



حوزه معاونت پژوهشی و فناوری

## گزارش پایانی طرح پژوهشی

با عنوان

پیش بینی فعالیت دارویی ضد ایدز (سیتو توکسیتی) مشتقات پیریل  
آریل سولفون به وسیله شبکه عصبی مصنوعی

کد: ۲۲۰۲۸

خرداد ۱۳۸۷

مجری: منصور عرب چم جنگلی

عضو هیئت علمی دانشکده شیمی

این طرح با استفاده از اعتبارات پژوهشی دانشگاه صنعتی شاهرود انجام شده است و تاریخ های تصویب و خاتمه آن به ترتیب ۸۶/۸/۱۳ و ۸۷/۸/۲۶ می باشد.

پیش بینی فعالیت داروئی ضد ایدز (سیتو توکسییتی) مشتقات ۱- (۵-کلروفنیل) سولفونیل]-۱-

## H- پیروول به وسیله شبکه عصبی مصنوعی

### چکیده:

در این پروژه یک مدل شبکه عصبی با آموزش به کمک الگوریتم لونیگ مارکوت برای مطالعات QSAR مشتقات ۱- (۵-کلروفنیل) سولفونیل]-۱- H- پیروول ارائه شده است. در این پروژه توصیفگرهای محاسبه‌ای مورد استفاده قرار گرفتند، لذا تمام ۱۴۸۱ توصیف کننده برای هر مولکول به کمک نرم افزار Dragon محاسبه گردید. از میان این توصیفگرها تعداد ۸ توصیفگر مهم که با داده‌های CC<sub>50</sub> ارتباط داشتند به روش رگرسیون مرحله به مرحله در نرم افزار SPSS انتخاب شده و در مدلسازی مورد استفاده قرار گرفتند. در این مطالعه، شبکه عصبی مصنوعی با جریان رو به جلو با الگوی آموزش پس انتشار بر اساس الگوریتم لونیگ-مارکوت انتخاب گردید و عوامل مؤثر در آموزش شبکه بررسی و بهینه شدند. در ارزیابی مدل ارائه شده در این پروژه برای پیش‌بینی داده‌های CC<sub>50</sub> داروهای ضد ایدز، روشهای استفاده از سری پیش بینی و روش رد مرحله‌ای تک تک بکار گرفته شد. خطای جذر میانگین (MSE) حاصله برای دو روش فوق به ترتیب ۰/۰۶۷ و ۰/۰۱۳ به دست آمد. نتایج حاصله نشان دهنده قدرت پیش‌بینی بالای مدل شبکه عصبی می‌باشد. نتایج آشکارا بیانگر این هستند که روش QSAR ارائه شده در این پروژه قادر است خاصیت داروئی ضد ایدز هر مشتقی از ۵- فینیل -۱- فینیل آمینو-۱- H- ایمیدازول با هر استخلافی را با قدرت بالا و خطای بسیار کم پیش‌بینی کند.

## فهرست مطالب

عنوان.....	صفحه.....
فصل اول - مقدمه.....	۱.....
۱-۱- مطالعات ارتباط کمی ساختار-فعالیت (خاصیت).....	۱.....
۱-۱-۱- جمع‌آوری لیستی از ترکیبات.....	۱.....
۱-۱-۲- بهینه‌سازی ساختار هندسی استخراج توصیفگرها.....	۲.....
۱-۲-۱- روشهای آغازین.....	۲.....
۱-۲-۲- روشهای نیمه تجربی.....	۴.....
۱-۲-۳- روشهای تابع دانسیته.....	۴.....
۱-۲-۴- روشهای مکانیک مولکولی.....	۵.....
۱-۳-۱- استخراج توصیفگرها.....	۵.....
۱-۴-۱- انتخاب بهترین توصیف کننده‌ها.....	۵.....
۱-۵-۱- مدلسازی.....	۶.....
۱-۶-۱- ارزیابی مدل.....	۹.....
۱-۲-۱- بیماری ایدز.....	۹.....
۱-۲-۱- تاریخچه کشف ویروس HIV.....	۱۰.....
۱-۲-۲- انواع HIV.....	۱۱.....
۱-۲-۳- میزان پایداری ویروس HIV.....	۱۲.....
۱-۲-۴- همانند سازی و چرخه تکثیر.....	۱۲.....
۱-۲-۵- بیماری‌زایی و یافته‌های ایمنی.....	۱۴.....
۱-۲-۶- مراحل بیماری.....	۱۴.....
۱-۲-۶-۱- مرحله مقدماتی (مرحله حاد).....	۱۴.....
۱-۲-۶-۲- مرحله بدون علامت (مزمین).....	۱۵.....
۱-۲-۶-۳- مرحله ایدز (بحرانی).....	۱۵.....

۱۶	۲-۱-۷- آمار مبتلایان به ایدز.....
۱۷	۲-۱-۸- پیشگیری و درمان .....
۱۹	۱-۳- مروری بر کارهای انجام شده .....
۲۱	فصل دوم-محاسبات و مدلسازی.....
۲۱	۲-۱- سری داده‌ها .....
۲۲	۲-۲- نرم افزارهای مورد استفاده .....
۲۴	۲-۳- انجام محاسبات و مدلسازی .....
۲۴	۲-۳-۱- رسم و بهینه‌سازی ساختار مولکولها.....
۲۴	۲-۳-۲- محاسبه توصیفگرهای مولکولی.....
۲۵	۲-۳-۳- انتخاب توصیفگرهای مهم .....
۲۷	۲-۳-۴- مدل شبکه عصبی و بهینه سازی آن .....
۲۸	۲-۳-۴-۱- انتخاب تعداد لایه‌های شبکه عصبی .....
۲۸	۲-۳-۴-۲- انتخاب الگوریتم آموزشی .....
۲۹	۲-۳-۴-۳- انتخاب نوع تابع تبدیل .....
۳۰	۲-۳-۴-۴- انتخاب سرعت آموزش و تعداد گره‌های لایه پنهان .....
۳۲	۲-۳-۴-۵- انتخاب تعداد دورهای آموزش.....
۳۳	۲-۳-۴-۶- ساختار و پارامترهای بهینه شده شبکه عصبی.....
۳۴	۲-۳-۵- ارزیابی مدل ارائه شده .....
۳۴	۲-۳-۵-۱- ارزیابی شبکه عصبی با استفاده از سری پیش‌بینی .....
۳۵	۲-۳-۵-۲- ارزیابی شبکه عصبی به روش رد مرحله ای تک تک.....
۳۹	فصل سوم - بحث و نتیجه‌گیری.....
۳۹	۳-۱- توصیف‌گرهای انتخاب شده .....
۴۰	۳-۲- ارزیابی آماری مدل ارائه شده .....
۴۱	۳-۳- نتیجه‌گیری نهایی .....

٢٣ ..... مراجع

# فصل اول

## مقدمه

### ۱-۱- مطالعات ارتباط کمی ساختار-فعالیت (خاصیت)

هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده مولکول انجام گیرد به آن ارتباط کمی ساختار- خاصیت (QSPR)<sup>۱</sup> می گویند مانند یافتن رابطه ای بین خواصی نظیر نقطه جوش و ساختار مولکول. وقتی از خواصی از نوع فعالیت بیولوژیکی (مانند فعالیت دارو) توصیف می شود اشاره به ارتباط کمی ساختار- فعالیت (QSAR)<sup>۲</sup> خواهد داشت. این ارتباط نیازمند اصول تئوری سخت و پیچیده نمی باشد و شامل مراحل زیر است [۱]:

- ۱- جمع آوری لیستی از ترکیبات.
- ۲- فراهم کردن ساختار هندسی.
- ۳- استخراج توصیفگرها
- ۴- انتخاب بهترین توصیف کننده ها.
- ۵- مدلسازی.
- ۶- ارزیابی مدل.

### ۱-۱-۱- جمع آوری لیستی از ترکیبات

اولین مرحله در توسعه مطالعات QSAR و QSPR جمع آوری لیستی از ترکیبات است که از لحاظ تجربی خواص محاسبه شده و شناخته شده داشته باشند. غالباً هزاران ترکیب

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

<sup>۲</sup>-Structure – activity relationship

در مطالعات QSAR و QSPR مورد استفاده قرار می‌گیرد. اگر تعداد ترکیبات نسبت به پارامترهای برازش شده<sup>۱</sup> در معادله کمتر باشد، این برازش صحیحی نمی‌باشد. اما اگر تعداد پارامترها و تعداد ترکیبات برابر باشد، برازش منحنی به طور صحیحی فراهم خواهد شد.

## ۱-۱-۲- بهینه سازی ساختار هندسی و استخراج توصیفگرها

مرحله بعدی فراهم کردن ساختار هندسی مولکولهاست که بهتر است از لحاظ ساختاری بهینه شوند. در این کار از شیمی محاسباتی<sup>۲</sup> استفاده می‌شود که در اینجا اشاره ای به شیمی محاسباتی و روشهای رایج در بهینه سازی ساختار مولکولها می‌شود. در سالهای اخیر روشهای محاسباتی در بین شیمیدانان بسیار رواج پیدا کرده است. در این روشها می‌توان به آسانی محاسبات خود را انجام داد، بدون اینکه از اصول اولیه و طریقه محاسبه، آگاهی و دانش لازم را داشت. شیمی محاسباتی ساختارها و واکنشهای شیمیایی را شبیه سازی<sup>۳</sup> می‌کند و این روش به شیمیدان اجازه می‌دهد که به مطالعه روشهای شیمیایی (با اجرای محاسبات با کامپیوتر) بپردازد، بدون اینکه از روشهای تجربی و آزمایشگاهی استفاده شود. روشهای محاسباتی در شیمی به چند دسته تقسیم می‌شوند:

### ۱-۱-۲-۱- روشهای آغازین<sup>۴</sup>

در این روش از هیچ داده تجربی استفاده نمی‌شود، بلکه از یک سری ثابت هایی نظیر بار الکترون، جرم الکترون و ثابت پلانک استفاده می‌شود و تمام معادلات لازم از حل معادلات کوانتومی از جمله، معادله شرودینگر<sup>۵</sup> بدست می‌آید به همین دلیل چون از هیچ داده تجربی استفاده نمی‌شود به آن « روش آغازین » گفته می‌شود [۲ و ۳].

<sup>۱</sup>-Fitted

<sup>۲</sup>- Computational Chemistry

<sup>۳</sup>- Simulation

<sup>۴</sup>-Ab initio methods

<sup>۵</sup>-Schrodinger Equation

اگر برای یک سیستم که تابع موج آن را داریم، تابع هامیلتونی<sup>1</sup> را به آن اثر دهیم می‌توانیم انرژی آن را به دست آوریم:

$$H\psi = E\psi \quad (1-1)$$

که در این معادله H معرف تابع هامیلتونی،  $\psi$  بیانگر تابع موج و E نشان دهنده انرژی سیستم مورد نظر می‌باشد. هر تابع هامیلتونی خود شامل دو قسمت انرژی پتانسیل و جنبشی می‌باشد. فرض کنید یک الکترون و یک هسته داریم، انرژی پتانسیل به فاصله الکترون از هسته و بار آنها بستگی دارد بنابراین وقتی معادله را می‌نویسیم در اوپراتور پتانسیل یک تابع فاصله به شکل  $\frac{1}{r}$  وجود دارد.

حال اگر تعداد الکترونها بیشتر شود در این صورت به غیر از یک فاصله و انرژی پتانسیل که ذکر شد یک فاصله و انرژی پتانسیل دیگر هم بین دو الکترون به وجود می‌آید، پس یک  $\frac{1}{r_{1,2}}$  ظاهر می‌شود.

برای حل این مشکل از روش هارتری - فاک<sup>2</sup> استفاده می‌شود، در اینجا بر همکنش دو الکترون مستقیماً در نظر گرفته نمی‌شود، بلکه از اوپراتور فاک استفاده می‌شود یعنی الکترون دوم به صورت ابر الکترونی در نظر گرفته می‌شود و بر همکنش الکترون اول را با ابری از الکترون دوم در نظر گرفته می‌شود. اما مشکلی که در این روش وجود دارد این است که بر همکنش لحظه‌ای بین الکترون اول و الکترون دوم در نظر گرفته نمی‌شود.

در روش هارتری - فاک باید نوع هارتری - فاک به کار برده شده را مشخص کرد یعنی باید مشخص شود که هارتری - فاک محدود شده<sup>3</sup> یا هارتری = فاک نامحدود<sup>4</sup> به کار رفته است.

وقتی از اوربیتالی که دو الکترون دارد صحبت می‌شود، یا می‌توان برای هر دو الکترون یک تابع موج یکسان تعریف کرد، که به این حالت هارتری - فاک محدود شده گفته می‌شود و

<sup>1</sup> -Hamiltonian function

<sup>2</sup> -Hartree-fock

<sup>3</sup> -Restricted hartree-fock

<sup>4</sup> - Unrestricted hartree-fock



یا اینکه برای هر کدام از الکترون‌ها یک تابع موج خاص تعریف کرد، که به این حالت هارتری-فاک نامحدود گفته می‌شود. در هارتری-فاک محدود شده حجم محاسبات به نصف کاهش می‌یابد، ولی مزیت هارتری-فاک نامحدود این است که اگر اختلافی بین تابع موج دو الکترون وجود داشته باشد را می‌توان مشاهده نمود. معمولاً اگر لایه آخر پر باشد تفاوت چندانی بین دو الکترون وجود ندارد، لذا از هارتری-فاک محدود شده استفاده می‌شود.

