



حوزه معاونت پژوهشی و فناوری

گزارش پایانی طرح پژوهشی

با عنوان

پیش بینی فعالیت داروئی ضد ایدز (سیتو توکسیتی) مشتقات ۵-
فنیل-۱-فنیل آمینو-H۱-ایمیدازول به وسیله شبکه عصبی مصنوعی

کد: ۲۲۰۲۳

آذرماه ۱۳۸۶

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 ۸۵/۱۰/۲۴ و ۸۶/۶/۲۵ می باشد.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

این گزارش نتیجه طرح پژوهشی با عنوان: "پیش بینی فعالیت داروئی ضد ایدز (سیتو توکسیتی) مشتقات ۵-فنیل-۱-فنیل آمینو-H۱-ایمیدازول به وسیله شبکه عصبی مصنوعی" می باشد که با کد ۲۲۰۲۳ به تصویب شورای پژوهشی شماره ۱۵۶ مورخ ۸۵/۱۰/۲۴ رسیده است.

پیش بینی فعالیت داروئی ضد ایدز (سیتو توکسیتی) مشتقات ۵-فنیل-۱-فنیل آمینو-H۱-

ایمیدازول به وسیله شبکه عصبی مصنوعی

چکیده:

در این پروژه یک مدل شبکه عصبی با آموزش به کمک الگوریتم لونیگ مارکوت برای مطالعات QSAR مشتقات ۵-فنیل-۱-فنیل آمینو-H۱-ایمیدازول 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 ۲۸ توصیفگر مهم که با داده‌های CC_{50} 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_{50} داروهای ضد ایدز، روشهای استفاده از سری پیش بینی و روش رد مرحله‌ای تک تک بکار گرفته شد. خطای مجذور جذر میانگین (RMSE) حاصله برای دو روش فوق به ترتیب ۰/۰۴۸ و ۰/۰۲۴ به دست آمد. نتایج حاصله نشان دهنده قدرت پیش‌بینی بالای مدل شبکه عصبی نسبت به مدل MLR می‌باشد. این نتایج آشکارا بیانگر این هستند که روش QSAR ارائه شده در این پروژه قادر است خاصیت داروئی ضد ایدز هر مشتقی از ۵-فنیل-۱-فنیل آمینو-H۱-ایمیدازول با هر استخلافی را با قدرت بالا و خطای بسیار کم پیش‌بینی کند.

فهرست مطالب

عنوان.....	صفحه.....
فصل اول - مقدمه.....	۱.....
۱-۱- بیماری ایدز.....	۱.....
۱-۱-۱- تاریخچه کشف ویروس HIV.....	۲.....
۱-۱-۲- انواع HIV.....	۳.....
۱-۱-۳- میزان پایداری ویروس HIV.....	۴.....
۱-۱-۴- همانند سازی و چرخه تکثیر.....	۴.....
۱-۱-۵- بیماری‌زایی ویافته‌های ایمنی.....	۶.....
۱-۱-۶- مراحل بیماری.....	۶.....
۱-۱-۶-۱- مرحله مقدماتی (مرحله حاد).....	۶.....
۱-۱-۶-۲- مرحله بدون علامت (مزمن).....	۷.....
۱-۱-۶-۳- مرحله ایدز (بحرانی).....	۷.....
۱-۱-۷- آمار مبتلایان به ایدز.....	۷.....
۱-۱-۸- پیشگیری و درمان.....	۹.....
۱-۲- مطالعات ارتباط کمی ساختار-فعالیت (خاصیت).....	۱۱.....
۱-۲-۱- جمع‌آوری لیستی از ترکیبات.....	۱۱.....
۱-۲-۲- بهینه سازی ساختار هندسی استخراج توصیفگرها.....	۱۲.....
۱-۲-۲-۱- روشهای آغازین.....	۱۲.....
۱-۲-۲-۲- روشهای نیمه تجربی.....	۱۴.....
۱-۲-۲-۳- روشهای تابع دانسیته.....	۱۴.....
۱-۲-۲-۴- روشهای مکانیک مولکولی.....	۱۴.....
۱-۲-۳- استخراج توصیفگرها.....	۱۵.....
۱-۲-۴- انتخاب بهترین توصیف کننده‌ها.....	۱۵.....

