشان می دهد.

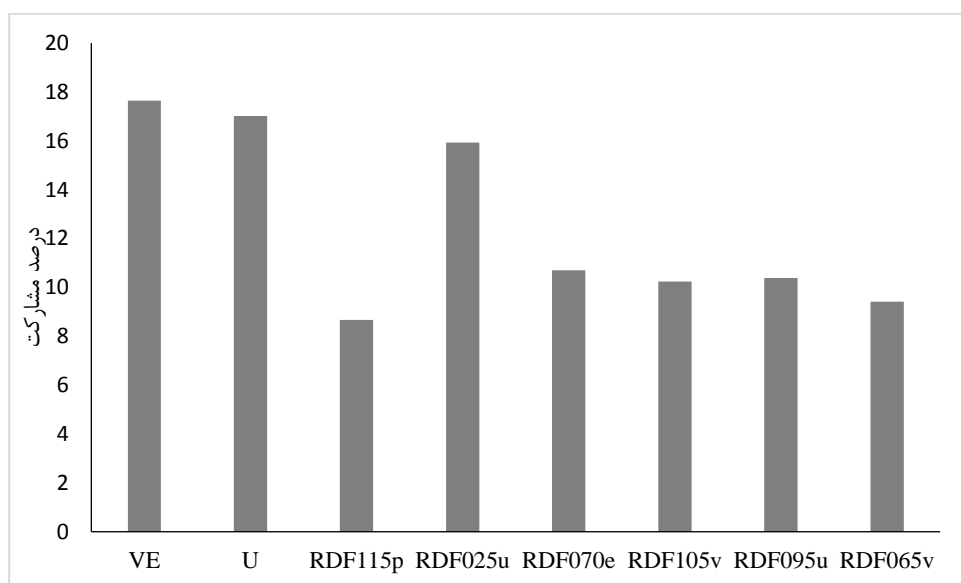
توصیف کننده VE مربوط به مجموع انرژی واندروالس، هیدروژنی و حلال پوشی است. هرچه این انرژی منفی تر باشد انرژی آزاد اتصال کمپلکس لیگاند-پروتئین منفی تر شده و در نتیجه کمپلکس پایدارتر خواهد شد. با توجه به ضریب منفی این توصیف کننده در معادله، مقادیر کوچک تر از این توصیف کننده  $pIC_{50}$  بزرگتر و در نتیجه فعالیت بازدارندگی بیشتری دارند. مقایسه فعالیت ترکیب شماره ۴۱ ( $pIC_{50}=11/22$ ) با  $VE=-9/24$  و فعالیت ترکیب شماره ۱۹ ( $pIC_{50}=5/21$ ) با  $VE=-6/35$  نشان می دهد که با کاهش مقدار این توصیف کننده فعالیت افزایش می یابد.

توصیف کننده انرژی درونی (U) از دسته توصیف کننده های داکینگ مولکولی است که با ضریب

---

<sup>۱</sup> Sanderson electronegativity

اثر مثبت در معادله وارد شده است. در نتیجه هرچه این انرژی مثبت تر شود فعالیت دارویی ترکیب افزایش می‌یابد. مقایسه فعالیت ترکیب شماره ۴۱ ( $pIC_{50}=11/22$ ) با  $U=-6/69$  و فعالیت ترکیب شماره ۱۹ ( $pIC_{50}=5/21$ ) با  $U=-7/03$  تاثیر مثبت این توصیف کننده بر روی فعالیت را نشان می‌دهد. میزان مشارکت توصیف کننده‌های منتخب در مدل شبکه عصبی مطابق با روشی که در بخش (۷-۷-۲) توضیح داده شد، برآورد شد، که نتایج آن در شکل (۱-۵) ارائه شده است. نتایج نمودار نشان می‌دهد که دو توصیف کننده داکینگ مولکولی دارای بیشترین اثر مشارکت در پیش‌بینی فعالیت تجربی ترکیبات ۴- آنیلینوکوئینازولین است و نقش موثری را در مدل شبکه عصبی بهینه دارند.



شکل (۱-۵) نمودار مشارکت توصیف کننده‌های منتخب در مدل شبکه عصبی توسعه یافته با ترکیبات ۴- آنیلینوکوئینازولین

۵-۱-۲ بررسی ارتباط توصیف کننده‌های وارد شده در مدل با فعالیت بازدارندگی مشتقات

### DAPY-Like به عنوان بازدارنده‌های HIV

در این قسمت بررسی اجمالی روی اثرات شش توصیف کننده وارد شده در مدل ANN بر فعالیت دارویی ترکیبات مورد مطالعه صورت خواهد گرفت. مدل MLR برای یافتن رابطه بین مؤثرترین توصیف کننده‌ها و مقادیر  $pEC_{50}$  به شرح زیر بدست آمد:

$$pEC_{50} = -0.331 - 0.439 VE + 0.359 nCp - 1.662 nRCN + 0.586 EE + 0.06 K_i$$

رابطه ( ۲-۵ )

$$- 0.376 nArNO_2$$

توصیف کننده VE مربوط به مجموع انرژی واندوالسی، هیدروژنی و حلال پوشی است. هر چقدر این انرژی منفی تر باشد کمپلکس لیگاند-پروتئین پایدارتر خواهد بود. با توجه به ضریب اثر متوسط منفی مربوط به این توصیف کننده با کوچکتر شدن مقدار توصیف کننده VE، فعالیت دارویی ترکیب افزایش می‌یابد. دو ترکیب شماره ۵۵ و ۶۰ مورد مقایسه قرار گرفت. ترکیب شماره ۵۵ دارای ۱۴/۷-  $pEC_{50} = 7/92$  و  $VE = -14/27$  و مولکول شماره ۶۰ با  $pEC_{50} = 6/98$  و  $VE = -14/27$  می‌باشد و ضریب اثر منفی این توصیف کننده بر روی فعالیت را نشان می‌دهد.

توصیف کننده nCP از دسته توصیف کننده‌های گروه عاملی می‌باشد که با اثر ضریب مثبت در مدل وارد شده است. با افزایش تعداد گروه‌های  $-CH_3$  متصل به بازوی سمت چپ لیگاند (شکل ( ۴-۱۱ )) و جای گیری در کیسه<sup>۱</sup> پیوندی هیدروفوبی، تعداد پیوندهای هیدروفوبی افزایش یافته و در نتیجه  $pEC_{50}$  لیگاند بهبود می‌یابد. مطالعات گزارش شده نشان می‌دهد که با افزایش تعداد گروه های  $-CH_3$  در لیگاند های مورد بررسی ، تعداد پیوندهای آگریز بین لیگاند و اسیدهای آمینه در کیسه پیوندی آب گریز افزایش می یابد [۹۹]. با مقایسه‌ی فعالیت ترکیب شماره ۱ ( $pEC_{50} = 5/19$ ) با دو استخلاف  $-CH_3$  بر روی حلقه سمت چپ و ترکیب شماره ۲ ( $pEC_{50} = 6/59$ ) با سه گروه  $-CH_3$ ، اثر مثبت این توصیف کننده بر روی فعالیت مشاهده می‌شود.