#### ۱-۱-۲-۲- روشهای نیمه تجربی<sup>۱</sup>

روشهایی هستند که در برنامه‌هایی نظیر MOPAC و Hyper Chem وارد شده‌اند و به شکل مکانیک کوانتومی می‌باشند. در این روشها یک سری از مقادیر تجربی نیز وارد شده‌اند و فقط الکترونهای لایه ظرفیت در محاسبات در نظر گرفته می‌شوند. ویژگی بارز این روشها این است که زمان محاسبات در این روشها کوتاهتر از روشهای آغازین می‌باشد. اما عیب این روشها این است که پیشگویی خیلی صحیح و خوبی را نمی‌توانند داشته باشند. این روشها شامل چندین دسته می‌باشند که در این پروژه از روش AM1<sup>۲</sup>: استفاده شده است. روش AM1 همان روش MNDO<sup>۳</sup> اصلاح شده می‌باشد و روی مولکولهای قطبی نتایج خوبی را می‌دهد اما با پیوند هیدروژنی مشکل دارد. معمولاً گرمای تشکیل و انرژی اکتیواسیون را بسیار دقیقتر از MNDO پیش‌بینی می‌کند.

#### ۱-۱-۲-۳- روشهای تابع دانسیته<sup>۴</sup>

در این روش به جای استفاده از تابع موج، از دانسیته الکترونی استفاده می‌شود و مزیتی که دارد این است که وابستگی الکترون‌ها را در نظر می‌گیرد، این روش حجم محاسباتی کمی دارد و برای کارهای کاربردی نظیر تعیین ساختار و به دست آوردن طیفهای NMR و IR استفاده می‌شود.

<sup>۱</sup>-Semi-empirical methods

<sup>۲</sup>-Austin methods

<sup>۳</sup>- Modified neglect of diatomic overlap

<sup>۴</sup>-Density functionl theory(DFT)

## ۱-۲-۲-۴- روشهای مکانیک مولکولی

در این روش هیچ ماهیت کوانتومی وجود ندارد و اتمها به صورت گلوله و پیوندها صورت فنر در نظر گرفته می‌شوند و بهینه شدن ساختار با توجه به ارتعاشات و حرکتهایی که وجود دارد انجام می‌شود.

## ۱-۱-۳- استخراج توصیفگرها

پس از بهینه سازی ساختار مولکولها بایستی توصیف کننده‌های وابسته به ساختار مولکول استخراج شوند. توصیفگرهایی که در مطالعات QSAR مورد استفاده قرار می‌گیرند شامل دو دسته توصیفگرهای تئوری و توصیفگرهای تجربی می‌باشند. توصیفگرهای تجربی از طریق آزمایش اندازه گیری می‌شوند بنابراین این توصیفگرها برای دسته ترکیبات محدودی در دسترس هستند در حالیکه توصیفگرهای تئوری بدون نیاز به دادهای تجربی بر راحتی به کمک نرم افزارهای مناسب برای طیف وسیعی از ترکیبات چه واقعی و چه تئوری قابل محاسبه هستند بنابراین در مطالعات QSAR از اهمیت زیادی برخوردار هستند.

## ۱-۱-۴- انتخاب بهترین توصیف کننده‌ها

مرحله بعدی در مطالعات QSAR انتخاب بهترین توصیف کننده‌های محاسبه شده می‌باشد زیرا توصیف کننده‌های زائد کار برآزش و مدلسازی را طولانی می‌کنند و تأثیری در بهبود نتایج نخواهند داشت. یکی از روشهای انتخاب توصیفگرهای مناسب روش برآزش مرحله ای<sup>۱</sup> میباشد که معمولاً با محاسبه ضرایب همبستگی مشخص انجام می‌شود. اگر توصیف کننده ای ضریب همبستگی برابر یک داشته باشد، بدان معنا است که می‌تواند خواص مورد نظر را به درستی توصیف کند. و ضریب همبستگی صفر بدین معنی است که هیچ ارتباطی بین توصیف

<sup>۱</sup>-Molecular mechanics

<sup>۲</sup>-Intercorrelation coefficient

کننده و خاصیت اصلی وجود ندارد. پس توصیف کننده‌هایی با ضرایب همبستگی بزرگ دربرازش منحنی<sup>۱</sup> برای ساختن مدل جهت پیشگویی خواص استفاده می‌شوند.

## ۱-۱-۵- مدلسازی

یکی از مهمترین مراحل در مطالعات QSAR و QSPR ساختن مدلی است که قادر باشد با استفاده از توصیفگرهای مولکولی خاصیت یا فعالیت مورد نظر را پیشگویی نماید. روشهای متفاوتی در مدلسازی بکار گرفته می‌شود که میتوان آنها را در دو دسته کلی زیر قرار داد:

الف) استفاده از روشهای آماری چند متغیره: در این شیوه از روشهایی کالیبراسیون چند متغیره نظیر رگرسیون خطی چند گانه [۴-۷] و روش حداقل مربعات جزئی<sup>۲</sup> [۸-۱۱] استفاده می‌شود. این روشها خطی می‌باشند و وقتی کارایی مناسبی دارند که داده‌ها دارای رفتار خطی باشند. یک رگرسیون با ایجاد معادله خطی انجام می‌شود:

$$property (activity) = c_0 + c_1d_1 + c_2d_2 \quad (1-1)$$

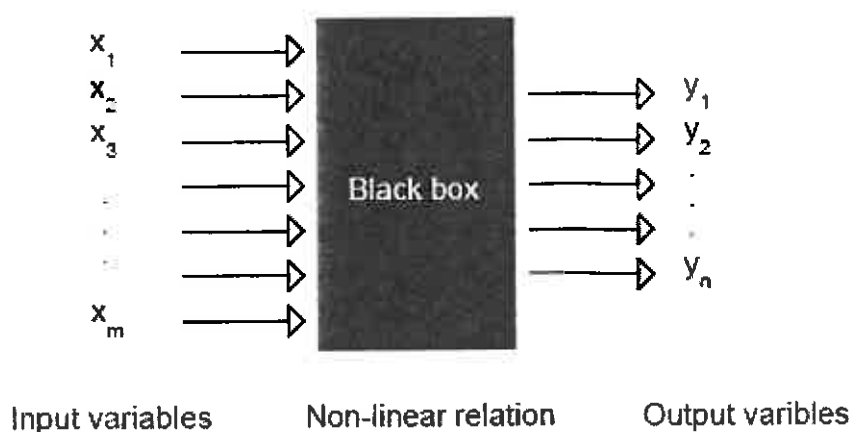
در این معادله  $c_i$  بیانگر پارامترهای برازش شده و  $d_i$  بیانگر توصیف کننده‌ها می‌باشد. دلیل استفاده از معادلات خطی، سادگی روش می‌باشد.

ب) استفاده از روشهای غیر خطی: ممکن است از یک برازش غیر خطی در مدلسازی استفاده شود که این کار جهت کم کردن انحراف بین مقدار برازش شده و مقادیر صحیح می‌باشد. از میان روشهای غیر خطی، روش استفاده از شبکه های عصبی مصنوعی به طور گسترده در مطالعات QSAR و QSPR بکار گرفته شده است که در اینجا به طور اجمالی شرح داده خواهد شد.

<sup>۱</sup>-Regression

<sup>۲</sup>-Partial least square

شبکه عصبی را می توان به صورت جعبه سیاهی در نظر گرفت که می تواند یک سری از داده ها را به صورت ورودی دریافت کند و یک یا چند خروجی تولید نماید (شکل ۱-۲).



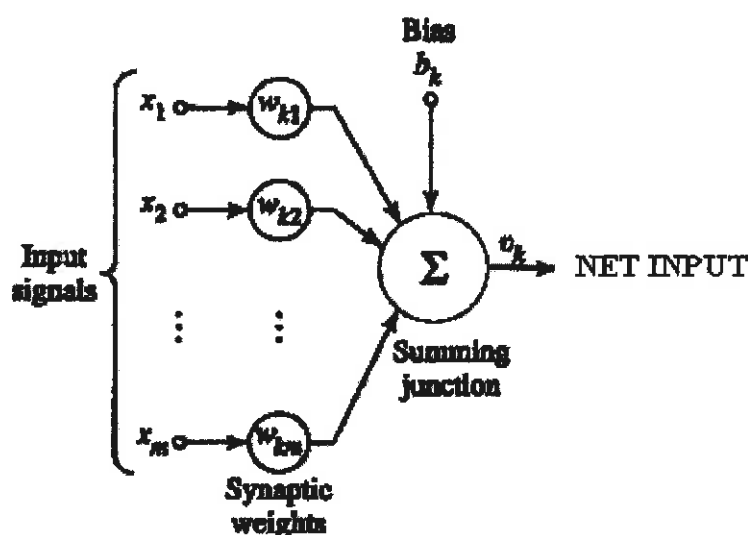
شکل (۱-۲) - نمایش یک شبکه عصبی نوعی

ورودی های زیادی را می توان برای شبکه در نظر گرفت. به عنوان مثال ورودی می تواند اطلاعات پزشکی مربوط به یک بیمار باشد و خروجی شبکه نوع مرضی وی باشد. یا از روی طیف یک ترکیب به عنوان ورودی می توان ساختار آن را (به عنوان خروجی) پیش بینی کرد.

خیلی از استفاده کنندگان از شبکه های عصبی با موفقیت از شبکه های عصبی استفاده می کنند بدون اینکه نیازی باشد که بدانند درون این جعبه سیاه چه اتفاقاتی می افتد. هدف ما در این نوشتار آشنایی نسبی با عملکرد این جعبه، اجزای درون آن و اتفاقاتی است که درون آن روی می دهد.

یک نرون به عنوان به عنوان کوچکترین واحد پردازشگر اطلاعات است که اساس عملکرد ANNs را تشکیل می دهد. نرون مصنوعی برای تقلید مشخصات اولیه نرون بیولوژیکی طراحی شد. از نظر مفهومی مجموعه ای از داده ها به عنوان ورودی نرون (که شاید خروجی نرونهای دیگر باشند) بکار برده می شوند. هر ورودی در وزن متناظرش ضرب می شود

(نظیر Synaptic Strength) و بعد همه ورودی های وزین شده به منظور تعیین سطح فعالیت نرون با هم جمع می شوند. شکل ۲-۲ مدلی را نشان می دهد که این ایده را به کار برده است.



شکل (۲-۲) - مدل نرون مصنوعی

$$NET = X_1 W_{K1} + X_2 W_{K2} + \dots + X_m W_{km} \quad (2-1)$$

ورودی متناظر وزن بایاس در واقع برای انتقال مقدار خروجی به میزان مناسب در مجموعه فضای خروجی می باشد. یادگیری برای نرون فوق بدین معنی است که پارامترهای  $W$  و  $b$  طوری تغییر می کنند که رابطه ورودی - خروجی نرون با هدف خاص مطابقت نماید.

با وجود تنوع نمونه های شبکه، تقریباً همه آنها روی این ساختار بنا نهاده شده اند. در اینجا ورودی ها که با  $X_1, X_2, \dots, X_n$  نشان داده شده اند و مجموعاً با بردار  $X$  نمایش داده می شود، متناظر با سیگنال ها در نرون بیولوژیکی می باشد. هر سیگنال به وزن وابسته  $W_1, W_2, \dots, W_n$  ضرب شده، بعد در بلوک  $\Sigma$  به کار برده می شود. هر وزن که با بردار  $W$  نمایش داده شده است متناظر با مقاومت اتصال سیناپتیکی بیولوژیکی است. بلوک جمع بندی

$\Sigma$  ورودی های وزین شده را به صورت جبری جمع می کند و یک خروجی ایجاد می کند که با NET نمایش داده شده است

### ۱-۱-۶- ارزیابی مدل

اعتبار و اهمیت معادلات پیشگویی وقتی است که خواص مولکول های مجهولی که در دسته آموزش موجود نیستند را پیشگویی کنند. یک معادله که به خوبی در دسته آموزش قادر به پیشگویی می باشد ممکن است برای مولکول های خارج از این دسته به چند دلیل به خوبی پیشگویی نکند. برخی از دلایل این امر عبارتند از: استفاده از مولکول های محدود شده در مجموعه آموزش و یا استفاده از مولکول های بسیار شبیه به هم در مجموعه آموزش. برای مثال یک معادله که برای مولکول های آلی آموزش دیده است به طور ناچیزی خواص مواد معدنی را پیشگویی کند.

### ۱-۲- بیماری ایدز

ایدز یا سندرم نقص ایمنی اکتسابی<sup>۱</sup> (AIDS) عبارت است از بوجود آمدن ضعف عمده در دستگاه ایمنی بدن (نقص ایمنی). این امر باعث کاهش توانایی بدن در مقابله با عفونتها و توانایی سرکوب سلول های غیرطبیعی مثل سلول های سرطانی می شود. ویروس ایدز سلول های ایمنی موجود در خون (لنفوسیت ها) و سلول های ایمنی موجود در بافت ها مانند مغز استخوان، طحال، کبد و گره های لنفاوی را درگیر می سازد. این سلول ها در تولید پادتن برای مقابله با بیماریها و سرطانها نقش دارند. در مجموع باید گفت که ایدز یک نوع نقص ایمنی ثانویه است که در سیر عفونت با ویروس ایدز ایجاد می شود.