۱۶	۱-۲-۵- مدلسازی
۱۹	۱-۲-۷- ارزیابی مدل
۱۹	۱-۳- مروری بر کارهای انجام شده
۲۲	فصل دوم-محاسبات و مدلسازی
۲۲	۱-۲- سری داده‌ها
۲۳	۲-۲- نرم افزارهای مورد استفاده
۲۵	۲-۳- انجام محاسبات و مدلسازی
۲۵	۲-۳-۱- رسم و بهینه‌سازی ساختار مولکولها
۲۶	۲-۳-۲- محاسبه توصیفگرهای مولکولی
۲۶	۲-۳-۳- انتخاب توصیفگرهای مهم
۲۷	۲-۳-۴- مدل شبکه عصبی و بهینه سازی آن
۲۹	۲-۳-۴-۱- انتخاب تعداد لایه‌های شبکه عصبی
۳۰	۲-۳-۴-۲- انتخاب الگوریتم آموزشی
۳۱	۲-۳-۴-۳- انتخاب نوع تابع تبدیل
۳۱	۲-۳-۴-۴- انتخاب تعداد ورودیها (تعداد توصیفگرها)
۳۲	۲-۳-۴-۵- انتخاب سرعت آموزش، تعداد گره‌های لایه پنهان و تعداد دورهای آموزش
۳۳	۲-۳-۴-۵- ساختار و پارامترهای بهینه شده شبکه عصبی
۳۳	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
۳۸	فصل سوم - بحث و نتیجه‌گیری
۳۸	۳-۱- توصیف‌گرهای انتخاب شده
۴۰	۳-۲- ارزیابی آماری مدل ارائه شده
۴۱	۳-۳- نتیجه‌گیری نهایی

فصل اول

مقدمه

۱-۱- بیماری ایدز

ایدز یا سندرم نقص ایمنی اکتسابی^۱ (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 سلولهای غیرطبیعی مثل سلولهای سرطانی می‌شود. ویروس ایدز سلولهای ایمنی موجود در خون (لنفوسیتها) و سلولهای ایمنی موجود در بافتها مانند مغز استخوان، طحال، کبد و گرههای لنفاوی را درگیر می‌سازد. این سلولها در تولید پادتن برای مقابله با بیماریها و سرطانها نقش دارند. در مجموع باید گفت که ایدز یک نوع نقص ایمنی ثانویه است که در سیر عفونت با ویروس ایدز ایجاد می‌شود.

عامل بیماری ایدز یک نوع ویروس از گروه رتروویروسها^۲ بنام ویروس نقص ایمنی انسانی^۳ (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 بعضا توسط پادتنها خنثی می‌شوند اما ژنهای

¹ - Acquired Immunodeficiency Syndrome

² - Reteroviruse

³ - Human Immundiciency Virus (HIV)

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. [۱]

۱-۱-۱- تاریخچه کشف ویروس HIV

در سال ۱۹۸۱، هشت مورد وخیم ابتلا به بیماری سارکوم کاپوسی^۱ 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 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^۲ 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 بیماری‌کننده ایدز به نتیجه رسید. گروهی تحقیقاتی به سرپرستی رابرت گالو^۳ 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. [۲]

^۱ - Kaposi's Sarcoma

^۲ - Slim

^۳ - Robert Gallo

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.

۱-۱-۲- انواع HIV

در حال حاضر ۲ گونه از HIV وجود دارد:

الف) HIV-1

ب) HIV-2

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

۱-۱-۳- میزان پایداری ویروس HIV

ویروس HIV در اثر خشک شدن به سرعت غیر فعال می‌شود 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 می‌ماند. در درجه حرارت ۶۰-۵۰ درجه سانتیگراد به مدت سی دقیقه عفونت زایی خود را از دست می‌دهد. همچنین این ویروس در اثر استفاده از بعضی عوامل به مدت ده دقیقه در درجه حرارت اتاق به طور کامل غیر فعال می‌شود. این عوامل هیپوکلریت سدیم ۰/۵٪، اتانول ۷۰٪، پارافرمالدئید ۰/۵٪، و قرار گرفتن در PH بیش از ۱۰ را شامل می‌شوند [۱].