توصیف کننده nRCN دارای ضریب اثر منفی در مدل می‌باشد. در نتیجه با حضور این گروه عاملی در ترکیب، فعالیت آن کاهش می‌یابد. حضور استخلاف‌های آب دوست تر در موقعیت پارا و جای گیری در سطح مشترک پروتئین-آب، باعث افزایش حلالیت ترکیبات می‌شوند. از این رو وجود گروه CN در موقعیت پارای بازوی راست ترکیبات مورد مطالعه (شکل ( ۴-۱۱ ))، باعث افزایش حلالیت

<sup>۱</sup> Pocket

و در نتیجه افزایش فعالیت ترکیب می‌شود [۲۴]. از آنجایی که خاصیت آب دوستی گروه CN نسبت به گروه متیلن سیانید بیشتر می‌باشد، مقایسه فعالیت ترکیب شماره ۲ ( $pEC_{50}=6/59$ ) با استخلاف CN در موقعیت پارا حلقه سمت راست و ترکیب شماره ۴ ( $pEC_{50}=5/04$ ) با استخلاف  $CH_2CN$  در موقعیت پارای حلقه سمت راست، تاثیر منفی حضور این توصیف کننده بر روی  $pEC_{50}$  را نشان می‌دهد.

انرژی الکترواستاتیک (EE) از دسته توصیف کننده‌های داکینگ مولکولی می‌باشد. انرژی الکترواستاتیک از مجموع جفت برهمکنش‌های کولومبی ایجاد شده بین اتم‌های مولکول‌های دارو و گیرنده و با استفاده از بارهای جزئی گاستیگر<sup>۱</sup> مشخص شده در Autodock4.2 مطابق معادله زیر محاسبه می‌شود [۱۱۲, ۱۱۱]:

$$EE = \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \frac{q_{a_i} q_{b_j}}{4\pi\epsilon_0 r_{a_i b_j}} \quad \text{رابطه (۳-۵)}$$

در این معادله EE مجموع انرژی الکترواستاتیک متقابل حاصل از برهم کنش‌های کولومبی بین اتم‌های مولکول‌های a و b می‌باشد.  $N_a$  و  $N_b$  به ترتیب تعداد اتم‌های مولکول a و b می‌باشند.  $q_{a_i}$  بار جزئی اتم i ام مولکول a،  $q_{b_j}$  بار جزئی اتم j ام مولکول b و  $r_{a_i b_j}$  فاصله بین اتم‌های a و b می‌باشد. انرژی الکترواستاتیک جهت‌گیری و موقعیت لیگاند در جایگاه فعال پروتئین را نشان می‌دهد و اولین نیرویی است که داروها را به طرف سطح گیرنده با بار مخالف می‌کشد [۱۱۳]. در این پژوهش توصیف کننده EE با ضریب اثر مثبت در مدل وارد شده است. بنابراین با افزایش انرژی الکترواستاتیک فعالیت دارویی افزایش می‌یابد. با مقایسه ترکیب شماره ۶۹ ( $EE=-0/44$  و  $pEC_{50}=5/64$ ) و ترکیب شماره ۴۲ ( $EE=0/61$  و  $pEC_{50}=7/30$ ) می‌توان به اثر مثبت این توصیف کننده بر  $pEC_{50}$  پی برد.

ضریب بازداري تئوری حاصل از شبیه سازی داکینگ مولکولی ( $K_i$ ) طبق رابطه زیر به دست

می‌آید:

$$K_i = \exp (\Delta G \times 1000) / (R \times T_K) \quad \text{رابطه (۴-۵)}$$

<sup>۱</sup> Gasteiger

$\Delta G$  انرژی آزاد اتصال داکینگ است،  $R$  ثابت گازها برابر با  $1/98719$  (cal/mol. K) و  $T_K$  نیز برابر با ۲۹۸ کلوین می‌باشد. این توصیف کننده دارای ضریب کوچکی در مدل MLR می‌باشد و از طرفی با توجه به شکل (۵-۲) اثر مشارکت کمتری نیز در مدل شبکه عصبی دارد، بنابراین  $K_i$  برای تصحیح برازش مدل ایجاد شده است و یافتن ترکیباتی که بتواند تاثیر این توصیف کننده را به خوبی نشان دهد امکان پذیر نیست.

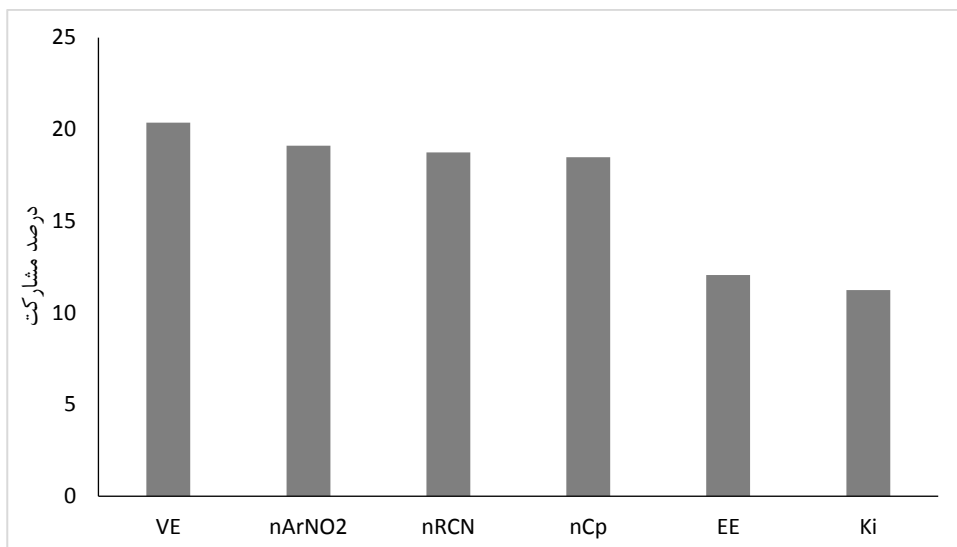
توصیف کننده  $nArNO_2$  توصیف کننده ساختاری دیگری است که با اثر منفی در مدل ظاهر شده است. حضور گروه  $NO_2$  بر روی حلقه آروماتیک بازوی راست ترکیبات مورد مطالعه باعث کاهش حلالیت آن‌ها می‌شود و در نتیجه  $pEC_{50}$  ترکیبات کاهش می‌یابد [۲۴، ۲۶]. مقایسه فعالیت ترکیب شماره ۳۲ با گروه آب دوست تر<sup>۱</sup>  $CN$  بر روی حلقه آروماتیک ( $pEC_{50}=7/30$ ) و ترکیب شماره ۳۳ با گروه  $NO_2$  بر روی حلقه آروماتیک ( $pEC_{50}=6/22$ ) اثر منفی این توصیف کننده بر روی  $pEC_{50}$  را نشان می‌دهد.