---

<sup>1</sup> - Acquired Immunodeficiency Syndrome

عامل بیماری ایدز یک نوع ویروس از گروه رتروویروس ها<sup>۱</sup> بنام ویروس نقص ایمنی انسانی<sup>۲</sup> (HIV) است که باعث کاهش توانایی سیستم ایمنی بدن میزبان می‌شود. علایمی که در بیماری ایدز مشاهده می‌شود مربوط به بیماریهایی است که در اثر نقص دستگاه ایمنی بدن تولید می‌شوند. در اکثر بیماریهای ویروسی وضع به این منوال است که سلولهای مملو از ویروس ویروسها را آزاد کرده و این ویروسها خود را با پادتنهای آماده مواجه می‌بینند. در چنین وضعی بیماری شخص برطرف می‌شود. اما در مورد ویروس ایدز وضع به گونه‌ای دیگر است. در اینجا اولین ویروسی که وارد اولین سلول میزبان می‌شوند از حمله پادتنها در امان می‌مانند. ویروسهای جدیدی که از سلول خارج می‌شوند بعضاً توسط پادتنها خنثی می‌شوند اما ژنهای اولیه در درون سلول میزبان به تولید ویروس ادامه می‌دهند. ممکن است که ویروس در بدو ورود به سلول میزبان به صورت غیر فعال درآید و بعد از گذشت چندین سال به مساعد شدن اوضاع فعالیت بیماریزایی خود را آغاز کند. این ویروس دستگاه ایمنی بدن را تضعیف می‌کند و در این یک سرماخوردگی ساده هم می‌تواند برای شخص مشکل ساز باشد. [۱۲]

## ۱-۲-۱- تاریخچه کشف ویروس HIV

در سال ۱۹۸۱، هشت مورد وخیم ابتلا به بیماری سارکوم کاپوسی<sup>۳</sup> که نوعی سرطان بدخیم عروقی است، در میان مردان همجنس‌گرای نیویورک گزارش شد. این نوع سرطان نادر است و معمولاً سالمندان را گرفتار می‌کند. همزمان با این رویداد، شمار مبتلایان به نوعی عفونت نادر ریوی نیز در نیویورک و کالیفرنیا افزایش یافت. مسئولان بهداشت و درمان آمریکا در آن زمان علت شیوع ناگهانی این دو بیماری را نمی‌دانستند؛ اما امروزه این دو واقعه پزشکی مصادف با تولد بیماری ایدز در زندگی بشری در نظر گرفته می‌شود. طی یک سال، این بیماری ناشناخته گسترش یافت و در سال ۱۹۸۲، AIDS به معنی «سندروم نقص ایمنی اکتسابی» نام

<sup>۱</sup> -Retroviruse

<sup>۲</sup> -Human Immunodeficiency Virus (HIV)

<sup>۳</sup> - Kaposi's Sarcoma

گرفت. رفته‌رفته، این بیماری نه تنها همجنس‌گرایان، بلکه گروه‌های مختلفی از مردم را مبتلا کرد. بیماران هموفیلی و معتادان گروه‌های بعدی بودند. در این سال‌ها، بیماری جدیدی در فقیرترین قاره جهان به معضلی جدی تبدیل شد. این بیماری را آفریقایی‌ها در زبان بومی اسلیم<sup>۱</sup> به معنی مرگ در نتیجه تحلیل تدریجی بدن می‌نامیدند، به دنبال تحقیقات بیشتر مشخص شد که این بیماری همان ایدز است. در ۱۹۸۴، تلاش دانشمندان و محققان برای کشف علت بیماری کشنده ایدز به نتیجه رسید. گروهی تحقیقاتی به سرپرستی رابرت گالو<sup>۲</sup> ادعا کرد که عامل این بیماری را شناسایی کرده است. البته این گفته به شدت با اعتراض محققان فرانسوی که ماه‌ها قبل ویروس HIV را شناسایی کرده بودند رو به رو شد. [۱۳]

براساس آخرین تحقیقات بر روی ویژگی‌های ژنتیکی ویروس HIV، این ویروس از ترکیب دو ویروس مختلف در شامپانزه به وجود آمده است. ویروس HIV از دیدگاه علم ویروس شناسی به گروهی از ویروس‌ها تعلق دارد که در میمون‌ها بیماری‌زا هستند. یافته ژنتیکی مذکور این نظریه را که ایدز طی یک قرن گذشته در جنگل‌های غرب آفریقا ظهور کرده است، تأیید می‌کند. به اعتقاد دانشمندان، انسان برای اولین بار در نیمه اول قرن گذشته در نتیجه شکار و تغذیه گوشت شامپانزه، سنتی که هنوز در آفریقا ادامه دارد، به این ویروس آلوده شده است. بسیاری از دانشمندان بر این باورند که انتقال ویروس بیش از یک بار اتفاق افتاده است، زیرا انواع متفاوتی از این ویروس باعث بروز بیماری ایدز در انسان می‌شود. جالب آنکه هرچند نخستین مورد ابتلا به ایدز در ۱۹۸۱ در آمریکا گزارش شده، شواهد نشان می‌دهد که اولین قربانی ایدز در این کشور یک جوان سیاه پوست بوده که در ۱۹۶۹ در گذشته استدرسال ۱۹۸۱ یک سندرم بالینی جدید که با نقص ایمنی واضحی مشخص می‌شود، ثبت شد. که نامیده شد.

## ۱-۲-۲- انواع HIV

<sup>۱</sup> -Slim

<sup>۲</sup> - Robert Gallo



در حال حاضر ۲ گونه از HIV وجود دارد:

HIV-1 (الف)

HIV-2 (ب)

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

### ۱-۲-۳- میزان پایداری ویروس HIV

ویروس HIV در اثر خشک شدن به سرعت غیر فعال می شود در محیط کشت عاری از سلول و در درجه حرارت اتاق تا پانزده روز فعال بوده و در همان محیط در حرارت ۳۷ درجه تا یازده روز زنده می ماند. در درجه حرارت ۶۰-۵۰ درجه سانتیگراد به مدت سی دقیقه عفونت زایی خود را از دست می دهد. همچنین این ویروس در اثر استفاده از بعضی عوامل به مدت ده دقیقه در درجه حرارت اتاق به طور کامل غیر فعال می شود. این عوامل هیپوکلریت سدیم ۰/۵٪، اتانول ۷۰٪، پارافرمالدئید ۰/۵٪، و قرار گرفتن در PH و بیش از ۱۰ را شامل می شوند [۱۲].

### ۱-۲-۴- همانند سازی و چرخه تکثیر

ویروس HIV یک رتروویروس است که دستخوش جهش های مکرر می شود و بر سلول های دستگاه ایمنی بدن حمله می کند. رتروویرویس ها ویروسهایی هستند که ژنوم (ماده ژنتیکی ویروس) آنها از RNA تشکیل شده است. بنابراین برای تکثیر خود به آنزیمی به

نام نسخه بردار معکوس<sup>۱</sup> وابسته هستند که ژنوم RNA آنها را به DNA نسخه برداری می کند تا بعد بتواند آن را با کمک آنزیمی (آنزیم اینتگراز) وارد ژنوم سلول میزبان کند و به این ترتیب امکان تکثیر ویروس به وجود می آید. کلمه رترو (معکوس) در نام این نوع ویروس ها هم به همین خاطر است چرا که معمولاً نسخه برداری از DNA به RNA انجام می شود و لی در این دسته ویروسها بر عکس است.

ویروس HIV با آلوده کردن سلول های ویژه ای از سیستم ایمنی به نام "لنفوسیت های T کمک کننده"<sup>۲</sup> یا CD4<sup>+</sup> یا TCD4<sup>+</sup> باعث بیماری می شود. این سلول های زیرگروهی از گلبول های سفید هستند که به طور طبیعی پاسخ ایمنی به عفونت را تنظیم می کنند. HIV با استفاده از سلول های T برای تکثیر خودش در سراسر بدن گسترش می یابد و در همان زمان باعث کاهش این سلول ها می شود که بدن برای دفاع از خود به آنها نیاز دارد [۱۲].

ویروس HIV به سلول های TCD4<sup>+</sup> از طریق مولکولهای خاصی (CXCR4) یا هر دو مولکول CXCR4 و CCR5 بسته به مرحله عفونت متصل می شوند. در مراحل اولیه عفونت HIV دو گیرنده (CCR5 و CXCR4) محل اتصال ویروس هستند اما در مراحل انتهایی عفونت که اغلب HIV دچار جهش می شود آنها تنها به یک گیرنده (CXCR4) متصل می شوند. هنگامی که HIV به سلول های TCD4<sup>+</sup> متصل می شود یک ساختار ویروسی (به نام GP۴۱) به داخل غشای سلول نفوذ می کند و RNA ویروس و آنزیم های مختلف (از جمله نسخه بردار معکوس، اینتگراز و پروتئاز) به داخل سلول تزریق می شوند. مرحله بعدی تولید DNA از روی RNA ویروس با کمک آنزیم مخصوصی (نسخه بردار معکوس) است. در صورت موفقیت این عمل، DNA اولیه ویروسی (پروویروس<sup>۳</sup>) با استفاده از آنزیمی (آنزیم اینتگراز) وارد DNA سلول میزبان می شود (مرحله الحاق<sup>۴</sup>). در این حال سلول میزبان کامل با HIV آلوده شده است اما به صورت فعال پروتئین های ویروس را تولید نمی کند. از این به بعد دوره

<sup>۱</sup> -Reverse Transcriptase

<sup>۲</sup> - T-Lymphocytes

<sup>۳</sup> - Provirus

<sup>۴</sup> -Integration

نهفته آغاز می شود که در آن سلول های آلوده مانند «بمب هایی منفجر نشده» برای مدت طولانی باقی می مانند. هنگامی که سلول میزبان تولید پروتئین های ویروس را از روی DNA پروویروسی آغاز می کنند، آنزیم خاصی (آنزیم پروتئاز) فراهم شده به وسیله HIV باید آنها را به صورت پروتئین های نوبنیاد HIV درآورند تا با اتصال آنها به هم ذرات ویروسی HIV به وجود آید. ذرات ویروسی تازه به وجود آمده با جوانه زدن بر روی سطح سلول میزبان از آن خارج می شوند [۱۳-۱۵].

### ۱-۲-۵- بیماری زایی و یافته های ایمنی

عفونت با HIV موجب واکنش ویروس با گیرنده  $CD4^+$  می گردد که با تراکم زیاد بر روی لنفوسیت های T کمکی قرار دارند. بعضی از مونوسیتها و ماکروفاژها نیز به میزان کمتری دارای ملکول  $CD4^+$  هستند، بنابراین می توانند به وسیله HIV آلوده گردند. سلولهای T4 آلوده به ویروس در سطح خود مقدار زیادی گلیکوپروتئین پوششی به نام gp 120 را تولید می کنند. که این امر منجر به اتصال آنها به سایر سلولهای T4 که آلوده به ویروس نیستند شده و در نتیجه سلولهای موجب حذف سریع تعداد زیادی از سلولهای غیر آلوده گردش خون می شود. پیامدهای ناشی از آلودگی لنفوسیت های T کمکی به وسیله ویروس HIV بسیار مخرب است زیرا این سلولها نقش مهمی را در دستگاه دفاعی بدن ایفا می کنند.

### ۱-۲-۶- مراحل بیماری

دوره کمون<sup>۱</sup> بیماری چندروز تا سه ماه است. عفونت با HIV، به سه مرحله تقسیم می شود [۳۶]:

۱-۲-۶-۱- مرحله مقدماتی (مرحله حاد)

<sup>۱</sup> \_Prion of Incubation

در این مرحله علائم بالینی عبارتند از : خستگی، احساس کسالت، سردرد، کاهش شدید وزن، تب و اسهال مزمن. در این مرحله ویروس در پلاسما مشخص بوده و تا ۱٪ لنفوسیت‌های  $CD4^+$  خون محیطی آلوده هستند. به نظر می‌رسد که تمام سلول‌های آلوده فعالانه در حال تکثیر ویروس نیستند.

حدود یک ماه پس از ظهور مرحله حاد علائم فوق فروکش می‌کند و در انتهای این مرحله، تعداد لنفوسیت‌های  $CD4^+$  نیز به میزان طبیعی برمی‌گردد. مرحله مقدماتی با پیدایش بادی در برابر آنتی ژن‌های ویرو پایان می‌یابد.

#### ۱-۲-۶-۲- مرحله بدون علامت (مزمن)

بعد از مرحله حاد دوره بدون علامت وجود دارد که ۱۱-۷ سال طول می‌کشد این دوره که در گذشته به نام دوره نهفتگی نامیده می‌شده است در حقیقت عفونت خاموشی را نمایش می‌دهد که طی آن تخریب مداوم و دائمی سیستم ایمنی انجام می‌گردد. در این دوره تعداد لنفوسیت‌های  $CD4^+$  آلوده افزایش می‌یابد و اثرات تخریب سیستم ایمنی مشخص‌تر می‌شود. پاسخ‌های ایمنی سلولی در این دوره به طور فزاینده‌ای کاهش می‌یابد و فرد آلوده قادر نیست که نسبت به آنتی ژن‌های میکروبی پاسخ دهد. بایستی خاطر نشان کرد که ویروس در این مرحله و در طی مرحله سوم خصوصیات را که در شروع بیماری داشت، دارا نمی‌باشد. در دوره مزمن بیماری HIV سریع‌تر تکثیر می‌یابد، محدوده سلولی آن فراتر است، برای  $CD4^+$  ها مخرب‌تر می‌باشد.

#### ۱-۲-۶-۳- مرحله ایدز (بحرانی)

پیشگویی آمایشگاهی در مورد پیشرفت بیماری به سوی ایدز به وسیله لنفوسیت‌های  $CD4^+$  انجام می‌پذیرد. هنگامی که تعداد این لنفوسیت‌ها خیلی پائین بیاید میزبان در معرض مبتلا به عفونت‌های فرصت طلب قرار می‌گیرد و اگر این تعداد بین ۴۹۹-۲۰۰ سلول در هر میکرولیتر باشد و علائم بالینی مختلفی مشاهده گردد، پزشکان بایستی برای بیماری برنامه

درمانی دقیقی را ترتیب دهند. هنگامی که شمارش لنفوسیت‌های  $CD4^+$  به کمتر از  $200$  سلول در هر میکرولتیر برسد، این افراد در طبقه بیماران ایدزی قرار می‌گیرند.