۱-۱-۴- همانند سازی و چرخه تکثیر

ویروس HIV یک رتروویروس است که دستخوش جهش‌های مکرر می‌شود و بر سلول‌های دستگاه ایمنی بدن حمله می‌کند. رتروویروس‌ها ویروس‌هایی هستند که ژنوم (ماده ژنتیکی ویروس) آنها از RNA تشکیل شده است. بنابراین برای تکثیر خود به آنزیمی به نام نسخه بردار معکوس^۱ وابسته هستند که ژنوم RNA آنها را به DNA نسخه برداری می‌کند تا بعد بتواند آن را با کمک آنزیمی (آنزیم اینتگرز) وارد ژنوم سلول میزبان کند و به این ترتیب امکان تکثیر ویروس به وجود می‌آید. کلمه رترو (معکوس) در نام این نوع ویروس‌ها هم به همین خاطر است چرا که معمولاً نسخه برداری از DNA به RNA انجام می‌شود و لی در این دسته ویروسها بر عکس است.

ویروس HIV با آلوده کردن سلول‌های ویژه‌ای از سیستم ایمنی به نام "لنفوسیت‌های T کمک کننده"^۲ یا CD⁺₄ یا TCD⁺₄ باعث بیماری می‌شود. این سلول‌های زیرگروهی از گلبول‌های سفید هستند که به طور طبیعی پاسخ ایمنی به عفونت را تنظیم می‌کنند. HIV با

^۱ -Reverse Transcriptase

^۲ - T-Lymphocytes

۱-۱-۵- بیماری‌زایی ویافته‌های ایمنی

عفونت با HIV موجب واکنش ویروس با گیرنده $CD4^+$ می‌گردد که با تراکم زیاد بر روی لنفوسیت‌های T کمکی قرار دارند. بعضی از مونوسیت‌ها و ماکروفاژها نیز به میزان کمتری دارای ملکول $CD4^+$ هستند، بنابراین می‌توانند به وسیله HIV آلوده گردند. سلول‌های T4 آلوده به ویروس در سطح خود مقدار زیادی گلیکوپروتئین پوششی به نام gp 120 را تولید می‌کنند. 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 می‌شود. پیامدهای ناشی از آلودگی لنفوسیت‌های T کمکی به وسیله ویروس HIV بسیار مخرب است زیرا این سلول‌ها نقش مهمی را در دستگاه دفاعی بدن ایفا می‌کنند.

۱-۱-۶- مراحل بیماری

دوره کمون^۱ بیماری چندروز تا سه ماه است. عفونت با HIV، به سه مرحله تقسیم می‌شود [۳۶]:

۱-۱-۶-۱- مرحله مقدماتی (مرحله حاد)

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 _Prion of Incubation

۱-۱-۶-۲- مرحله بدون علامت (مزمن)

بعد از مرحله حاد دوره بدون علامت وجود دارد که ۱۱-۷ سال طول می کشد این دوره که در گذشته به نام دوره نهفتگی نامیده می شده است درحقیقت عفونت خاموشی را نمایش می دهد که طی آن تخریب مداوم و دائمی سیستم ایمنی انجام می گردد. در این دوره تعداد لنفوسیت های $CD4^+$ آلوده افزایش می یابد و اثرات تخریب سیستم ایمنی مشخص تر می شود. پاسخ های ایمنی سلولی در این دوره به طور فزاینده ای کاهش می یابد و فرد آلوده قادر نیست که نسبت به آنتی ژن های میکروبی پاسخ دهد. بایستی خاطر نشان کرد که ویروس در این مرحله و در طی مرحله سوم خصوصیتی را که در شروع بیماری داشت، دارا نمی باشد. در دوره مزمن بیماری HIV سریعتر تکثیر می یابد، محدوده سلولی آن فراتر است، برای $CD4^+$ ها مخرب تر می باشد.

۱-۱-۶-۳- مرحله ایدز (بحرانی)

پیشگویی آزمایشگاهی در مورد پیشرفت بیماری به سوی ایدز به وسیله لنفوسیت های $CD4^+$ انجام می پذیرد. هنگامی که تعداد این لنفوسیتها خیلی پائین بیاید میزان در معرض مبتلاء به عفونت های فرصت طلب قرار می گیرد 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 لنفوسیت های $CD4^+$ به کمتر از ۲۰۰ سلول در هر میکرولیتر برسد، این افراد در طبقه بیماران ایدزی قرار می گیرند.