درصد مشارکت توصیف کننده‌های منتخب در ترکیبات مورد بررسی در شکل (۵-۲) نشان داده شده است. بر اساس این نمودار ۴ توصیف کننده  $VE$ ،  $nArNO_2$ ،  $nRCN$  و  $nCp$  دارای بیشترین اثر مشارکت در مدل شبکه عصبی برتر می‌باشند. بنابراین می‌توان نتیجه گرفت که اضافه شدن توصیف کننده‌های شمارش گروه عاملی به توصیف کننده‌های داکینگ مولکولی نقش موثری در بهبود قدرت پیش‌بینی مدل شبکه عصبی دارند.

---

<sup>۱</sup> Hydrophilic





شکل (۲-۵) نمودار مشارکت توصیف کننده‌های منتخب در مدل شبکه عصبی توسعه یافته با ترکیبات DAPY-like

### ۳-۱-۵ بررسی ارتباط توصیف کننده‌های وارد شده در مدل با فعالیت بازدارندگی مشتقات آزینی

#### به عنوان بازدارنده‌های HIV

در این قسمت با توجه به توصیف کننده‌های وارد شده در مدل ANN یک بررسی اجمالی روی اثرات این متغیرها بر فعالیت دارویی ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده شامل ۴ توصیف کننده منتخب از مجموع توصیف کننده‌های داکینگ مولکولی و شاخص‌های شبه دارویی می‌باشد. ارتباط بین مؤثرترین توصیف کننده‌ها و مقادیر  $pEC_{50}$  با استفاده از مدل MLR بدست آمد:

$$pEC_{50} = 5.439 + 0.728 EE - 0.764 \text{ Neoplastic80} + 0.843 \text{ TI} + 0.276 \text{ pK}_i \quad \text{رابطه (۵-۵)}$$

توصیف کننده EE مربوط به انرژی الکترواستاتیک حاصل از داکینگ لیگاند در پروتئین می‌باشد و جهت‌گیری لیگاند در پروتئین را نشان می‌دهد [۱۱۳, ۱۱۴] این توصیف کننده دارای ضریب اثر مثبت در مدل می‌باشد، بنابراین ترکیباتی با EE بزرگتر،  $pEC_{50}$  بیشتری خواهند داشت. با مقایسه ترکیب شماره ۲۶ ( $pEC_{50}=7/97$  و  $EE=0/48$ ) و ترکیب شماره ۲۷ ( $pEC_{50}=7/92$  و  $EE=0/43$ ) می‌توان تاثیر اثر مثبت این توصیف کننده بر روی فعالیت را مشاهده کرد.

توصیف کننده TI مربوط به انرژی نهایی درونی کل حاصل از شبیه سازی داکینگ مولکولی است. این نوع از انرژی مربوط به مجموع تغییرات عبارت‌های انرژی در گیر در فرایند تشکیل کمپلکس و لیگاند است. این توصیف کننده دارای ضریب اثر مثبت در مدل است به این معنا که هرچه انرژی درونی کل بزرگتر باشد سبب افزایش مقدار  $pEC_{50}$  می‌شود. ترکیب شماره ۲ دارای  $TI = -0/66$  و  $pEC_{50} = 7/42$  و مولکول شماره ۳ با  $TI = -1/17$  و  $pEC_{50} = 6/72$  می‌باشد. مقایسه این دو ترکیب ضریب اثر مثبت این توصیف کننده بر روی فعالیت را نشان می‌دهد.

Neoplastic-80 توصیف کننده‌ای از مجموعه شاخص‌های شبه دارویی می‌باشد که به‌عنوان یک توصیف کننده موثر در مدل توسعه یافته وارد شده است. محدوده قابل قبول DRAGON برای ALOGP، AMR، MW و nAT جهت بروز شباهت ۸۰ درصدی ترکیب به یک عامل آنتی نئوپلاستیک [۱۰۶، ۱۱۰، ۱۱۵] و محدوده مقادیر محاسبه شده این ویژگی‌ها برای ۷۳ مولکول مورد مطالعه در این پژوهش در جدول (۱-۵) گزارش شده است. این توصیف کننده با ضریب منفی در مدل ظاهر شده است. در نتیجه ترکیبات با Neoplastic-80 پایین‌تر فعالیت بازدارندگی بیشتری نشان می‌دهند. مقایسه فعالیت ترکیب شماره ۴۷ ( $pEC_{50} = 6/7$  و Neoplastic-80=۰) و ترکیب شماره ۶۰ ( $pEC_{50} = 6/59$ ) و فعالیت ترکیب شماره ۱ (Neoplastic-80=۱) تاثیر ضریب اثر منفی این توصیف کننده بر فعالیت ضد ایدز ترکیبات مورد مطالعه را نشان می‌دهد.

جدول (۱-۵) محدوده قابل قبول ویژگی‌های مولکولی برای بروز خاصیت آنتی نئوپلاستیک در DRAGON و محدوده این خواص برای ترکیبات مورد مطالعه

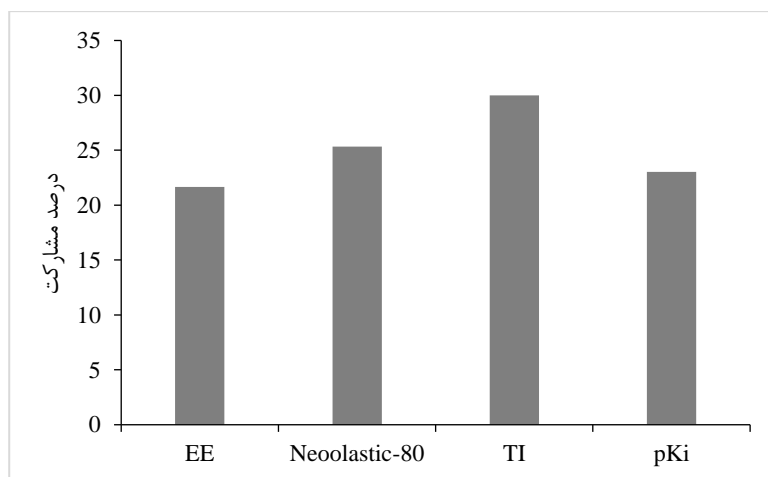
خصوصیت	ALOGP	AMR	MW	nAT
محدوده تعریف شده برای Neoplastic-80 در دراگون	۱/۵ تا ۴/۷	۴۳ تا ۱۲۸	۱۸۰ تا ۴۷۵	۶۳ تا ۲۱
محدوده تعریف شده برای ترکیبات مورد مطالعه	۳/۲ تا ۷/۱	۹۲ تا ۱۷۴	۳۳۱ تا ۶۰۸	۷۶ تا ۴۲

توصیف کننده  $pK_i$  مربوط به منفی لگاریتم ضریب بازدارندگی تئوری ( $K_i$ ) حاصل از شبیه سازی داکینگ مولکولی است.  $K_i$  طبق رابطه (۴-۵) به دست می‌آید.