## ۱-۲-۷- آمار مبتلایان به ایدز

براساس تازه‌ترین آمار سازمان ملل متحد درباره گسترش بیماری ایدز در جهان، هم‌اکنون نزدیک به ۳۸ میلیون نفر به ویروس HIV آلوده‌اند. این ویروس از زمان کشف تاکنون بیش از ۲۰ میلیون قربانی داده است. همچنین این تحقیق نشان می‌دهد که بیماری ایدز به سرعت در تمامی مناطق جهان در حال گسترش است. فقط در سال گذشته میلادی، نزدیک به پنج میلیون نفر به بیماری ایدز مبتلا شده‌اند. میزان شیوع بیماری ایدز در مناطق جهان متفاوت است. ساکنان برخی از کشورها بیش از کشورهای دیگر به ایدز مبتلا شده‌اند و در سطح یک کشور، حتی بین استان‌ها، تفاوت وجود دارد. اما سازمان ملل تعداد افراد آلوده به ویروس ایدز در قاره آسیا را در حدود ۷/۴ میلیون نفر اعلام کرده است. اعتقاد بر این است که در حدود نیم میلیون نفر از جمعیت قاره کهن در سال گذشته به علت ابتلا به ایدز جان خود را از دست داده‌اند و ۱/۱ میلیون نفر به ویروس آن آلوده شده‌اند. در میان مبتلایان سال گذشته، ۰/۳٪ از زنان و ۰/۴٪ از مردان در گروه سنی ۱۵ تا ۲۴ سال قرار دارند. جالب آنکه میزان آلودگی مردم به ویروس HIV در دو کشور پرجمعیت جهان یعنی چین و هند بسیار کم است.

در ایران، براساس آخرین آمار وزارت بهداشت، درمان و آموزش پزشکی، تعداد افراد آلوده به HIV، تا ابتدای تیر ۱۳۸۳، ۷۱۰۸ نفر بوده است. از این تعداد، ۶۷۶۱ نفر (۹۵/۱٪) مرد و ۳۴۷ نفر (۴/۹٪) زن هستند. از تعداد ۶۷۶۱ نفر، ۱۱ پسر و ۱۳ دختر زیر ۴ سال (در مجموع، ۰/۵٪ کل افراد آلوده) همچنین ۱۵ مرد و ۴ زن بالای ۶۵ سال (به ترتیب ۰/۳٪ و ۱/۴٪) به این ویروس آلوده‌اند. براساس همین نمونه گیری، بیشترین موارد آلودگی به HIV در رده سنی ۲۵ تا ۳۴ سال است. به نحوی که ۱۹۱۳ مرد آلوده و مبتلا (۴۲/۳٪) و ۹۴ زن آلوده و مبتلا به ایدز (۳۳/۶٪) در این رده سنی گزارش شده‌اند. بر طبق گزارش وزارت بهداشت، درمان و

آموزش پزشکی، بیشترین راه انتقال این بیماری در کشور اعتیاد تزریقی است که ۴۱۰۷ نفر (۵۷/۸٪) از این طریق مبتلا شده‌اند. پس از آن، ۲۲۹۲ نفر (۳۲/۲٪) به شیوه‌های نامشخص، ۴۷۶ نفر (۶/۷٪) از راه جنسی و ۲۰۱ نفر (۲/۸٪) از طریق خون و فرآورده‌های خونی به این بیماری مبتلا شده‌اند. در ۳۲ مورد (۰/۵٪) نیز انتقال از طریق مادر بوده است. وزارت بهداشت تعداد موارد فوت‌شده بر اثر بیماری ایدز را تاکنون ۷۷۱ نفر اعلام کرده است که از این جمعیت، ۷۴۱ نفر مرد و ۳۰ نفر زن بوده‌اند. اما آمارهای اعلام شده و رسمی موجود درباره بیماری ایدز تعداد کمی از جمعیت واقعی افراد آلوده به ویروس HIV را شامل می‌شود. دکتر محمد مهدی گویا، معاون وزارت بهداشت، درمان و آموزش پزشکی و کارشناس مسائل ایدز، پس از انتشار گزارش اخیر سازمان ملل متحد درباره وضع ایدز در جهان، در گفت‌وگو با یکی از جراید روز، تعداد افراد آلوده به ویروس این بیماری را در ایران ۳۰ تا ۳۵ هزار نفر اعلام کرده است.

### ۱-۲-۸- پیشگیری و درمان

در حال حاضر واکنش مناسبی برای پیشگیری از ایدز در دسترس نیست و پیشگیری ایدز به وسیله آموزش انجام می‌گیرد. یکی از مشکلات عمده برای درمان بیماران مبتلا به ایدز این است که این قبیل افراد میزبان عوامل میکروبی فرصت طلب می‌باشند. برای درمان بیماری ایدز داروهای مختلفی وجود دارد که این داروها به دسته‌های زیر تقسیم می‌شوند.

#### الف) مهارکننده نوکلئوزیدی<sup>۱</sup> (NRTIs)

این دسته دارویی از سال ۱۹۸۷ در درمان HIV موجود است و بیشتر بعنوان Nucleosid analogues یا nukes شناخته می‌شود. HIV به آنزیم reverse Transcriptase برای تکثیر خود نیاز دارد و این دسته دارویی، این آنزیم را مهار می‌کند و روند تکثیر آنزیم را آهسته کرده و HIV قادر به آلوده کردن سلولها و تکثیر خود نیست [۱۶ و ۱۷].

---

<sup>۱</sup>-Nucleosid Reverse Transcriptase Inhibitors

### ب) مهار کننده غیر نوکلئوزیدی<sup>۱</sup> (NNRTIs)

این دسته داروها از سال ۱۹۹۷ تأیید شده‌اند و بطور کلی بعنوان غیر نوکلئوزیدی شناخته می‌شوند. این دسته دارویی بوسیله تداخل با Transcriptase جلوی تکثیر ویروس HIV را می‌گیرند داروهای غیر نوکلئوزیدی تا حدودی متفاوت از مشابه های نوکلئوزیدی عمل می‌کنند و با روشی متفاوت، جلوی آلوده کردن سلولها بوسیله ویروس HIV را می‌گیرند. [۷]

### ج) مهار کننده آنزیم پروتئاز<sup>۲</sup> (PI)

این دسته دارویی اولین بار در سال ۱۹۹۵ تأیید شدند. پروتئاز تقریباً در هر سلول زنده‌ای وجود دارد و یک آنزیم گوارشی است که پروتئین را تجزیه می‌کند و یکی از آنزیمهایی است که HIV برای تکثیر خود بکار می‌برد، پروتئاز در HIV به زنجیره طویل و سالم آنزیمها و پروتئینها در سلولها حمله کرده و آنها را به قطعات کوچکتری تقسیم می‌کند این قطعات کوچک آلوده پروتئین و آنزیم به آلوده کردن سلولهای جدید ادامه می‌دهد. مهار کننده پروتئاز قبل از اینکه آنزیم پروتئاز شانس برای تقسیم کردن پروتئینها و آنزیمها داشته باشد عمل می‌کند. در این روش آنزیم مهار کننده پروتئاز، تکثیر ویروس و در نتیجه آلوده کردن سلولهای جدید را آهسته می‌کند و NRTIs و NNRTIs فقط بر روی سلولهایی که جدیداً آلوده شده‌اند موثر است آنزیم مهار کننده پروتئاز قادر است فرآیند تبدیل ویروس نابالغ غیر عفونی به ویروس بالغ عفونی را آهسته کند. آنزیم مهار کننده پروتئاز در سلولهایی که مدت طولانی آلوده شده‌اند اثر خود را از طریق آهسته کردن تکثیر ویروس اعمال می‌کند.

### د) مهار کننده ورود<sup>۳</sup>

این دسته دارویی در حال انجام مطالعات بالینی در انگلستان و آمریکا می‌باشد و در شرف گرفتن تأییدیه است. در سطح HIV پروتئینهایی بنام gp-۴۱ و gp-۱۲۰ وجود دارند که بوسیله آنها HIV به سلول متصل شده و وارد سلول می‌گردد، با بلوک کردن یکی از این

<sup>۱</sup>-Non- Nucleosid Reverse Transcriptase Inhibitors

<sup>۲</sup>- Protease Inhibitors

<sup>۳</sup>- Fusion or Entry Inhibitors

پروتئینها، تکثیر ویروس آهسته می‌شود. بعنوان مثال T-20 که مهار کننده ورود است و در شرف گرفتن تأییدیه است به پروتئین gp41 متصل می‌گردد. T-20 یک پروتئین است و باید بصورت تزریقی مصرف شود و از طریق خوراکی نمی‌توان مصرف کرد چونکه در معده تجزیه می‌شود.

### ۱-۳- مروری بر کارهای انجام شده

درسال ۱۹۹۱ کوکلا<sup>۱</sup> پیشگویی فعالیت دارویی ضد HIV-1 مربوط به سه سری مختلف از مشتقات TIBO را با استفاده از سه معادله خطی را گزارش داده است [۳].

درسال ۱۹۹۷، ترونکت<sup>۲</sup> با استفاده از مطالعات QSAR برهمکنش HERT را با HIR، با استفاده از شبکه عصبی مصنوعی پیشگویی نمود [۱۹].

کای<sup>۳</sup> و همکارانش درسال ۱۹۹۸ با استفاده از مدل شبکه عصبی سایت‌های گسسته پروتئاز مربوط به HIV-1 در پروتئین را پیشگویی کردند [۲۰].

جلالی هروی و همکارش درسال ۲۰۰۰ در یک مطالعه QSAR، فعالیت ترکیبات ضدایدز را برای یک گروه بزرگ از مشتقات HEPT بررسی کردند. در این کار روش رگرسیون خطی با شبکه عصبی مقایسه شده است که در نتیجه مدل شبکه عصبی به عنوان پیشگویی کننده بهتری نسبت به مدل رگرسیون خطی معرفی گردیده است [۲۱].

در سال ۲۰۰۱ مقاله‌ای تحت عنوان تشخیص ایمنی آزمایشگاهی HIV با استفاده از شبکه عصبی توسط هات زاکیس<sup>۴</sup> ارائه شده است [۲۲].

درسال ۲۰۰۲، بازوئی<sup>۵</sup> و همکارانش با استفاده از مطالعات QSAR فعالیت ضدایدز را برای مشتقات HEPT بررسی نموده‌اند. در این کار هر دو مدل شبکه عصبی و رگرسیون خطی با

---

<sup>۱</sup> -Kukla  
<sup>۲</sup> -Tronchet  
<sup>۳</sup> -Cai  
<sup>۴</sup> -Hatzakis  
<sup>۵</sup> -Bazoui



هم مقایسه شده است. نتایج بدست آمده بیانگر این واقعیت است که فعالیت ترکیبات HEPT به

طور موثری وابسته به خصوصیات هیدروفوبیک و گروههای استخلاق می باشد [۲۳]

میشل<sup>۱</sup> در سال ۲۰۰۳ وانیر<sup>۲</sup>، با استفاده از شبکه عصبی فعالیت ضدایدز را برای یک سری از ترکیبات تتراپیرول پیشگویی کرد و شبکه عصبی به عنوان مناسب ترین مدل معرفی شده است [۲۴].

چایو<sup>۳</sup> در سال ۲۰۰۴ مقاله ای تحت عنوان توسعه مدل های QSPR مربوط به شبکه عصبی برای ثابت های استخلاق هانس<sup>۴</sup> و کاربردهای آن در مطالعات QSAR مربوط به HIV-1 و بازدارنده های دی هیدروفیلات، ارائه داد [۲۵]

در سال ۲۰۰۴ دالی<sup>۵</sup> و همکارانش به مطالعات QSAR غیر خطی مشتقات HEPT پرداختند و هشت توصیف کننده ساختمانی و متغیر فیزیکی شیمیایی را در نظر گرفتند و به روابط ویژه ای بین پارامترهای ساختاری و فعالیت ضدایدز پی بردند [۲۶].

در همین سال (۲۰۰۴) دالی، مجدداً با استفاده از شبکه عصبی سه لایه ای (۸-۴-۱) بر روی ۸۲ ترکیب از مشتقات TIBO مطالعه کرد و  $R^2=0/861$  را برای سری پیشگویی گزارش داد [۲۷].

یانگ<sup>۶</sup> و همکارانش در سال ۲۰۰۵ با استفاده از شبکه عصبی و توابع اساسی بیو<sup>۷</sup>، موفق به پیشگویی سایت های گسسته پروتئاز در HIV و سایت های گسسته تریپسین<sup>۸</sup> در پروتئین شدند [۲۸].

---

<sup>1</sup> Michele

<sup>2</sup> Vanyur

<sup>3</sup> Chiu

<sup>4</sup> Hansch substituent constant

<sup>5</sup> Dauoli

<sup>6</sup> Yang

<sup>7</sup> Bio

<sup>8</sup> Trypsin

## فصل دوم

### محاسبات و مدلسازی

در این پروژه، شبکه عصبی مصنوعی به عنوان یک شیوه مدلسازی در مطالعه رابطه ساختار - فعالیت برای پیش‌بینی داده‌های سیتوتوکسی سی تی<sup>۱</sup> ( $CC_{50}$ ) مشتقات ۱-[(۵-کلروفنیل) سولفونیل]-H-۱-پیرول<sup>۲</sup> بکار گرفته شده است. و توانایی مدل حاصله در پیش‌گویی داده‌های سیتوتوکسی سی تی مورد ارزیابی قرار گرفته است.