۱-۱-۷- آمار مبتلایان به ایدز

بر اساس تازه ترین آمار سازمان ملل متحد درباره گسترش بیماری ایدز در جهان، هم اکنون نزدیک به ۳۸ میلیون نفر به ویروس HIV آلوده اند. این ویروس از زمان کشف تاکنون بیش از ۲۰ میلیون قربانی داده است. همچنین این تحقیق نشان می دهد 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.

۱-۱-۸- 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^۱ (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 [۵ و ۶].

ب) مهارکننده غیر نوکلئوزیدی^۲ (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 مشابه‌های نوکلئوزیدی عمل می‌کنند و با روشی متفاوت، جلوی آلوده کردن سلولها بوسیله ویروس HIV را می‌گیرند. [۷]

ج) مهارکننده آنزیم پروتئاز^۳ (PI)

^۱-Nucleosid Reverse Transcriptase Inhibitors

^۲-Non- Nucleosid Reverse Transcriptase Inhibitors

^۳- Protease Inhibitors

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 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^۱

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 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- Fusion or Entry Inhibitors

۲-۱- 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)^۱ 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)^۲ خواهد داشت. این ارتباط نیازمند اصول تئوری سخت و پیچیده نمی باشد و شامل مراحل زیر است [۸]:

- 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.

۱-۲-۱- 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 در مطالعات QSAR و QSPR مورد استفاده قرار می گیرد. اگر تعداد ترکیبات نسبت به پارامترهای برازش شده^۳ در معادله کمتر باشد، این برازش، صحیحی نمی باشد. اما اگر تعداد پارامترها و تعداد ترکیبات برابر باشد، برازش منحنی به طور صحیحی فراهم خواهد شد.

^۱-Quantitative structure –property relationship

^۲-Structure – activity relationship

^۳-Fitted

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

مرحله بعدی فراهم کردن ساختار هندسی مولکولهاست که بهتر است از لحاظ ساختاری بهینه شوند. در این امر از شیمی محاسباتی^۱ استفاده می شود که در اینجا اشارهای به شیمی محاسباتی و روشهای رایج در بهینه سازی به ساختار مولکولها می شود. در سالهای اخیر روشهای محاسباتی در بین شیمیدانان بسیار رواج پیدا کرده است. در این روشها می توان 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^۲ می کند و این روش به شیمیدان اجازه می دهد 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 می شوند:

۱-۲-۲-۱- روشهای آغازین^۳

در این روش از هیچ داده تجربی استفاده نمی شود، بلکه از یک سری ثابت هایی نظیر بار الکترون، جرم الکترون و ثابت پلانک استفاده می شود و تمام معادلات لازم از حل معادلات کوانتومی از جمله، معادله شرودینگر^۴ بدست می آید به همین دلیل چون از هیچ داده تجربی استفاده نمی شود به آن « روش آغازین » گفته می شود [۱۰ و ۹].

اگر برای یک سیستم که تابع موج آن را داریم، تابع هامیلتونی^۵ را به آن اثر دهیم می توانیم انرژی آن را به دست آوریم:

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

¹- COMPUTATIONAL CHEMISTRY

²- Simulation

³-Ab initio methods

⁴-Schrodinger Equation

⁵-Hamiltonian function

که در این معادله H معرف تابع هامیلتونی، ψ 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 - فاک¹ 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 شده² یا هارتری = فاک نامحدود³ 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 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

¹ -Hartree-fock

² -Restricted hartree-fock

³ - Unrestricted hartree-fock

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.

۱-۲-۲-۲- روشهای نیمه تجربی^۱

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 AM1^۲: استفاده شده است. روش AM1 همان روش MNDO^۳ 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.

۱-۲-۲-۳- روشهای تابع دانسیته^۴

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.

۱-۲-۲-۴- روشهای مکانیک مولکولی^۵

^۱ -Semi-empirical methods

^۲ -Austin methods

^۳ - Modified neglect of diatomic overlap

^۴ -Density functionl theory(DFT)

^۵ -Molecular mechanics

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.

۱-۲-۳- 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 براحتی 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.