این توصیف کننده دارای ضریب اثر مثبت در مدل می‌باشد. مقدار  $pEC_{50}$  برای ترکیباتی که  $pK_i$

بیشتری دارند، زیاد است. ترکیباتی با منفی لگاریتم ضریب بازداری تئوری بزرگتر با احتمال بیشتری توانایی برقراری بر هم کنش‌های هیدروژنی با اسیدهای آمینه موجود در جایگاه فعال را دارند. در حالت کلی لیگاندهایی که قادر باشند با اسیدهای آمینه موجود در جایگاه فعال گیرنده در تشکیل پیوندهای هیدروژنی و هیدروفوبی مشارکت نمایند، از نظر طراحی مولکول‌های زیست فعال حائز اهمیت هستند، زیرا از طریق روش‌های اصلاح مولکولی، بهینه سازی جایگاه‌های فارماکوفوری احتمالی و ارتقا به مولکول‌های زیست فعال قوی‌تر امکان پذیر خواهد بود. مقایسه ترکیب شماره ۲۸ ( $pK_i=7/44$ ) و  $pEC_{50}=6/24$  و ترکیب شماره ۲۹ ( $pK_i=7/86$  و  $pEC_{50}=7/25$ ) تاثیر مثبت این توصیف کننده بر فعالیت بازدارندگی ترکیبات مورد مطالعه را نشان می‌دهد.

میزان مشارکت توصیف کننده‌های منتخب مطابق رابطه (۲-۱۹) به دست آمد. درصد مشارکت توصیف کننده‌های منتخب ترکیبات مورد مطالعه در شکل (۵-۳) نشان داده شده است. با توجه به این نمودار مشاهده می‌شود که از بین ۴ توصیف کننده موثر بر فعالیت تجربی، ۳ توصیف کننده داکینگ مولکولی بیشترین سهم مشارکت (۷۵٪) در مدل شبکه عصبی برتر را دارند.



شکل (۵-۳) نمودار مشارکت توصیف کننده‌های منتخب در مدل شبکه عصبی توسعه یافته با ترکیبات آزمینی

## ۲-۵ آینده نگری

- ✓ استفاده از دسته‌های دیگری از توصیف کننده‌های ساختاری حاوی اطلاعات مولکولی برای افزودن به توصیف کننده‌های داکینگ مولکولی برای بررسی تاثیر حضور این دسته از توصیف کننده‌ها در بهبود مدل شبکه عصبی توسعه یافته
- ✓ بررسی برهم کنش لیگاند- پروتئین و استخراج توصیف کننده‌ها با استفاده از روش‌های دیگری همچون دینامیک مولکولی، Molergo ، iGEMDOCK و BINANA
- ✓ استفاده از مجموعه داده‌های دیگر برای بررسی اثر ترکیب توصیف کننده‌های داکینگ مولکولی- ساختاری (تابع توزیع شعاعی، شمارش گروه عاملی و ویژگی‌های مولکولی) برای پیش‌بینی فعالیت دارویی با به کارگیری مدل شبکه عصبی مصنوعی
- ✓ استفاده از روش‌های کارآمد انتخاب متغیر جدید برای انتخاب موثرترین توصیف کننده‌ها
- ✓ استفاده از روش‌های مدل‌سازی جدید برای توسعه مدل QSAR

- [1] Zhang A. (2009), "**Protein interaction networks: computational analysis**", Vol. 1, Cambridge University Press. pp. 89-90.
- [2] Grace E.S. (2006), "**Biotechnology unzipped: Promises and realities**", Vol. 1, Joseph Henry Press. pp. 25.
- [3] Brocchieri L. and Karlin S. (2005) "Protein length in eukaryotic and prokaryotic proteomes" **Nucleic Acids Res.**, **33,10**, pp 3390-3400.
- [۴] مکاری م.س.، (۱۳۹۲) "داروشناسی"، نشر فرهیختگان، ص ۱۲-۱۵.
- [۵] صبوری ع.، (۱۳۷۵) "سینتیک آنزیمی"، انتشارات دانشگاهی، تهران ص ۲۰۲-۲۲۰.
- [۶] اشنايدر گ.، (۲۰۰۸) "طراحی مولکولی مفاهیم و کاربردها"، نشر پژوهشی نوآوران شریف، ص ۱۴.
- [7] Siegel R., Ward E., Brawley O. and Jemal A. (2011) "Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths" **CA Cancer J. Clin.**, **61,4**, pp 212-236.
- [8] Zeinalzadeh A., Hosseini R. and Abdullahi L. (2015) "Survey of epidemiology of cancers in the patients above 15 years old in east Azerbaijan province, Iran 2013" **JArUMS**. **15,1**, pp 46-55.
- [9] Zhong S., Zhang C. and Johnson D.L. (2004) "Epidermal growth factor enhances cellular TATA binding protein levels and induces RNA polymerase I-and III-dependent gene activity" **Mol. Cell. Biol.**, **24,12**, pp 5119-5129.
- [10] Qiao L. and Wong B.C. (2009) "Targeting apoptosis as an approach for gastrointestinal cancer therapy" **Drug Resist. Updat.**, **12,3**, pp 55-64.
- [11] Feldinger K. and Kong A. (2015) "Profile of neratinib and its potential in the treatment of breast cancer" **Breast Cancer (Auckl)**., **7**, pp 147-162.
- [12] Zandi R., Larsen A.B., Andersen P., Stockhausen M.-T. and Poulsen H.S. (2007) "Mechanisms for oncogenic activation of the epidermal growth factor receptor" **Cell. Signal.**, **19,10**, pp 2013-2023.
- [13] Paul M.K. and Mukhopadhyay A.K. (2004) "Tyrosine kinase—role and significance in cancer" **Int. J. Med. Sci.**, **1,2**, pp 101-115.
- [14] Schwartz P.A., Kuzmic P., Solowiej J., Bergqvist S., Bolanos B., Almaden C., Nagata A., Ryan K., Feng J. and Dalvie D. (2014) "Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance" **PNAS**. **111,1**, pp 173-178.
- [15] Finzi D., Blankson J., Siliciano J.D., Margolick J.B., Chadwick K., Pierson T., Smith K., Lisziewicz J., Lori F. and Flexner C. (1999) "Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy" **Nat. Med.**, **5,5**, pp 512-517.
- [16] Esposito D. and Craigie R. (1999), "**HIV integrase structure and function**", Vol. 52, Elsevier. pp. 68.
- [17] Sanders R.W. (2012) "HIV takes double hit before entry" **BMC Biol.**, **10,1**, pp 99-100.