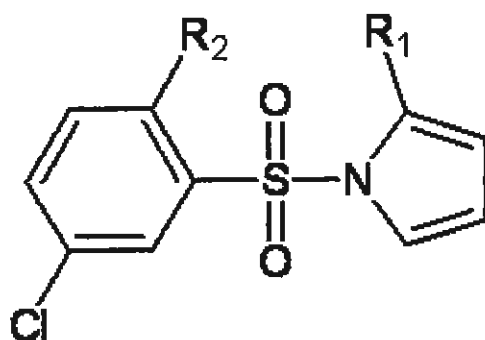
#### ۱-۲- سری داده‌ها

سری داده‌های مورد استفاده در این پروژه، داده‌های  $CC_{50}$  برای ۲۷ ترکیب از مشتقات ۱-[(۵-کلروفنیل) سولفونیل]-H-۱-پیرول می‌باشد که از مرجع شماره [۲۹] گردآوری شده است. مقادیر منهای لگاریتم داده‌های  $CC_{50}$  ( $-\log CC_{50}$ ) بر حسب مولار به عنوان متغیر پاسخ در این مدلسازی مورد استفاده قرار گرفت. ساختار پایه مشتقات در شکل (۱-۲) آمده است که دو استخلاف  $R_1$  و  $R_2$  در جایگاه‌های متفاوتی روی حلقه‌های فنیل و پیرول قرار گرفته اند و با تغییر این گروه‌ها مقدار  $CC_{50}$  ترکیب مورد نظر تغییر می‌کند. در جدول (۱-۲)، جزئیات استخلافها و مقادیر عددی  $CC_{50}$  متناظر هر ترکیب آورده شده است. همانطور که ملاحظه

<sup>۱</sup>- Cytotoxicity

<sup>۲</sup> - 1-[(5-chlorophenyl) sulfonyl]-1H-pyrrole derivatives

می‌شود، نوع استخلافها و نحوه آرایش آنها در اطراف اسکلت اصلی مولکول تأثیر زیادی بر روی فعالیت دارویی ترکیب دارد. لذا در این پروژه سعی شده است مدلی ارائه گردد که بتواند خاصیت دارویی هر ترکیب را با هر نوع استخلاف در هر موقعیتی پیش‌بینی نماید، بدون اینکه نیازی به سنتز و آزمایش سلولی ترکیب مورد نظر باشد.



شکل (۱-۲)- اسکلت مولکولی ترکیبات مورد بررسی

## ۲-۲- نرم افزارهای مورد استفاده

برای انجام مدل‌سازی لازم است، ساختار واقعی مولکول ترسیم شده و سپس پارامترهای عددی مربوط به ساختار هر مولکول استخراج گردد. لذا برای ترسیم و بهینه‌سازی ساختار سه بعدی مولکولها از نرم افزار Hyper Chem 7 [۳۰] استفاده گردید. برای محاسبه توصیفگرهای مولکولی از نرم‌افزار Dragon که به وسیله گروه QSPR و کمومتریکس میلانو<sup>۱</sup> [۳۱] ارائه شده بکار گرفته شده است. برای انتخاب توصیفگرهای مولکولی مناسب نرم افزار SPSS-13 [۳۲] مورد استفاده قرار گرفته و برنامه شبکه عصبی در نرم افزار MATLAB 5.5 [۳۳] نوشته شده است. تمام محاسبات با استفاده از یک کامپیوتر پنتیوم ۴- با سیستم عامل ویندوز xp انجام گردید.

جدول (۱-۲) - جزئیات ساختار مولکولی ترکیبات مورد بررسی

شماره	R <sub>1</sub>	R <sub>2</sub>	داده های CC <sub>50</sub>
۱	COO(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	NO <sub>2</sub>	۴/۶۶
۲	COOCH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	NO <sub>2</sub>	۴/۶۲
۳	COOCH(CH <sub>3</sub> )C <sub>2</sub> H <sub>5</sub>	NO <sub>2</sub>	۴/۴۹
۴	COOCH <sub>2</sub> -Cyclo-C <sub>3</sub> H <sub>5</sub>	NO <sub>2</sub>	۴/۰۲
۵	COOCH <sub>2</sub> CCH	NO <sub>2</sub>	۵/۳۱
۶	COO-Cyclo-C <sub>5</sub> H <sub>9</sub>	NO <sub>2</sub>	۴/۹۶
۷	COO-Cyclo-C <sub>6</sub> H <sub>11</sub>	NO <sub>2</sub>	۳/۸۰
۸	COOCH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	NO <sub>2</sub>	۴/۹۲
۹	CN	NO <sub>2</sub>	۵/۸۵
۱۰	COOC <sub>6</sub> H <sub>5</sub>	NO <sub>2</sub>	۴/۸۱
۱۱	H	NH <sub>2</sub>	۶/۱۶
۱۲	COOCH(CH <sub>3</sub> )C <sub>2</sub> H <sub>5</sub>	NH <sub>2</sub>	۳/۸۰
۱۳	COOCH <sub>2</sub> -Cyclo-C <sub>3</sub> H <sub>5</sub>	NH <sub>2</sub>	۴/۴۱
۱۴	COOCH <sub>2</sub> CCH	NH <sub>2</sub>	۳/۹۳
۱۵	COO-Cyclo-C <sub>5</sub> H <sub>9</sub>	NH <sub>2</sub>	۴/۲۷
۱۶	CN	NH <sub>2</sub>	۴/۸۹
۱۷	COOC <sub>6</sub> H <sub>5</sub>	NH <sub>2</sub>	۴/۶۲
۱۸	H	NH <sub>2</sub>	۳/۳۶
۱۹	COOCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	NH <sub>2</sub>	۳/۹۶
۲۰	COOCH <sub>2</sub> CHCH <sub>2</sub>	NH <sub>2</sub>	۴/۰۰
۲۱	CONH <sub>2</sub>	NO <sub>2</sub>	۴/۲۸
۲۲	CONHC <sub>2</sub> H <sub>5</sub>	NO <sub>2</sub>	۴/۷۸
۲۳	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	NO <sub>2</sub>	۴/۸۹
۲۴	CON(CH <sub>3</sub> )C <sub>6</sub> H <sub>5</sub>	NO <sub>2</sub>	۴/۷۷
۲۵	CONH <sub>2</sub>	NH <sub>2</sub>	۳/۷۰
۲۶	CON(C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	NH <sub>2</sub>	۳/۸۴
۲۷	CON(CH <sub>3</sub> )C <sub>6</sub> H <sub>5</sub>	NH <sub>2</sub>	۴/۰۶

## ۳-۲- انجام محاسبات و مدلسازی

### ۳-۲-۱- رسم و بهینه‌سازی ساختار مولکولها

همانطور که در فصل قبل اشاره گردید، دومین مرحله در مطالعات QSAR پس از جمع‌آوری داده‌ها، بهینه‌سازی ساختار هندسی مولکولهای مورد بررسی می‌باشد. بدین منظور ساختار دوبعدی مشتقات ۱-[(۵-کلروفنیل) سولفونیل]-H-۱-پیرول در نرم افزار Hyper chem7 ترسیم شدند و سپس ساختار نهایی به روش نیمه تجربی و با استفاده از روش AM1 بهینه گردید. تمامی محاسبات در سطح محدود شده هارتری-فاک انجام گردید. در بهینه سازی، الگوریتم پلاک-ربیر مورد استفاده قرار گرفت و بهینه سازی تا زمانی ادامه یافت تا اینکه گرادیان به  $0/001$  کیلوگالری بر مول<sup>ک</sup> برسد.

### ۳-۲-۲- محاسبه توصیفگرهای مولکولی

سومین مرحله در مطالعات ساختار - فعالیت، استخراج توصیفگرهای مولکولی باشد. در این پروژه توصیفگرهای محاسبه‌ای مورد استفاده قرار گرفتند لذا برای محاسبه این توصیفگرها، ساختار بهینه شده هر مولکول به نرم افزار Dragon منتقل گردید و تمام ۱۴۸۱ توصیف کننده برای هر مولکول محاسبه گردید. این پارامترها در چهار دسته کلی صفر بعدی، یک بعدی، دو بعدی و سه بعدی قرار می‌گیرند. این چهار دسته، مجموعاً به ۱۸ گروه کوچکتر تقسیم‌بندی می‌شوند.

توصیفگرهای صفر بعدی در برگیرنده توصیف کننده‌های اساسی هستند که مستقل از صورتبندی مولکولی می‌باشند. دسته توصیفگرهای یک بعدی شامل گروههای عاملی، خصوصیات مولکولی محاسبه شده از مدل‌های رگرسیون و پارامترهای تجربی می‌باشند. توصیف کننده‌های دو بعدی در برگیرنده توصیف کننده‌های ساختمانی، خود همبستگی دو بعدی،

شاخصهای بار و مقادیر مثبت و منفی ماتریس بارگیری می‌باشند. شاخصهای آروماتیکی، توصیفگرهای مربوط به بار، پروفایل مولکولی، توصیف کننده‌های هندسی و توابع اساسی شعاعی، توصیف کننده‌های شامل شاخص‌های آماری اجزاء اصلی و توصیف کننده‌های محاسبه شده از عناصر ماتریس‌های مربوط به نفوذ مولکولی در دسته توصیف کننده‌های سه بعدی قرار می‌گیرند.

### ۲-۳-۳- انتخاب توصیفگرهای مهم

انتخاب توصیفگرهای مهم که داده‌های  $CC_{50}$  را به ساختار مولکولی ارتباط می‌دهند یکی از مراحل مهم در مطالعات QSAR می‌باشد. از آنجایی که تعداد توصیف کننده‌های محاسبه شده با استفاده از نرم افزار Dragon بسیار زیاد می‌باشد (۱۴۸۱ توصیفگر) و تمام این توصیفگرها با داده‌های  $CC_{50}$  ارتباط ندارند لذا لازم است از میان این تعداد زیاد از توصیفگرها، توصیف کننده‌های مهم استخراج شوند. بدین منظور مراحل زیر در استخراج پارامترهای مؤثر به ترتیب بکار گرفته شد:

الف) تمام توصیف کننده‌هایی که مقادیر یکسانی برای تمام مولکولهای مورد بررسی داشتند از لیست توصیفگرها حذف شدند.

ب) با توجه به اینکه نیابستی متغیرهای ورودی در روش رگرسیون خطی چندگانه، با یکدیگر همبستگی نداشته باشند، لذا از میان جفت توصیف کننده‌ای که ضریب همبستگی بزرگتر از  $0.90$  ( $r^2 > 0.90$ ) داشتند، یکی حذف گردید. این کار با استفاده از نرم افزار SPSS برای هر گروه از توصیفگرها انجام گردید.

ج) برای انتخاب توصیفگرهای مهم که با داده‌های  $CC_{50}$  ارتباط دارند روش رگرسیون مرحله به مرحله در نرم افزار SPSS برای تمام مولکولها و تمام توصیفگرهای باقیمانده پس از مراحل الف و ب انجام گردید. برای اینکار، توصیفگرها به عنوان متغیرهای مستقل و داده‌های  $CC_{50}$  به

عنوان متغیر وابسته در نظر گرفته شد. نتایج حاصله نشان داد که از میان تعداد زیادی توصیف کننده، فقط تعداد ۸ توصیف کننده دارای ارتباط معناداری با داده‌های CC<sub>50</sub> مولکولهای مورد بررسی می‌باشند. همانطور که بعداً خواهیم گفت این توصیف کننده ها در مدل سازی مورد استفاده قرار گرفت. این توصیف گرها به همراه کلاس مربوطه و نام کامل آنها در جدول (۲-۲) آمده است. جدول (۳-۲) بیانگر ماتریس همبستگی بین توصیف گرها می باشد که نشان دهنده فقدان همبستگی در بین توصیف گرها است.

جدول (۲-۲) - توصیفگرهای بکار رفته در مدل سازی

شماره	Symbol	Class	Meaning
۱	Q2	Charge	Total squared charge
۲	Mor30p	3D-MoRSE	3D-MoRSE-signal 30
۳	Mor32m	3D-MoRSE	3D-MoRSE-signal 32
۴	MWC10	Walk and path count	Molecular walk count of order 10
۵	R4m	GETAWAY	R autocorrelation of lag 4
۶	R7m	GETAWAY	R autocorrelation of lag 7
۷	H8e	GETAWAY	H autocorrelation of lag 8
۸	E1e	WHIM	1 <sup>st</sup> component symmetry directional WHIM index

جدول (۳-۲) - ماتریس همبستگی توصیفگرهای بکار رفته در مدل سازی

	Q2	Mor30p	E1e	MWC10	Mor32m	R7m	R4m	H8e
Q2	۱							
Mor30p	۰/۵۲۶	۱						
E1e	۰/۳۰۸	۰/۰۹۸	۱					
MWC10	۰/۰۷۸	۰/۱۶۵	۰/۱۳۸	۱				
Mor32m	۰/۵۰۶	۰/۳۵۴	۰/۰۹۴	۰/۱۸۵	۱			
R7m	۰/۷۱۵	۰/۴۷۵	۰/۰۸۸	۰/۱۲۹	۰/۵۲۹	۱		
R4m	۰/۱۶۲۵	۰/۴۴۰	۰/۱۶۴	۰/۰۱۷	۰/۲۵۴	۰/۵۸۰	۱	
H8e	۰/۷۴۷	۰/۷۲۴	۰/۰۷۴	۰/۰۸۵	۰/۵۵۶	۰/۴۸۶	۰/۴۰۸	۱

## ۲-۳-۴- مدل شبکه عصبی و بهینه سازی آن

پس از استخراج توصیفگرهای مناسب، مرحله اساسی دیگر ساختن مدلی است که به خوبی بتواند یک ارتباط کمی بین ساختار و فعالیت دارویی ترکیبات شیمیایی مورد نظر فراهم آورد. در این مطالعه شبکه عصبی مصنوعی که یک روش مدلسازی غیر خطی است انتخاب گردید. شبکه انتخاب شده یک شبکه عصبی با جریان رو به جلو<sup>۱</sup> با الگوی آموزش پس انتشار<sup>۲</sup> می‌باشد. به منظور آموزش موفق شبکه و در نتیجه افزایش قدرت پیش‌بینی شبکه بایستی عوامل مؤثر در آموزش شبکه بررسی و بهینه شوند. پارامترهای مؤثر و بهینه شده عبارتند از:

۱- تعداد لایه‌های پنهان

۲- نوع الگوریتم آموزشی

۳- نوع تابع تبدیل در لایه خروجی و لایه پنهان

۴- تعداد ورودیها

۵- سرعت آموزش

۶- تعداد دوره‌های آموزش

۷- تعداد نرونها در لایه پنهان

به منظور بهینه سازی پارامترهای فوق، ۲۷ سری داده مشتقات ۱- (۵-کلروفنیل) سولفونیل]-

H-۱-پیرول، به صورت تصادفی به دو زیر گروه تقسیم شدند:

الف ( تعداد ۲۰ ترکیب به عنوان سری آموزش انتخاب شدند که این داده‌ها برای آموزش شبکه

مورد استفاده قرار گرفتند.