۱-۲-۴- 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^۱ 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 موردنظر را به درستی توصیف کند. و ضریب همبستگی صفر بدین معنی است که هیچ ارتباطی بین توصیف

^۱-Intecorrelation coefficient

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

۱-۲-۵- 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 [۱۴-۱۱] و روش حداقل مربعات جزئی^۲ [۱۸-۱۵] 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:

$$property (activity) = c_0 + c_1d_1 + c_2d_2 \quad (9-1)$$

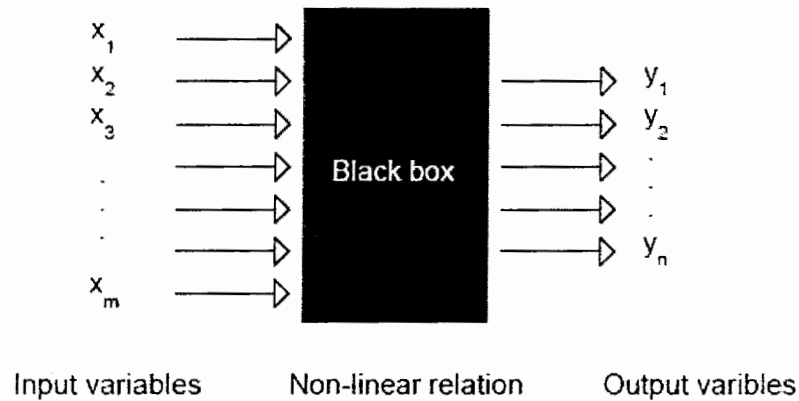
در این معادله c_i بیانگر پارامترهای برازش شده و d_i بیانگر توصیف کننده‌ها می‌باشد. دلیل استفاده از معادلات خطی، سادگی روش می‌باشد.

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 شبکه‌های عصبی مصنوعی به طور گسترده در مطالعات QSAR و QSPR بکار گرفته شده است که در اینجا به طور اجمالی شرح داده خواهد شد.

¹-Curve fitting

²-Nonlinear fitting

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 ۲-۱).



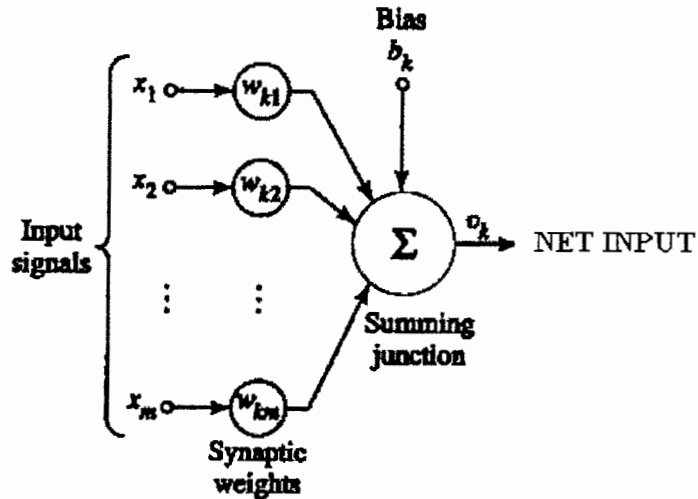
شکل (۲-۱) - نمایش یک شبکه عصبی نوعی

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

$$NET = X_1 W_{K1} + X_2 W_{K2} + \dots + X_m W_{km}$$

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 بافتار 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 Σ 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

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

۱-۲-۷- 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.

۱-۳- مروری بر کارهای انجام شده

در سال ۱۹۹۱ کوکلا^۱ پیشگویی فعالیت دارویی ضد HIV-1 مربوط به سه سری مختلف از مشتقات TIBO را با استفاده از سه معادله خطی با گزارش داده است [۱۱].

در سال ۱۹۹۷، ترونکت^۲ با استفاده از مطالعات QSAR برهمکنش HERT را با HIR، با استفاده از شبکه عصبی مصنوعی پیشگویی نمود [۱۹]

کای^۳ و همکارانش در سال ۱۹۹۸ با استفاده از مدل شبکه عصبی سایت های گسسته پروتئاز مربوط به HIV-1 در پروتئین را پیشگویی کردند [۲۰]

جلالی هروی و همکارش در سال ۲۰۰۰ در یک مطالعه QSAR، فعالیت ترکیبات ضدایدز را برای یک گروه بزرگ از مشتقات HEPT بررسی کردند. در این کار روش رگرسیون

^۱-Kukla

^۲-Tronchet

^۳-Cai

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^۱ ارائه شده است [۲۲].