- [18] Richman D.D. (2001) "HIV chemotherapy" **Nature**. **410,6831**, pp **995-1001**.
- [19] Sarafianos S.G., Marchand B., Das K., Himmel D.M., Parniak M.A., Hughes S.H. and Arnold E. (2009) "Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition" **J. Mol. Biol.**, **385,3**, pp **693-713**.
- [20] Pomerantz R.J. and Horn D.L. (2003) "Twenty years of therapy for HIV-1 infection" **Nat. Med.**, **9,7**, pp **867-873**.
- [۲۱] شیرازیان م.، (۱۳۸۸)، پایان نامه کارشناسی ارشد: "تخمین پارامترهای مدل ریاضی HIV/AIDS با استفاده از داده های بالینی"، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد.
- [22] Fabian J. and Lewars E. (2004) "Azabenzenes (azines) The nitrogen derivatives of benzene with one to six N atoms: Stability, homodesmotic stabilization energy, electron distribution, and magnetic ring current; a computational study" **Can. J. Chem.**, **82,1**, pp **50-69**.
- [23] Miller C., Crain J., Tran B. and Patel N. (2011) "Ralpivirine: a new addition to the anti-HIV-1 armamentarium" **Drugs. today.**, **47,1**, pp **5-15**.
- [24] Huang B., Li C., Chen W., Liu T., Yu M., Fu L., Sun Y., Liu H., De Clercq E. and Pannecouque C. (2015) "Fused heterocycles bearing bridgehead nitrogen as potent HIV-1 NNRTIs. Part 3: optimization of [1, 2, 4] triazolo [1, 5-a] pyrimidine core via structure-based and physicochemical property-driven approaches" **Eur. J. Med. Chem.**, **92**pp **754-765**.
- [25] Tian Y., Du D., Rai D., Wang L., Liu H., Zhan P., De Clercq E., Pannecouque C. and Liu X. (2014) "Fused heterocyclic compounds bearing bridgehead nitrogen as potent HIV-1 NNRTIs. Part 1: design, synthesis and biological evaluation of novel 5, 7-disubstituted pyrazolo [1, 5-a] pyrimidine derivatives" **Biorg. Med. Chem.**, **22,7**, pp **2052-2059**.
- [26] Wang L., Tian Y., Chen W., Liu H., Zhan P., Li D., Liu H., De Clercq E., Pannecouque C. and Liu X. (2014) "Fused heterocycles bearing bridgehead nitrogen as potent HIV-1 NNRTIs. Part 2: Discovery of novel [1, 2, 4] Triazolo [1, 5-a] pyrimidines using a structure-guided core-refining approach" **Eur. J. Med. Chem.**, **85**, pp **293-303**.
- [27] Todeschini R., Consonni V., Mauri A. and Pavan M., (2007), DRAGON Web version 5.5,
- [28] Quinn J., Molecular Modeling Pro Plus, in, Norgwyn Montgomery Software, Version 5.
- [29] Katritzky A. (1995) "Pure & Appl. Chem. 1997, 245.(b) Katritzky, AR; Lobanov, VS; Karelson, M" **Chem. Soc. Rev.** **24**, pp **279-287**.
- [30] Wang M.-J., Zhao X.-B., Wu D., Liu Y.-Q., Zhang Y., Nan X., Liu H., Yu H.-T., Hu G.-F. and Yan L.-T. (2014) "Design, synthesis, crystal structure, insecticidal activity, molecular docking, and QSAR studies of novel N 3-substituted imidacloprid derivatives" **J. Agric. Food Chem.**, **62,24**, pp **5429-5442**.
- [31] Wu J., Wang Y. and Shen Y. (2014) "Molecular docking and QSAR analysis on maleimide derivatives selective inhibition against human monoglyceride lipase based on

various modeling methods and conformations" **Chemometrics Intellig. Lab. Syst.**, **131**, pp **22-30**.

[32] Quesada-Romero L. and Caballero J. (2014) "Docking and quantitative structure-activity relationship of oxadiazole derivatives as inhibitors of GSK3  $\beta$ " **Mol. Divers.**, **18,1**, pp **149-159**.

[33] Chakraborty S., Ramachandran B. and Basu S. (2014) "Encompassing receptor flexibility in virtual screening using ensemble docking-based hybrid QSAR: discovery of novel phytochemicals for BACE1 inhibition" **Mol. Biosyst.**, **10,10**, pp **2684-2692**.

[34] Sheikhpour R., Sarram M.A. and Gharaghani S. (2017) "Constraint score for semi-supervised feature selection in ligand-and receptor-based QSAR on serine/threonine-protein kinase PLK3 inhibitors" **Chemometrics Intellig. Lab. Syst.**, **163**, pp **31-40**.

[35] Amini Z., Fatemi M.H. and Gharaghani S. (2016) "Hybrid docking-QSAR studies of DPP-IV inhibition activities of a series of aminomethyl-piperidones" **Comput. Biol. Chem.**, **64**, pp **335-345**.

[36] Asadollahi-Baboli M. (2016) "In silico evaluation, molecular docking and QSAR analysis of quinazoline-based EGFR-T790M inhibitors" **Mol. Divers.**, **20,3**, pp **729-739**.

[37] Ghasemi J.B., Nazarshodeh E. and Abedi H. (2015) "Molecular docking, 2D and 3D-QSAR studies of new indole-based derivatives as HCV-NS5B polymerase inhibitors" **J. Iran. Chem. Soc.**, **12,10**, pp **1789-1799**.

[38] Xi L., Li S., Yao X., Wei Y., Li J., Liu H. and Wu X.a. (2014) "In silico study combining docking and QSAR methods on a series of matrix metalloproteinase 13 inhibitors" **Arch. Pharm.**, **347,11**, pp **825-833**.

[39] Garg A., Tewari R. and Raghava G.P. (2010) "K<sub>i</sub> DoQ: using docking based energy scores to develop ligand based model for predicting antibacterials" **BMC Bioinformatics**, **11,1**, pp **125**.

[40] Singla D., Anurag M., Dash D. and Raghava G.P. (2011) "A web server for predicting inhibitors against bacterial target GlmU protein" **BMC Pharmacol.**, **11,1**, pp **5-14**.

[41] Ebrahimi M., Khayamian T. and Gharaghani S. (2012) "Interactions between Activin-Like Kinase 5 (ALK5) receptor and its inhibitors and the construction of a Docking Descriptor-Based QSAR model" **J. Braz. Chem. Soc.**, **23,11**, pp **2043-2092**.

[42] Coi A. and Bianucci A.M. (2013) "Combining structure-and ligand-based approaches for studies of interactions between different conformations of the hERG K<sup>+</sup> channel pore and known ligands" **J. Mol. Graphics Model.**, **46**, pp **93-104**.

[43] Gharaghani S., Khayamian T. and Ebrahimi M. (2013) "Molecular dynamics simulation study and molecular docking descriptors in structure-based QSAR on acetylcholinesterase (AChE) inhibitors" **SAR QSAR Environ. Res.**, **24,9**, pp **773-794**.