---

<sup>۱</sup>- Feed-forward ANNs

<sup>۲</sup>- Back propagation algorithm



ب) تعداد ۷ ترکیب به عنوان سری پیش‌بینی انتخاب شدند و براری ارزیابی قدرت پیش‌بینی شبکه پس از آموزش مورد استفاده قرار گرفتند. معیار در نظر گرفته شده در بهینه‌سازی به حداقل رساندن خطای مجذور میانگین<sup>۱</sup> (MSE) حاصله در سری پیش‌بینی می‌باشد.

### ۲-۳-۴-۱- انتخاب تعداد لایه‌های شبکه عصبی

هر شبکه عصبی دارای حداقل یک لایه ورودی، یک لایه خروجی و تعدادی لایه‌های پنهان می‌باشد. متأسفانه هیچ قاعده تئوری یا تجربی که بتوان به کمک آنها تعداد لایه‌های پنهان در یک شبکه عصبی را تعیین نمود وجود ندارد. اما نتایج حاصله از مطالعات مختلف نشان می‌دهد که دربرکارگیری شبکه عصبی درشیمی، انتخاب یک لایه پنهان نتایج رضایتبخشی را ایجاد می‌کند [۳۴]. بنابراین در این مطالعه یک شبکه عصبی سه لایه‌ای متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی بکار گرفته شد.

### ۲-۳-۴-۲- انتخاب الگوریتم آموزشی

پس از طراحی اولیه، شبکه عصبی بایستی آموزش داده شود تا اینکه قادر به پیش‌بینی صحیحی باشد. یکی از الگوهای آموزشی که استفاده زیادی در آموزش شبکه‌های عصبی دارد، روش پس انتشار می‌باشد [۳۵]. این روش اساساً از الگوریتم کاهش گرادیان استفاده می‌کند. بکارگیری تکنیک ساده کاهش گرادیان، دارای مشکلاتی در همگرایی شبکه می‌باشد زیرا احتمال به دام افتادن در می‌نیم‌های محلی وجود دارد در نتیجه یادگیری شبکه سخت و کند است. برای حل مشکل همگرایی الگوریتم آموزش کاهش گرادیانی، روشهای زیر ارائه شده است: الف) افزودن عبارت ممنوم که تا حدودی مشکل عدم همگرایی را حل می‌کند.

---

<sup>۱</sup> - Root mean squares errors

ب) بکارگیری روشهای مرتبه دوم برای یادگیری شبکه می‌باشد. از میان روشهای مرتبه دوم الگوریتم لونیبرگ - مارکوت<sup>۱</sup> (LM) یکی از رایجترین شیوه‌ها می‌باشد که هنگامی که پارامترها کارایی شبکه RMSE می‌باشد بسیار مناسب است [۳۶-۳۹].

الگوریتم LM شکل تغییر یافته روش نیوتن می‌باشد که علاوه بر مزیت سرعت روش نیوتن، دارای مزیت همگرایی سریع است. در این مطالعه، الگوریتم‌های مختلف آموزش پس‌انتشار مورد بررسی قرار گرفتند و نتایج حاصله نشان داد که از میان تمام الگوریتم‌های مورد بررسی قرار گرفته، الگوریتم LM دارای کمترین خطای پیش‌بینی و بعلاوه سرعت آموزش بسیار بالا می‌باشد. بنابراین الگوریتم یادگیری پس‌انتشار لونیبرگ - مارکوت (تابع Trainlm در جعبه ابزار نرم افزار مطلب<sup>۲</sup>) در مطالعات بعدی مورد استفاده قرار گرفت.

### ۲-۳-۴-۳- انتخاب نوع تابع تبدیل

در ساختار شبکه برای هر لایه خروجی و هر لایه پنهان یک تابع تبدیل لازم است. انتخاب تابع تبدیل بستگی به خروجی مورد نظر دارد. اگر خروجی‌ها کیفی باشند، معمولاً تابع انتقال زیگموئیدی استفاده می‌شود و اگر خروجی‌ها کمی باشند، تابع انتقال خطی مورد استفاده قرار می‌گیرد. در این مطالعه با توجه به اینکه خروجی‌های شبکه عصبی داده‌های کمی CC<sub>50</sub> می‌باشند لذا تابع انتقال خطی برای لایه خروجی انتخاب گردید.

مطالعات لازم برای انتخاب تابع انتقال مناسب برای لایه پنهان انجام گرفت. به این منظور توابع خطی، زیگموئیدی و تانژانت هیپربولیک به عنوان توابع انتقال لایه پنهان بکار گرفته شدند. با هر تابع، شبکه هائی با تعداد نرونهای لایه پنهان ۱ تا ۶ و تعداد دوره‌های آموزشی یکسان و برابر با ۱۰ آموزش داده شدند و سپس شبکه‌های آموزش داده شده برای پیش‌بینی داده‌های CC<sub>50</sub> در سری پیش‌بینی مورد استفاده قرار گرفتند. نتایج حاصله (جدول ۲-۴) نشان می‌دهند که با

<sup>۱</sup> - Levenberg-Marquardt back propagation algorithm

<sup>۲</sup> - Matlab toolbox

بکارگیری تابع تبدیل زیگوئیدی در لایه پنهان کمترین خطای پیش‌بینی مشاهده می‌شود. بنابراین تابع تبدیل زیگوئیدی (LOGSIG) در جعبه ابزار نرم افزار مطلب) برای لایه پنهان در نظر گرفته شد.

جدول (۲-۴) - مقادیر MSE برای توابع تبدیل متفاوت با تعداد گره های مختلف در لایه پنهان

تعداد گره ها	نوع تابع تبدیل		
	LOGSIG	TANSIG	PURELIN
۱	۰/۵۶۸۹	۰/۴۲۹۰	۰/۰۷۸۷
۲	۰/۰۱۰۸	۰/۱۵۰۲	۰/۰۴۶۲
۳	۰/۰۳۲۷	۰/۰۴۱۱	۰/۱۷۵۹
۴	۰/۴۰۸۲	۰/۰۳۳۴	۰/۲۴۲۲
۵	۰/۳۱۸۸	۰/۳۸۲۸	۰/۲۰۴۷
۶	۰/۳۱۷۸	۰/۳۷۹۷	۰/۰۱۴۱

### ۲-۳-۴-۴- انتخاب سرعت آموزش و تعداد گره‌های لایه پنهان

مطالعات اولیه نشان داد که تغییر در سرعت آموزش در ناحیه (۰/۱ - ۰/۰۰۱)، با استفاده از تعداد گره‌های متفاوت در لایه پنهان تأثیر چندانی روی MSE سری داده‌های پیش‌بینی ندارد. بنابراین سرعت آموزش ۰/۰۱۰ برای آموزش شبکه عصبی مورد استفاده قرار گرفت.

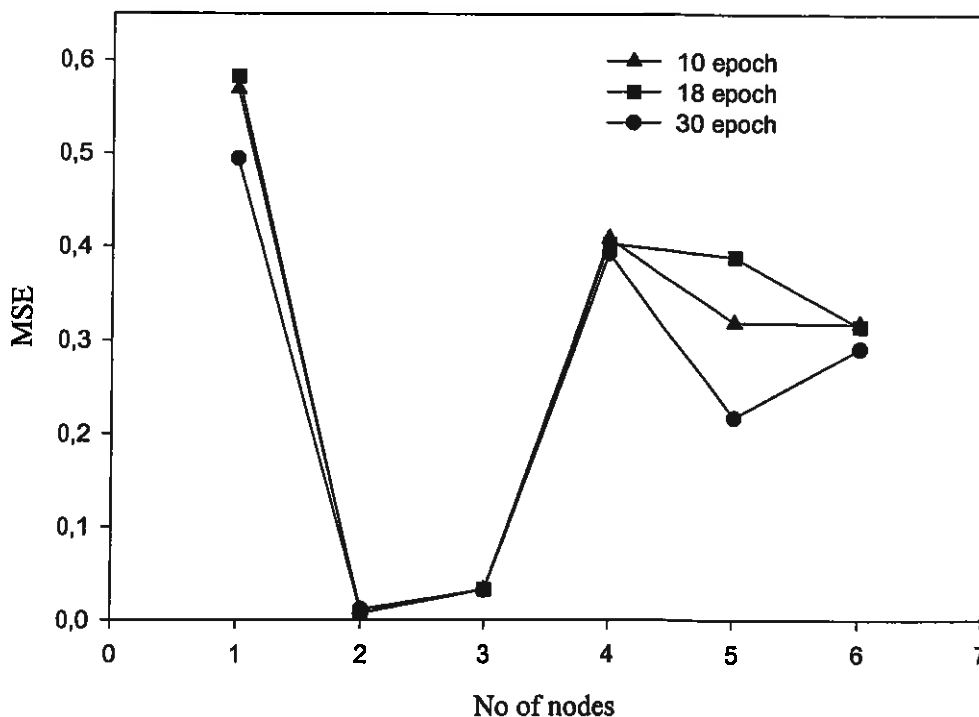
برای انتخاب بهترین تعداد گره‌های لایه پنهان این پارامتر مورد بررسی و بهینه سازی قرار گرفت. به این منظور شبکه هایی با تعداد گره‌های لایه پنهان در محدوده ۱۰-۲ با فواصل یک و با تعداد دوره‌های آموزش در ناحیه ۱۰-۱۰۰ با فواصل ۱۰ طراحی و آموزش داده شدند. شبکه‌های آموزش داده شده در پیش‌بینی داده‌های CC<sub>50</sub> سری پیش‌بینی بکارگرفته شدند. نتایج حاصله نشان در جدول (۲-۵) و شکل (۲-۲) آمده است. نتایج نشان میدهد که مقدار

MSE برای تعداد دو گره در لایه پنهان در تمام دوره‌های آموزشی می‌نیمم مقدار است. بنابراین

تعداد گره‌های برابر با ۲ که حداقل MSE را ایجاد می‌کند انتخاب گردید.

جدول (۲-۵) - مقادیر MSE برای تعداد گره‌های مختلف در لایه پنهان

تعداد دوره‌های آموزشی							
تعداد گره	۶	۱۰	۱۴	۱۸	۲۲	۲۶	۳۰
۱	۰/۶۱۴۰	۰/۵۶۸۹	۰/۵۸۲۴	۰/۵۸۲۴	۰/۴۴۰۶	۰/۵۰۲۸	۰/۴۹۴۰
۲	۰/۱۲۷۹	۰/۰۱۰۸	۰/۰۱۰۲	۰/۰۰۶۷	۰/۰۰۷۴	۰/۰۰۷۴	۰/۰۱۲۰
۳	۰/۰۲۱۰	۰/۰۳۲۷	۰/۰۳۲۷	۰/۰۳۲۷	۰/۰۳۲۷	۰/۰۳۲۷	۰/۰۳۲۷
۴	۰/۴۱۹۴	۰/۴۰۸۲	۰/۴۰۳۰	۰/۴۰۲۶	۰/۴۰۲۴	۰/۴۰۲۴	۰/۳۹۲۰
۵	۰/۱۸۶۸	۰/۳۱۸۸	۰/۶۰۵۷	۰/۳۸۸۲	۰/۲۱۷۴	۰/۲۱۷۴	۰/۲۱۷۴
۶	۰/۲۹۳۲	۰/۳۱۷۸	۰/۳۲۷۰	۰/۳۱۸۰	۰/۳۱۰۳	۰/۲۹۷۴	۰/۲۹۱۵

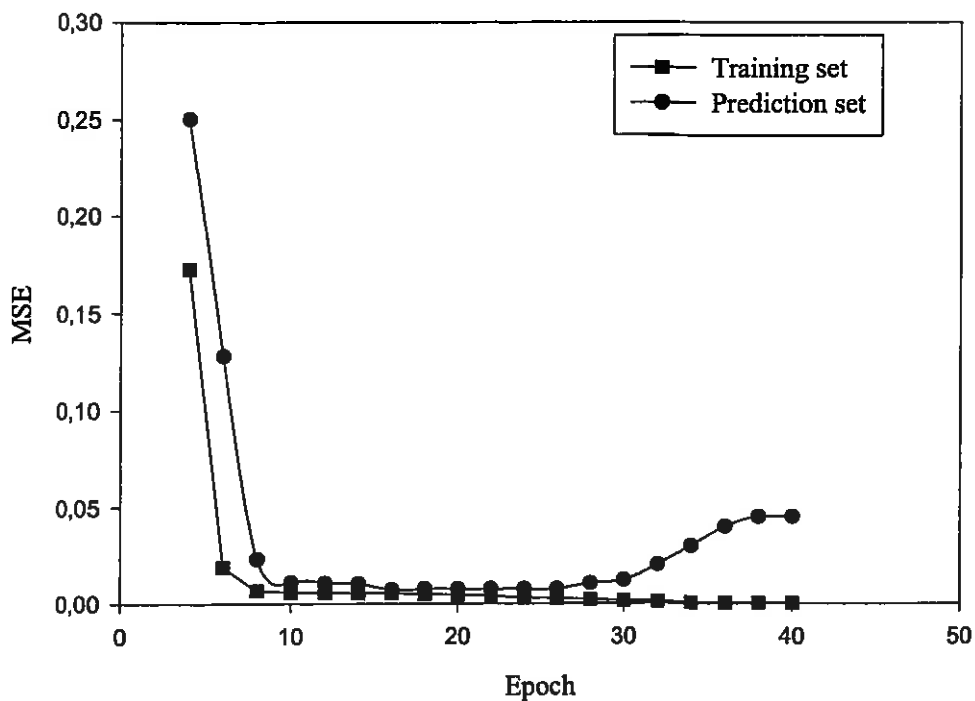


شکل (۲-۲) - تغییرات مقادیر MSE بر حسب تعداد گره های مختلف در لایه پنهان

### ۲-۳-۴-۵- انتخاب تعداد دوره‌های آموزش

در آموزش شبکه های عصبی با افزایش میزان آموزش (تعداد دوره‌های آموزشی) صحت پیش بینی برای داده‌های موجود در سری آموزش افزایش می یابد در حالی که ممکن است صحت پیش بینی برای داده‌های جدید خارج از سری آموزش کاهش یابد. برای جلوگیری از بروز چنین پدیده، شبکه هایی با تعداد دوره‌های آموزش در ناحیه ۴۰-۵ و با تعداد گره‌های برابر با ۲ در لایه پنهان طراحی و آموزش داده شدند و MSE برای داده‌های سری آموزش و سری پیش بینی در تمام دوره‌های آموزشی ثبت شدند. نتایج حاصل در شکل (۲-۳) نشان میدهد که با تعداد دوره‌های آموزشی MSE برای داده‌های موجود در سری آموزش کاهش می یابد در حالیکه