در سال ۲۰۰۲، بازوئی^۲ 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 هم مقایسه شده است. نتایج بدست آمده بیانگر این واقعیت است که فعالیت ترکیبات HEPT به طور موثری وابسته به خصوصیات هیدروفوبیک و گروه‌های استخلاقی می‌باشد [۲۳].

میشل^۳ در سال ۲۰۰۳ وانیر^۴، با استفاده از شبکه عصبی فعالیت ضدایدز را برای یک سری از ترکیبات تتراپیرول پیشگویی کرد و شبکه عصبی به عنوان مناسب‌ترین مدل معرفی شده است [۲۴].

چایو^۵ در سال ۲۰۰۴ مقاله‌ای تحت عنوان توسعه مدل‌های QSPR مربوط به شبکه عصبی برای ثابت‌های استخلاقی هانس^۶ و کاربردهای آن در مطالعات QSAR مربوط به HIV-1 و بازدارنده‌های دی هیدروفیلات، ارائه داد [۲۵].

در سال ۲۰۰۴ دالی^۷ و همکارانش به مطالعات QSAR غیر خطی مشتقات HEPT پرداختند و هشت توصیف کننده ساختمانی و متغیر فیزیکوشیمیایی را در نظر گرفتند و به روابط ویژه‌ای 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 $R^2=0/۸۶۱$ را برای سری پیشگویی گزارش داد [۲۷].

¹ -Hatzakis

² -Bazoui

³ -Michele

⁴ -Vanyur

⁵ -Chiu

⁶ -Hansch substituent constant

⁷ - Dauoli

یانگ^۱ و همکارانش در سال ۲۰۰۵ با استفاده از شبکه عصبی و توابع اساسی بیو^۲، موفق به پیشگویی سایتهای گسسته پروتئاز در HIV و سایتهای گسسته تریپسین^۳ در پروتئین شدند
[۲۸].

^۱ -Yang
^۲ -Bio
^۳ -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^۱ (CC_{50}) مشتقات ۵- فینیل -۱- فینیل آمینو-۱-*H*-ایمیدازول^۲ بکارگرفته شده است و نتایج حاصله با روش رگرسیون خطی چندگانه مقایسه گردیده است.

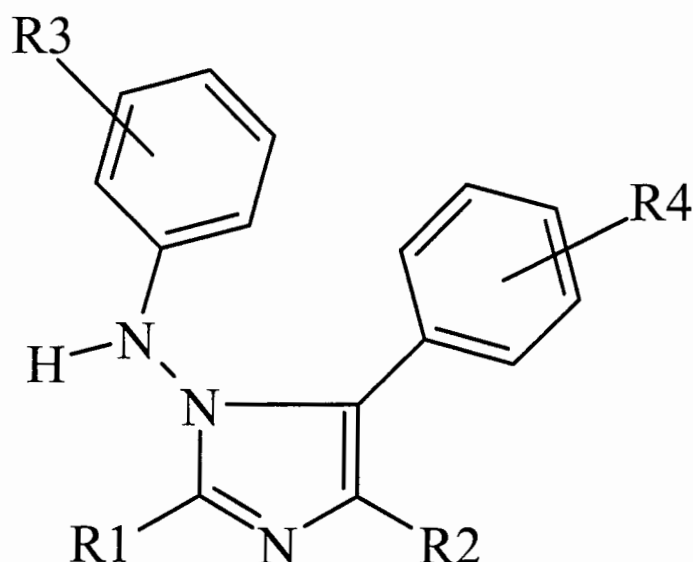
۱-۲- سری داده‌ها

سری داده‌های مورد استفاده در این پروژه، داده‌های CC_{50} برای ۴۲ ترکیب از مشتقات ۵- فینیل -۱- فینیل آمینو-۱-*H*-ایمیدازول می‌باشد که از مرجع شماره [۲۹] گردآوری شده است. مقادیر منهای لگاریتم داده‌های CC_{50} ($-\log CC_{50}$) به عنوان متغیر پاسخ در این مدل‌سازی مورد استفاده قرار گرفت. ساختار پایه مشتقات در شکل (۱-۲) آمده است که چهار استخلاف R_1 ، R_2 ، R_3 و R_4 در جایگاه‌های متفاوتی روی حلقه‌های فنیل و ایمیدازول 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