[44] Ebrahimi M. and Khayamian T. (2014) "Interactions of G-quadruplex DNA binding site with berberine derivatives and construct a structure-based QSAR using docking descriptors" **Med. Chem. Res.**, **23,3**, pp **1327-1339**.

[45] Zheng F., Zhan M., Huang X., Hameed M.D.M.A. and Zhan C.-G. (2014) "Modeling in vitro inhibition of butyrylcholinesterase using molecular docking, multi-linear regression and artificial neural network approaches" **Biorg. Med. Chem.**, **22,1**, pp **538-**

549.

[46] Safarizadeh H. and Garkani-Nejad Z. (2019) "Molecular docking, molecular dynamics simulations and QSAR studies on some of 2-arylethenylquinoline derivatives for inhibition of Alzheimer's amyloid-beta aggregation: Insight into mechanism of interactions and parameters for design of new inhibitors" **J. Mol. Graphics Model.**, **87**, pp 129-143.

[47] Davood A. and Iman M. (2011) "Docking and QSAR studies of  $\beta$ -phenylethylidenehydrazine derivatives as a Gamma-aminobutyric acid aminotransferase inhibitor" **Med. Chem. Res.**, **20**,7, pp 955-961.

[48] Brereton R.G. (2003), "**Chemometrics: data analysis for the laboratory and chemical plant**", Vol. 1, John Wiley & Sons. pp. 215.

[49] Massart D.L., Vandeginste B.G., Buydens L., De Jong S., Lewi P.J., Smeyers-Verbeke J. and Mann C.K. (1998) "**Handbook of chemometrics and qualimetrics: Part A**" *Appl. Spectrosc.*, **52**, pp 302.

[50] Frank I.E. and Kowalski B.R. (1982) "Chemometrics" **Anal. Chem.**, **54**,5, pp 232-243.

[51] Kowalski B.R. (1980) "Chemometrics" **Anal. Chem.**, **52**,5, pp 112-122.

[52] Arab Chamjangali M., Beglari M. and Bagherian G. (2007) "Prediction of cytotoxicity data (CC50) of anti-HIV 5-phenyl-1-phenylamino-1H-imidazole derivatives by artificial neural network trained with Levenberg–Marquardt algorithm" **J. Mol. Graphics Model.**, **26**,1, pp 360-367.

[53] Arab Chamjangali M. (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.

[۵۴] یانگ د. (۲۰۰۹) "طراحی محاسباتی دارو"، انتشارات جامعه نگر تهران، ص ۱۸-۱۹.

[۵۵] قرقانی س.، (۱۳۹۱)، رساله دکتری: "روابط کمی ساختار فعالیت (QSAR) بر مبنای ساختار پروتئین و استفاده از مدل سازی همسانی، داکینگ مولکولی و شبیه سازی دینامیک مولکولی، طراحی داروهای جدید بر مبنای قطعات مولکولی (fragment) با تاثر همزمان بر پروتئین های استیل کولین استراز و آمین اکسیداز حساس به سمی کاربازید"، دانشکده شیمی، دانشگاه صنعتی اصفهان.

[56] Thomsen R. and Christensen M.H. (2006) "MolDock: a new technique for high-accuracy molecular docking" **J. Med. Chem.**, **49**,11, pp 3315-3321.

[۵۷] مهدوی آ.، (۱۳۹۴)، رساله دکتری: "پیش بینی تمایل پیوندی نانوفولرن ها به عنوان مهارکننده های پروتئاز HIV-1 با استفاده از روش QSAR مبتنی بر داکینگ مولکولی"، دانشکده شیمی، دانشگاه مازندران.

[58] Wang Z., Cheng L., Kai Z., Wu F., Liu Z. and Cai M. (2014) "Molecular modeling studies of atorvastatin analogues as HMGR inhibitors using 3D-QSAR, molecular docking and molecular dynamics simulations" **Bioorg. Med. Chem. Lett.**, **24**,16, pp



3869-3876.

[۵۹] یانگ د. (۲۰۰۹) "طراحی محاسباتی دارو"، انتشارات جامعه نگر، تهران ص ۳۴.

[60] Wermuth C., Ganellin C., Lindberg P. and Mitscher L. (1998) "Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)" **Pure Appl. Chem.**, **70**,5, pp 1129-1143.

[61] Stroud R. and Finer-Moore J. (2007), "**Computational and structural approaches to drug discovery: ligand-protein interactions**", Vol. 1, Royal Society of Chemistry. pp. 186.

[62] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. (2000) "The protein data bank" **Nucleic. Acids. Res.**, **28**,1, pp 235-242.

[63] Feinstein W.P. and Brylinski M. (2015) "Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets" **J. Cheminform.**, **7**,1, pp 18-39.

[64] Goodsell D.S., Morris G.M. and Olson A.J. (1996) "Automated docking of flexible ligands: applications of AutoDock" **J Mol Recognit.**, **9**,1, pp 1-5.

[65] He L. and Jurs P.C. (2005) "Assessing the reliability of a QSAR model's predictions" **J. Mol. Graphics Model.**, **23**,6, pp 503-523.

[66] Yasri A. and Hartsough D. (2001) "Toward an optimal procedure for variable selection and QSAR model building" **J. Chem. Inf. Comput. Sci.**, **41**,5, pp 1218-1227.

[67] Mendelsohn L.D. (2004) "ChemDraw 8 ultra, windows and macintosh versions" **J. Chem. Inf. Comput. Sci.**, **44**,6, pp 2225-2226.

[68] Stewart J.J. (1989) "Optimization of parameters for semiempirical methods II. Applications" **J. Comput. Chem.**, **10**,2, pp 221-264.

[69] ahmadzadeh S. and Nazari M. (2016) "Prediction and Investigation of Role of a Hotspot Residue on binding Site Properties of HER2 Inhibitor (Trastuzumab) in Breast Cancer" یازدهمین کنگره بین المللی سرطان پستان، تهران.

[70] Lapin L.L. (1983), "**Probability and statistics for modern engineering**", Vol. 1, Thomson Brooks/Cole., pp. 12-13.

[71] Xu L. and Zhang W.-J. (2001) "Comparison of different methods for variable selection" **Anal. Chim. Acta.** **446**,1-2, pp 475-481.

[72] Inc S., (2017), SPSS 25 for Windows student.

[۷۳] فرشادفر ع. (۱۳۸۰) "اصول و روش های پیشرفته آماری (تجزیه رگرسیون)" ، انتشارات طاق بوستان، کرمانشاه، ص ۱۸.

[۷۴] اشرفی م.س.، (۱۳۸۹)، "مطالعه ارتباط کمی ساختار- فعالیت مشتقات تیوکربامات ها به عنوان دسته ی جدیدی از بازدارنده های غیر نوکلئوزیدی "HIV" ، دانشکده شیمی، دانشگاه صنعتی شاهرود

[75] Goodarzi M., Deshpande S., Murugesan V., Katti S.B. and Prabhakar Y.S. (2009) "Is feature selection essential for ANN modeling?" **QSAR. COMB. SCI.**, **28**,11-12, pp 1487-1499.