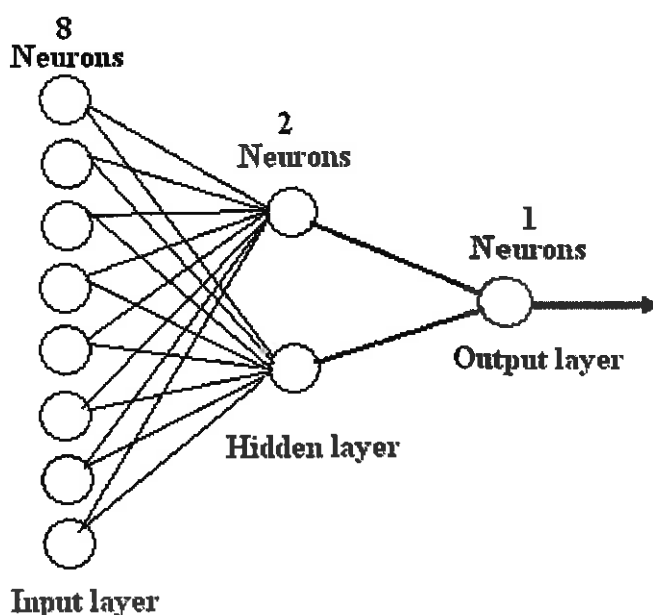
MSE برای سری پیش بینی در دوره‌های آموزشی بیشتر از ۱۸ افزایش می‌یابد. تعداد دور آموزش ۱۸ که حداقل MSE را ایجاد می‌کند انتخاب گردید.



شکل (۳-۲) - تغییرات مقادیر MSE بر حسب تعداد دوره‌های آموزش

### ۲-۳-۴-۵ - ساختار و پارامترهای بهینه شده شبکه عصبی

با توجه به نتایج حاصل در بهینه سازی شبکه، یک شبکه عصبی سه لایه با حرکت رو به جلو و آموزش پس انتشار با الگوریتم لونیبرگ-مارکوت انتخاب گردید. ساختار هندسی شبکه بهینه شده و پارامترهای بهینه آن شکل (۴-۲) آمده است. در شرایط بهینه حاصل حداکثر زمان لازم برای آموزش و پیش‌بینی در حدود ۱۵ ثانیه است که نشان دهنده سرعت بسیار بالای آموزش و پیش‌بینی مدل ارائه شده است.



شکل (۲-۴) - ساختار هندسی شبکه عصبی استفاده شده

### ۲-۳-۵- ارزیابی مدل ارائه شده

یکی از مهمترین و اساسی‌ترین مرحله در مطالعات QSAR ارزیابی قدرت پیش‌بینی شبکه می‌باشد. قدرت پیش‌بینی یک شبکه عصبی به صورت توانایی آن در ارائه خروجی‌های رضایتبخش برای مولکولهایی که در سری آموزش وجود نداشته‌اند، تعریف می‌شود. در ارزیابی مدل ارائه شده در این پروژه برای پیش‌بینی داده‌های  $CC_{50}$  داروهای ضد ایدز، دو روش بکار گرفته شد.

### ۲-۳-۵-۱- ارزیابی شبکه عصبی با استفاده از سری پیش‌بینی

در این شیوه، شبکه عصبی بهینه شده با استفاده از داده‌های ۲۰ سری از مولکولها آموزش داده شد و سپس شبکه آموزش داده شده برای پیش‌بینی داده‌های  $CC_{50}$ ، ۷ سری از مولکولها که در سری آموزش وجود نداشتند بکار گرفته شد. نتایج حاصله در جدول (۲-۶) نشان داده شده است. ترسیمی از مقادیر پیش‌بینی شده بر حسب مقادیر واقعی نیز در شکل

(۵-۲) نشان داده شده است. نتایج حاصله بیانگر این است که مدل ارائه شده در این پروژه خطاهای پیش‌بینی بسیار کم و صحت بالایی دارد.

### ۲-۳-۵-۲- ارزیابی شبکه عصبی به روش رد مرحله ای تک تک<sup>۱</sup>

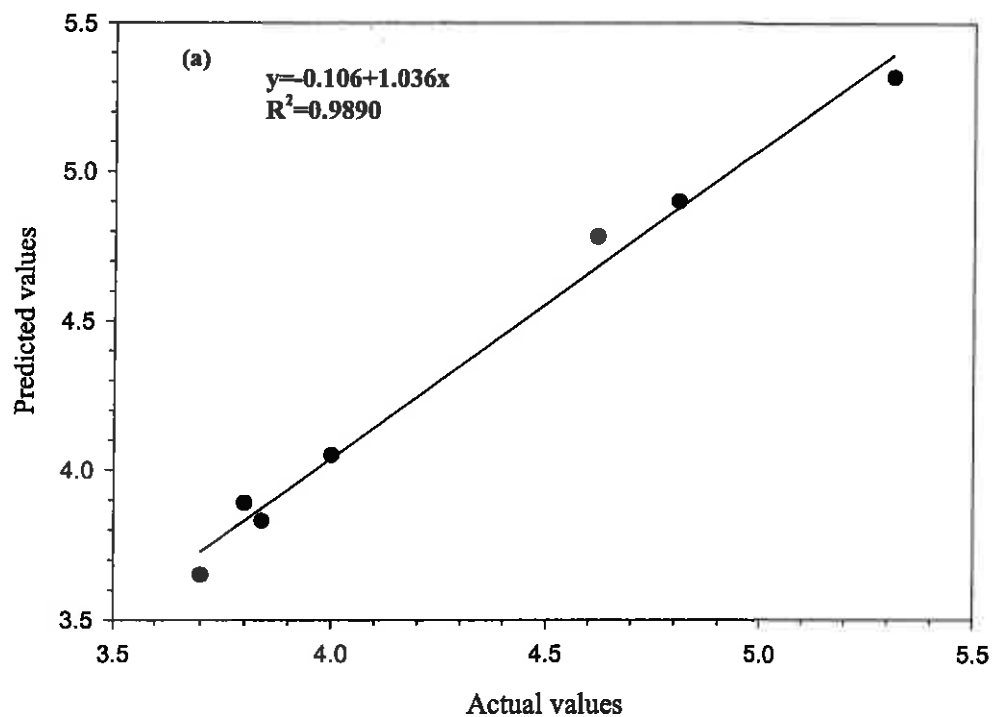
هر چند روش استفاده از سری پیش‌بینی برای ارزیابی شبکه عصبی یک روش مناسب و قابل قبول می‌باشد ولی با توجه به اینکه سری پیش‌بینی مورد استفاده در ارزیابی شبکه با سری پیش‌بینی بکار گرفته شده در مرحله بهینه سازی شبکه یکسان بود، بنابراین این احتمال وجود داشت که شبکه عصبی بهینه شده، به گونه‌ای بهینه شده باشد که فقط قابلیت بکارگیری برای سری داده‌های پیش‌بینی داشته باشد و با عوض شدن سری پیش‌بینی و وارد شدن مولکولهای جدید، شبکه قدرت پیش‌بینی خود را از دست بدهد، به عبارت دیگر شبکه قدرت تعمیم نداشته باشد. برای رفع این ابهام، روش رد مرحله‌ای تک تک برای ارزیابی شبکه عصبی ارائه شده، مورد استفاده قرار گرفت. در این روش به اینصورت عمل گردید که ابتدا یک مولکول از داده‌ها کنار گذاشته شد و شبکه عصبی در شرایط بهینه با ۲۶ ترکیب باقیمانده آموزش داده شد، سپس شبکه آموزش داده شده برای پیش‌بینی داده CC<sub>50</sub> ترکیب کنار گذاشته شده بکار گرفته شد. این عمل به ترتیب برای تمام ۲۷ مولکول مورد استفاده انجام گرفت. نتایج حاصله در جدول (۷-۲) خلاصه شده‌اند. ترسیمی از مقادیر پیش‌بینی شده بر حسب مقادیر واقعی نیز در شکل (۶-۲) نشان داده شده است. نتایج بیانگر قدرت پیش‌بینی بالای مدل و قدرت تعمیم شبکه عصبی ارائه شده می‌باشد. لذا با این مدل می‌توان داده‌های CC<sub>50</sub> را برای هر ترکیب جدیدی با خطای بسیار کم و قابل قبول، پیش‌بینی کرد.

<sup>۱</sup> - Leave-One-Out (LOO) method



جدول (۲-۶) - نتایج حاصله از ارزیابی مدل با استفاده از سری پیش بینی

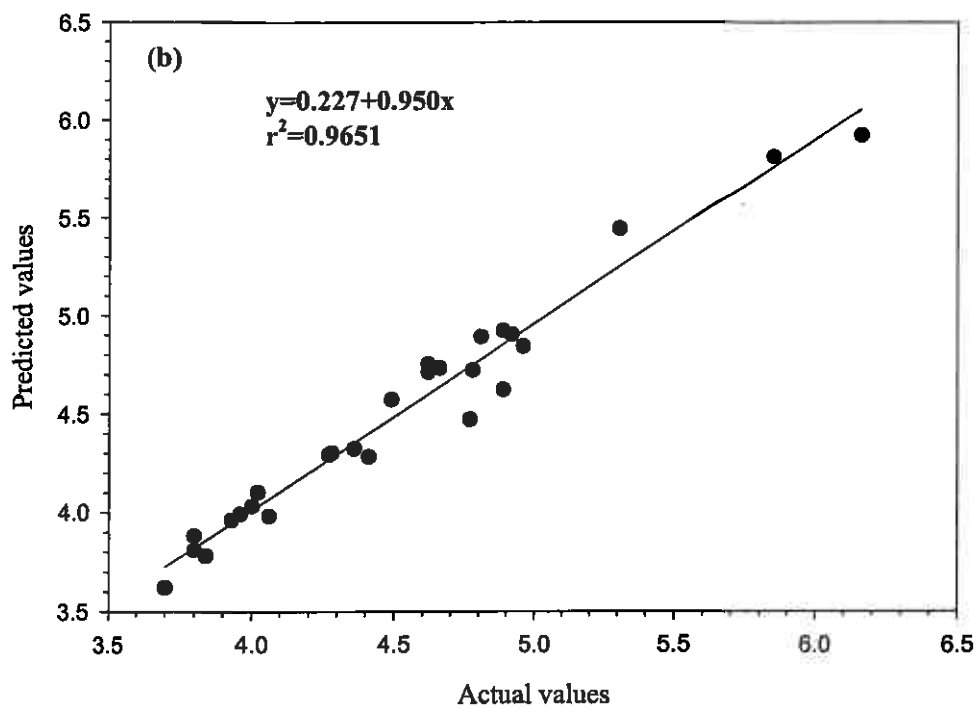
شماره	مقدار واقعی	مقدار پیش بینی شده	درصد خطا
۵	۵/۳۱	۵/۳۲	-۰/۱۹
۱۰	۴/۸۱	۴/۹۰	-۱/۸۷
۱۲	۳/۸۰	۳/۸۹	-۲/۳۷
۱۷	۴/۶۲	۴/۷۸	-۳/۴۶
۲۰	۴/۰۰	۴/۰۵	-۱/۲۵
۲۵	۳/۷۰	۳/۶۵	۱/۳۵
۲۶	۳/۸۴	۳/۸۳	۰/۲۶



شکل (۲-۵) - ترسیم از مقادیر پیش بینی شده بر حسب مقادیر واقعی در سری پیش بینی

جدول (۷-۲) - نتایج حاصله از ارزیابی مدل با استفاده از روش رد تک تک

شماره	مقادیر واقعی	مقادیر پیش بینی شده	درصد خطا
۱	۴/۶۶	۴/۷۳	-۱/۵۰
۲	۴/۶۲	۴/۷۱	-۱/۹۵
۳	۴/۴۹	۴/۵۷	-۱/۷۸
۴	۴/۰۲	۴/۱۰	-۱/۹۹
۵	۵/۳۱	۵/۴۴	-۲/۴۵
۶	۴/۹۶	۴/۸۴	۲/۴۱
۷	۳/۸۰	۳/۸۱	-۰/۲۶
۸	۴/۹۲	۴/۹۰	۰/۴۱
۹	۵/۸۵	۵/۸۱	۰/۶۸
۱۰	۴/۸۱	۴/۸۹	-۱/۶۶
۱۱	۶/۱۶	۵/۹۲	۳/۹۰
۱۲	۳/۸۰	۳/۸۸	-۲/۱۰
۱۳	۴/۴۱	۴/۲۸	۲/۹۵
۱۴	۳/۹۳	۳/۹۶	-۰/۷۶
۱۵	۴/۲۷	۴/۲۹	-۰/۴۷
۱۶	۴/۸۹	۴/۶۲	۵/۵۲
۱۷	۴/۶۲	۴/۷۵	-۲/۸۱
۱۸	۳/۳۶	۴/۳۲	۰/۹۲
۱۹	۳/۹۶	۳/۹۹	-۰/۷۶
۲۰	۴/۰۰	۴/۰۳	-۰/۷۵
۲۱	۴/۲۸	۴/۳۰	-۰/۴۷
۲۲	۴/۷۸	۴/۷۲	۱/۲۵
۲۳	۴/۸۹	۴/۹۲	-۰/۶۱
۲۴	۴/۷۷	۴/۴۷	۶/۲۹
۲۵	۳/۷۰	۳/۶۲	۲/۱۶
۲۶	۳/۸۴	۳/۷۸	۱/۵۶
۲۷	۴/۰۶	۳/۹۸	۱/۹۷



شکل (۲-۵) - ترسیمی از مقادیر پیش بینی شده بر حسب مقادیر واقعی برای کل داده ها

## فصل سوم

### بحث و نتیجه‌گیری

با توجه به اینکه تا به حال هیچ گزارشی دربارهٔ بکارگیری شبکه‌های عصبی مصنوعی در مطالعات QSAR مشتقات ۱-[(۵-کلروفنیل) سولفونیل]-H-۱-پیرول داده نشده است. بنابراین در این پروژه مدلی بر مبنای شبکه عصبی مصنوعی برای پیش‌بینی فعالیت دارویی ضد ایدز این ترکیبات ارائه گردیده است.

#### ۳-۱- توصیف‌گرهای انتخاب شده

انتخاب توصیف‌گرهای مناسب مطابق روش ارائه شده در بخش (۲-۳-۳) انجام گرفت. و در نهایت ۸ توصیف‌گر از میان کل توصیف‌گرها در مدلسازی شبکه عصبی مورد استفاده قرار گرفت که لیست کامل آنها در جدول (۲-۲) آمده است. در اینجا به بررسی علت اهمیت این توصیف‌گرها و رابطه احتمالی آنها با فعالیت ضد ایدز ترکیبات بررسی شده می‌پردازیم.

توصیف‌گرهای 3D-MoRSE (نمایش سه بعدی ساختار مولکول بر اساس پراش الکترون) بر اساس اطلاعات به دست آمده از مختصات سه بعدی اتم‌ها در اثر تبدیلات مورد استفاده در مطالعات پراکنش الکترون استوار هستند [۴۱]. برخی از این توصیف‌گرها (Mor30p و Mor32m) در مدلسازی مورد استفاده قرار گرفته‌اند زیرا بیانی از آرایش سه‌بعدی

اتم‌ها بدون توجه به اندازه مولکول بوده و قابل کاربرد برای بسیاری از مولکولها با تفاوت‌های ساختاری زیاد می‌باشند.

توصیف‌گرهای WHIM حاوی اطلاعات سه بعدی نظیر نحوه توزیع اتمی با توجه یک مرجع در مولکول می‌باشند [۴۲]. یکی از این توصیف‌گرها (E1e) در مدل ظاهر شده است و بیانگر تأثیر کنفورماسیون ترکیب در حین بر هم کنش با ویروس می‌باشد.

توصیف‌گرهای GETAWAY توصیف‌گرهایی که با توجه به مختصات فضایی اتم‌ها در یک مولکول براحتی قابل محاسبه هستند [۶۲و۶۱]. توصیف‌گرهای R4m و R7m که در مدل ظاهر شده اند، به ساختار سه بعدی مولکولها حساس بوده و در بر گیرنده اطلاعاتی نظیر شکل، سائز و تقارن مولکول و نحوه توزیع اتمها در مولکول می‌باشند. توصیف‌گر H8e به اندازه و محل اتمها در مولکول وابسته بوده به گونه ای که با افزایش اندازه اتم و فاصله اتم از مرکز مولکول، مقدار آن افزایش می‌یابد.

با توجه به تئوری شیمیائی کلاسیک، بر هم کنشهای شیمیائی از نوع کووالانسی یا الکترواستاتیک هستند. در بر هم کنشهای الکترواستاتیک، بارهای الکتریکی در مولکولها اهمیت زیادی دارند. توصیف‌گر Q2 یکی از توصیف‌گرهای بر پایه بار [۶۴] می‌باشد و اهمیت آن در مدل به این علت است که بیانی از بر هم کنشهای الکترواستاتیک می‌باشد.

### ۳-۲- ارزیابی آماری مدل ارائه شده

همانگونه که در بخش (۲-۳-۵) اشاره شده برای ارزیابی کارایی شبکه عصبی در شرایط بهینه، روشهای استفاده از سری پیش‌بینی و روش طرد یکی یکی به کار گرفته شد که نتایج حاصله در جداول (۲-۶) و (۲-۷) آمده است. برای بررسی میزان کارایی مدل ارائه شده و مقایسه آن با روش رگرسیون خطی چندگانه چهار پارامتر آماری زیر محاسبه شدند.

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

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

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \times 100$$

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

نتایج حاصل در جدول (۱-۳) آمده است. نتایج حاصله نشان دهنده قدرت پیش‌بینی بالای مدل شبکه عصبی می‌باشد. این نتایج آشکار بیانگر این هستند که روش QSAR ارائه شده در این پروژه قادر است فعالیت سیتوتوکسی سیتی ضد ایدز هر مشتقی از ۱-]۵- (کلروفنیل) سولفونیل-۱-H-پیرول با هر استخلافی را با قدرت بالا و خطای بسیار کم پیش‌بینی کند.

جدول (۱-۳) - پارامترهای آماری محاسبه شده

پارامتر	مدل پیشنهادی	
	سری پیش‌بینی	کل داده‌ها
MAE	۱/۵۴	۱/۸۷
REP	۱/۹۱	۲/۵۲
MSE	۰/۰۰۶۷	۰/۰۱۳
R <sup>2</sup>	۰/۹۸۹۰	۰/۹۶۵۱

### ۳-۳- نتیجه‌گیری نهایی

در این پروژه یک مدل شبکه عصبی با آموزش به کمک الگوریتم لونیگ مارکوت برای مطالعات QSAR مشتقات ۱-]۵- (کلروفنیل) سولفونیل-۱-H-پیرول ارائه شده است. بر اساس اطلاعات

موجوده، این اولین گزارش در مورد بکارگیری شبکه عصبی با الگوریتم آموزش لونبرگ مارکوت برای پیش‌بینی فعالیت دارویی ترکیبات شیمیایی است. نتایج بدست آمده نشان می‌دهد که شبکه عصبی قادر است یک رابطه صحیح و رضایت‌بخش بین توصیف‌گرهای مولکولی محاسبه‌ای و فعالیت ضد ایدز ترکیبات ۵۱-[(۵-کلروفنیل) سولفونیل]-۱-H-پیرول برقرار نماید. این مدل می‌تواند کمک بسیار زیادی در طراحی داروهای ضد ایدز از این مشتقات با استخلافهای مختلف نماید.

- [1]- 11. Yasri, A. and Hartsough, D., J. Chem. Inf. Comput. Sci. 41 (2001) 1218.
- [2]- D.C. Young, "Computational Chemistry: A", John Wiley & Sons Inc., New York, 2001.
- [3]- J.B. Foresman, "Exploring Chemistry with Electronic Structure methods", Gaussian Inc., Pittsburgh, 1996.
- [4]- Kukla, M., Breslin, H., Pauwels, R., Fedde, C., Miranda, M., Scott, M., Sherrill, R., Raeymaekers, A., Van Gelder, J., Andries, K., Janssen, M. A. C., De Clercq, E. and Janssen, P., J. Med. Chem. 34 (1991) 746.
- [5]- Miri, R., Javidnia, K., Hemmateenejad, B., Azarpira, A. and Amirghofran, Z., Bioorg. Med. Chem. 12 (2004) 2529.
- [6]- Bajaj, S., Sambhi, S.S. and Madan, A.K., Bioorg. Med. Chem. Lett. 15 (2005) 467.
- [7]- Pillai, A.D., Rani, S., Rathod, P. D., Xavier, F. P., Vasu, K. K., Padh, H. and Sudarsanam, V., Bioorg. Med. Chem. 13 (2005) 1275.
- [8]- Leonard, J. T. and Roy, K., Bioorg. Med. Chem. 14 (2006) 1039.
- [9]- Lin, L., Lin, W. Q., Jiang, J. H., Shen, G. L. and Yu, R. Q., Eur. J. Pharmaceut. Sci. 25 (2005) 245.
- [10]- De Lima Ribeiro, F.A. and Castro Ferreira, M.M., J. Mol. Struct.: THEOCHEM 719 (2005) 191.
- [11]- De Paula da Silva, C.H.T., Sanches, S.M. and Taft, C.A., J. Mol. Graph. Model. 23 (2004) 89.



[۱۲]- موسوی جاهد، "ویروس شناسی پزشکی"، انتشارات دانشگاه علوم پزشکی و پیراپزشکی شهید بهشتی، ۱۳۷۸.

[۱۳]- مرعشی، م.، "ویروس شناسی انسانی"، انتشارات دانشگاه فردوسی مشهد، ۱۳۸۲.  
[14]- R. Garg, S.P. Gupta, H. Gao, M.S. Babu, A.K. Debnath, C. Hansch, Chem. Rev., 99 (1999) 3525.

[15]- <http://www.tthhiveclinic.com/life cycle.htm>

[16]- Z. Debyser, R. Pauwels, K. Anderies, J. Desmyter, M.J. Kukla, P.A.J. Janseen, E. De Clercq, Proc. Natl. Acad. Sci. U.S.A., 88 (1991) 1451.

[17]- D.L. Romero, M. Busso, C.K. Tan, F. Reusser, J.R. Palmer, S.M. Pop, P.A. Aristoff, K.M. Doney, Proc. Natl. Acad. Sci. U.S.A., 88 (1991) 6863.

[18]- R. Pauwels, K. Andries, Z. Debyser, P. Van Daele, D. Schols, P. Staffels, K. De Vreese, R. Woestenborghs, A.M. Vandamme, C.G.M. Janssen, J. Anne, G. Cauwenbergh, Proc. Natl. Acad. Sci. USA, 90 (1993) 1711.

[19]- J.M.J. Tronchet, M. Grigorov, N. Dolatshahi, F. Moriaud, J. Weber, Europ. J. Med. Chem., 32 (1997) 279.

[20]- Y.D. Cai, H. Yu, K.C. Chou, J. Protein Chem. 17 (1998) 607.

[21]- Jalali-Heravi, M. and Parastar, F., J. Chem. Inf. Comput. Sci. 40 (2000) 147.

[22]- G. Hatzakis, C. Tsoukas, Proc. AMIA Symp., (2001) 249.

[23]- Bazoui, H., Zahouily, M., Boulaajaj, S., Sebti, S. and Zakarya, D., SAR QSAR Environ. Res. 13 (2002) 567.

[24]- Vanyur, R., Heberger, K. and Jakus, J., J. Chem. Inf. Comput. Sci. 43 (2003) 1829.

[25]- T.L. Chiu, S.S. So, J. Chem. Inf. Comput. Sci., 44 (2004) 154.

[26]- Douali, L. Villemin, D., Zyad, A. and Cherqaoui, D., Mol. Divers. 8 (2004)

1.

- [27]- Douali, L., Villemin, D. and Cherqaoui, D., *Int. J. Mol. Sci.* 5 (2004) 48.
- [28]- Z.R. Yang, R. Thomson, *IEEE Trans Neural Netw.* 16 (2005) 263.
- [29]- Silvestri, R., Artico, M., La Regina, G., De Martino, G., La Colla, M. Loddo, R., La Colla, P. *Farmaco* 59 (2004) 201.
- [30]- HyperChem 7.0 HyperCube Inc., Toronto, Canada, 2002
- [31]- R. Todeschini, Milano Chemometrics and QSPR Group.  
<http://www.disat.unimib.it/vhml>.
- [32]- SPSS for windows, Statistical Package for IBM PC, SPSS Inc.,  
<http://WWW.SPSS.com>.
- [33]- MATLAB 6.1, The Math Works, Inc., Natick, MA.
- [34]- Zupan, J.J, and Gasteiger, J. "Neural Networks for Chemists. An Introduction", VCH Publishers: Weinheim (Germany), 1993.
- [35]- Rumelhart, D. E., Hinton, G.E. and McClelland, J.L. A General Framework for Parallel distributed Processing, Rumelhart, D.E. and McClelland, J.L., Eds., Cambridge, MA: Foundations, MIT Press, 1986, vol. 1.
- [36]- Hagan, M.T. and Menhaj, M.B., *IEEE Transactions on Neural Networks* 5 (1994) 989.
- [37]- Adeloye, A.J. and De Munari, A., *J. Hydro.* 326 (2006) 215.
- [38]- Tourwe, E., Pintelon, R. and Hubin, A., *J. Electroanal. Chem.* 594 (2006) 50.
- [39]- Arab Chamjangali, M. Beglari, M., Bagherian, G. *J Mol. Graph. Model.* 26 (2007) 360.
- [40]- Schuur, J.H, Selzer, P and Gasteiger, J., *J. Chem. Inf. Comput. Sci.* 36 (1996) 334.
- [41]- Todeschini, R., Lasagni, M. and Marengo, E., *J. Chemom.* 8 (1994) 263.

[42]- Consonni, V., Todeschini R., Pavan M. (2002) J. Chem. Inf. Comp. Sci. 42 (2002) 682-692.

[43]- Consonni, V., Todeschini, R., Pavan M., Gramatica, P. J. Chem. Inf. Comp. Sci. 42 (2002) 693-705.

[44]- Karelson, M., Lobanov, V.S., Katritzky, A.R. Chem. Rev. 96 (1996) 1027-1043