- [76] Golbraikh A. and Tropsha A. (2002) "Beware of q<sup>2</sup>!" **J. Mol. Graphics Model.**, **20,4**, pp 269-276.
- [77] Veerasamy R., Rajak H., Jain A., Sivadasan S., Varghese C.P. and Agrawal R.K. (2011) "Validation of QSAR models-strategies and importance" **Int. J. Drug Discov.**, **3**, pp 511-519.
- [78] Roy P.P. and Roy K. (2008) "On some aspects of variable selection for partial least squares regression models" **QSAR. COMB. SCI.**, **27,3**, pp 302-313.
- [79] Gramatica P. and Sangion A. (2016) "A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology" **J. Chem. Inf. Model.**, **56,6**, pp 1127-1131.
- [80] Golbraikh A. (2003) "M.; Xiao, Z.; Xiao, YD; Lee, KH; Tropsha, A" **Rational selection of training and test sets for the development of validated QSAR models J. Comput.-Aided Mol.** **17**, pp 2-4.
- [81] Fatemi M.H. and Izadiyan P. (2011) "Cytotoxicity estimation of ionic liquids based on their effective structural features" **Chemosphere.** **84,5**, pp 553-563.
- [82] Caballero J., Garriga M. and Fernández M. (2006) "2D Autocorrelation modeling of the negative inotropic activity of calcium entry blockers using Bayesian-regularized genetic neural networks" **Biorg. Med. Chem.**, **14,10**, pp 3330-3340.
- [83] Joy S., Nair P.S., Hariharan R. and Pillai M.R. (2006) "Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware" **In Silico Biol.**, **6,6**, pp 601-605.
- [84] ViewerLite, (2001), Accelrys Inc, <https://viewerlite.software.informer.com/>
- [85] Mathworks I., (2012), MATLAB and statistics toolbox release 2012b, <https://www.mathworks.com/>
- [86] Henkelman R.M., Kay I. and Bronskill M.J. (1990) "Receiver operator characteristic (ROC) analysis without truth" **Med. Decis. Making.** **10,1**, pp 24-29.
- [87] Parida P., Deka P., Shankar B. and Yadav R. (2014) "In silico antigenic site evaluation and antiviral therapy against dengue serotypes" **Bangladesh J. Pharmacol.**, **9,1**, pp 83-95.
- [88] Zhou Q., Zhang N., Zhang C., Huang L., Niu Y., Zhang Y. and Liu W. (2010) "Molecular mechanism of enantioselective inhibition of acetolactate synthase by imazethapyr enantiomers" **J. Agric. Food Chem.**, **58,7**, pp 4202-4206.
- [89] Toor J.S., Sharma A., Kumar R., Gupta P., Garg P. and Arora S.K. (2011) "Prediction of drug-resistance in HIV-1 subtype C based on protease sequences from ART naive and first-line treatment failures in North India using genotypic and docking analysis" **Antiviral Res.**, **92,2**, pp 213-218.
- [90] MacDougall I.J. and Griffith R. (2006) "Selective pharmacophore design for  $\alpha$ 1-adrenoceptor subtypes" **J. Mol. Graphics Model.**, **25,1**, pp 146-157.
- [91] Awale M., Kumar V., Saravanan P. and Mohan C.G. (2010) "Homology modeling and atomic level binding study of Leishmania MAPK with inhibitors" **J. Mol. Model.**, **16,3**, pp 475-488.

[۹۲] مظفری ز.، (۱۳۹۵)، پایان نامه ارشد: "مطالعه ارتباط کمی ساختار-فعالیت مشتقات پیریدین و پیریمیدین به عنوان بازدارندهی HIV"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

[93] Arab Chamjangali M. and Ashrafi M. (2013) "QSAR study of necroptosis inhibitory activities (EC 50) of [1, 2, 3] thiadiazole and thiophene derivatives using Bayesian regularized artificial neural network and calculated descriptors" **Med. Chem. Res.**, **22**,1, pp 392-400.

[94] Zahouily M., Rakik J., Lazar M., Bahlaoui M., Rayadh A. and Komiha N. (2007) "Exploring QSAR of non-nucleoside reverse transcriptase inhibitors by artificial neural networks: HEPT derivatives" **ARKIVOC**. **14**, pp 245-256.

[95] Zupan J. and Gasteiger J. (1993), "**Neural networks for chemists: an introduction**", Vol. 1, John Wiley & Sons, Inc. pp. 21.

[96] González M.P., Gándara Z., Fall Y. and Gómez G. (2008) "Radial Distribution Function descriptors for predicting affinity for vitamin D receptor" **Eur. J. Med. Chem.**, **43**,7, pp 1360-1365.

[97] Noolvi M.N., Patel H.M., Bhardwaj V. and Chauhan A. (2011) "Synthesis and in vitro antitumor activity of substituted quinazoline and quinoxaline derivatives: Search for anticancer agent" **Eur. J. Med. Chem.**, **46**,6, pp 2327-2346.

[98] Hemmer M.C., Steinhauer V. and Gasteiger J. (1999) "Deriving the 3D structure of organic molecules from their infrared spectra" **Vib. Spectrosc.** **19**,1, pp 151-164.

[99] Huang B., Liang X., Li C., Chen W., Liu T., Li X., Sun Y., Fu L., Liu H. and De Clercq E. (2015) "Fused heterocycles bearing bridgehead nitrogen as potent HIV-1 NNRTIs. Part 4: design, synthesis and biological evaluation of novel imidazo [1, 2-a] pyrazines" **Eur. J. Med. Chem.**, **93**, pp 330-337.

[100] Burden F. and Winkler D. (2008), "**Bayesian regularization of neural networks**", Vol. 1, Springer, pp. 23-42.

[101] Chen W., Zhan P., Rai D., De Clercq E., Pannecouque C., Balzarini J., Zhou Z., Liu H. and Liu X. (2014) "Discovery of 2-pyridone derivatives as potent HIV-1 NNRTIs using molecular hybridization based on crystallographic overlays" **Bioorg. Med. Chem.**, **22**,6, pp 1863-1872.

[102] Chen X., Liu X., Meng Q., Wang D., Liu H., De Clercq E., Pannecouque C., Balzarini J. and Liu X. (2013) "Novel piperidinylamino-diarylpyrimidine derivatives with dual structural conformations as potent HIV-1 non-nucleoside reverse transcriptase inhibitors" **Bioorg. Med. Chem. Lett.**, **23**,24, pp 6593-6597.

[103] Chen X., Zhan P., Liu X., Cheng Z., Meng C., Shao S., Pannecouque C., De Clercq E. and Liu X. (2012) "Design, synthesis, anti-HIV evaluation and molecular modeling of piperidine-linked amino-triazine derivatives as potent non-nucleoside reverse transcriptase inhibitors" **Bioorg. Med. Chem.**, **20**,12, pp 3856-3864.

[104] Li D., Zhan P., Liu H., Pannecouque C., Balzarini J., De Clercq E. and Liu X. (2013) "Synthesis and biological evaluation of pyridazine derivatives as novel HIV-1 NNRTIs" **Bioorg. Med. Chem.**, **21**,7, pp 2128-2134.

[105] Wang J., Zhan P., Li Z., Liu H., De Clercq E., Pannecouque C. and Liu X. (2014) "Discovery of nitropyridine derivatives as potent HIV-1 non-nucleoside reverse

transcriptase inhibitors via a structure-based core refining approach" **Eur. J. Med. Chem.**, **76**, pp 531-538.

[106] Ghose A.K., Viswanadhan V.N. and Wendoloski J.J. (1999) "A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases" **J. Comb. Chem.**, **1,1**, pp 55-68.

[107] Ghose A.K. and Crippen G.M. (1987) "Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions" **J. Chem. Inf. Comput. Sci.**, **27,1**, pp 21-35.

[108] Kujawski J., Popielarska H., Myka A., Drabińska B. and Bernard M.K. (2012) "The log P parameter as a molecular descriptor in the computer-aided drug design—an overview" **CMST**. **18,2**, pp 81-88.

[109] Wildman S.A. and Crippen G.M. (1999) "Prediction of physicochemical parameters by atomic contributions" **J. Chem. Inf. Comput. Sci.**, **39,5**, pp 868-873.

[110] Chen C.-P., Chen C.-C., Huang C.-W. and Chang Y.-C. (2018) "Evaluating molecular properties involved in transport of small molecules in stratum corneum: a quantitative structure-activity relationship for skin permeability" **Molecules**. **23,4**, pp 911-928.

[111] Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K. and Olson A.J. (1998) "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function" **J. Comput. Chem.**, **19,14**, pp 1639-1662.

[112] Kitchen D.B., Decornez H., Furr J.R. and Bajorath J. (2004) "Docking and scoring in virtual screening for drug discovery: methods and applications" **Nat. Rev. Drug Discov.**, **3,11**, pp 935.

[۱۱۳] ساسانی ن. (۱۳۹۵) "شبیه سازی دینامیک مولکولی پروتئین همولیزین باسیلوس پومیلوس و داکینگ آن با کلسترول و پیشگویی تغییرات کنفورماسیونی این پروتئین در طراحی دارو با کمک بیوانفورماتیک" *مجله علمی - پژوهشی دانشگاه الزهراء (س) / زیست شناسی کاربردی*. ۲۹، ۲، ص ۷۹-

۱۰۰

[114] Kukić P. and Nielsen J.E. (2010) "Electrostatics in proteins and protein-ligand complexes" **Future Med. Chem.**, **2,4**, pp 647-666.

[115] Khajeh A. and Modarress H. (2014) "Linear and nonlinear quantitative structure-property relationship modelling of skin permeability" **SAR QSAR Environ. Res.**, **25,1**, pp 35-50.

## Abstract

Molecular docking was used to construct QSAR models for the prediction of the biological activities of three groups of medicinal compounds, and the molecular descriptors were extracted from the enzyme-ligand interactions using Autodock4.2 software. Since that the molecular docking descriptors do not represent the structural properties of ligands alone and also there is no strong relationship between molecular docking and biological activity, using structural descriptors (calculated by DRAGON software) in addition to docking descriptors is essential. In the first part, the study of the structure-activity of 43 quinazoline derivatives as cancer inhibitors was studied. 10 molecular docking descriptors were extracted for all compounds. The data set was divided into two training (35 compounds) and test (8 compounds) sets. The stepwise regression (SR) as the variable selection method was applied to the train set data. The artificial neural network (ANN) model was developed with two selected descriptors. The low determination coefficient ( $R^2$ ) and high mean square error (MSE) obtained for the test set (0.52 and 0.66 respectively) indicate the inappropriate predictability of the ANN model. In order to improve the results of the ANN model, radial distribution function (RDF) structural descriptors were added to the molecular docking descriptors. RDF descriptors represent the difference of the distribution of atoms in the molecule and the effect of this distribution on the inhibitory activity of the compounds. SR method was implemented on 160 descriptors (150 RDF and 10 molecular docking descriptors) and 8 effective descriptors were selected and used as the inputs of the ANN model. The high  $R^2$  and low MSE of the test set (0.90 and 0.15, respectively) prove the predictability and generalizability of the developed ANN model with a mixture of molecular-structure docking descriptors. In the second part, the structure-activity of 73 inhibitors of AIDS that are analog structure of Diarylpyrimidine (DAPY like) was studied. After extracting the molecular docking descriptors, the data set was divided into three training (53 compounds), validation (10 compounds) and test (10 compounds) sets. Two SR selected descriptors were introduced to the ANN model and all parameters of the ANN were optimized.  $R^2$  and MSE parameters of the test set were obtained 0.50 and 0.24, respectively. In order to improve the ANN model, the functional group counts structural descriptors were calculated for HIV inhibitors. SR method was applied on 164 descriptors

(154 functional group counts and 10 molecular docking descriptors), and the ANN model was developed using 7 selected descriptors. The biological activities of the test set data were predicted in the optimum ANN model.  $R^2$  and MSE of the test set were obtained 0.89 and 0.16 respectively that showed the improvement of the results of the ANN model with a mixture of molecular-structure docking descriptors. In the third section, a quantitative structure- activity relationship of the azine derivative as HIV inhibitors, was investigated. For this purpose, 10 molecular docking descriptors were obtained using the docking of the studied compounds in the active site of the protein. The data set was divided into three training (53 compounds), validation (10 compounds), and test (10 compounds) sets. Among the 10 molecular docking descriptors, 2 descriptors were selected as the most effective descriptors using SR method on the train set data. The ANN model was optimized using docking derived descriptors. The biological activity values of the test set compounds were predicted using the optimum ANN model.  $R^2$  and MSE parameters of the test set were obtained 0.79 and 0.23 respectively. Then, for the purpose of improving the ANN model, 16 drug like indexes were calculated for the studied compounds and added to 10 molecular docking descriptors. SR was employed to select the most effective descriptors on the train set data. 4 important descriptors selected from the total molecular docking and drug like indexes and they were used as ANN model inputs. The biological activities of the test set data were predicted using the optimum ANN model.  $R^2$  and MSE parameters of the test set are obtained and equal to 0.86 and 0.11 respectively. Therefore, using molecular docking descriptors and drug like indexes simultaneously, improved the efficiency of the ANN model.

**Keywords:** QSAR, Anti cancer, Anti HIV, molecular docking, stepwise regression, artificial neural network



**Faculty of Chemistry**

**Ph.D. Thesis in Analytical Chemistry**

**The application of new molecular descriptors in the development of  
QSAR models for some drug like compounds**

By:

**Mozhgan Beglari Amniye**

Supervisor:

**Dr. Nasser Goudarzi**

Advisor:

**Dr. Davood Shahsavani**

September 2020