¹ - Cytotoxicity

² - 5-phenyl-1-phenylamino-1*H*-imidazole derivatives

ملاحظه 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 7 Hyper Chem [۳۰] استفاده گردید. برای محاسبه توصیفگرهای مولکولی از نرم‌افزار Dragon که به وسیله گروه QSPR و کمومتریکس میلانو^۱ [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

^۱ - Milano Chemometrics and QSPR Group

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 ₁	R ₂	R ₃	R ₄	داده های CC ₅₀
۱	CH ₃	SH	3-Cl	H	۱/۱۷۸
۲	CH ₃	SH	4-F	H	۰/۶۵۷
۳	CH ₃	SH	4-Cl	H	۰/۸۷۲
۴	CH ₃	SH	3-Cl	4-Br	۱/۴۷۸
۵	CH ₃	SH	3-Cl	3-Cl	۱/۵۰۹
۶	CH ₃	SH	3-Cl	4-Cl	۱/۳۸۹
۷	CH ₃	SH	3-Cl	4-OCH ₃	۱/۴۱۲
۸	CH ₃	SH	H	H	۰/۷۱۸
۹	CH ₃	SH	3-Br	H	۱/۳۵۴
۱۰	CH ₃	SH	3-NO ₂	H	۱/۳۱۶
۱۱	CH ₃	SH	3-F	H	۰/۹۶۶
۱۲	CH ₃	SH	3-CH ₃	H	۱/۰۵۴
۱۳	(CH ₃) ₂ CH	SH	3-CH ₃	H	۱/۳۷
۱۴	C ₂ H ₅	SH	3-CH ₃	H	۱/۲۷۱
۱۵	C ₆ H ₅	SH	3-Cl	H	۱/۴۴
۱۶	CH ₃	SH	3-OCH ₃	H	۱/۴۲
۱۷	CH ₃	SH	3-Cl	3-CN	۱/۰۰۳
۱۸	CH ₃	SH	3-Cl	3-OCOCH ₃	۱/۱۷۲
۱۹	CH ₃	SH	3-Cl	3-COOH	۰/۹۲۱
۲۰	CH ₃	SH	3-CH ₃	3-COOH	۰/۷۲۸
۲۱	CH ₃	SH	4-C ₂ H ₅	H	۱/۴۶۳
۲۲	CH ₃	SH	4-CH ₃ S	H	۱/۲۷۵
۲۳	CH ₃	H	3-Cl	H	۱/۷۵۷
۲۴	CH ₃	H	3-Cl	3-Cl	۱/۹۲۳
۲۵	CH ₃	H	3-CH ₃	H	۱/۵۶۸
۲۶	C ₂ H ₅	H	3-CH ₃	H	۱/۵۵۵
۲۷	CH ₃	H	3-OCH ₃	H	۱/۴۶۳

۲۸	CH ₃	H	3-Cl	3-CN	۱/۴۳۴
۲۹	CH ₃	H	3-Cl	3-CONH ₂	۰/۵۸۴
۳۰	CH ₃	H	3-CH ₃	3-CONH ₂	۰/۷۹۰
۳۱	CH ₃	H	3-Cl	3-COOH	۰/۷۹۴
۳۲	CH ₃	SH	2-Cl	H	۰/۷۹۱
۳۳	CH ₃	SH	3-Cl	3-Br	۱/۳۵۷
۳۴	CH ₃	SH	2,5-Cl	H	۰/۸۳۲
۳۵	C ₂ H ₅	SH	3-Cl	H	۱/۴۰۹
۳۶	CH ₃	SH	3-CH ₃	3-CN	۱/۳۴۱
۳۷	CH ₃	H	3-Cl	3-Br	۱/۷۸۵
۳۸	CH ₃	H	H	H	۱/۲۸۲
۳۹	CH ₃	H	4-F	H	۱/۵۱۱
۴۰	CH ₃	H	3,5-CH ₃	H	۱/۶۹۰
۴۱	CH ₃	H	3-CH ₃	3-CN	۱/۳۵۰
۴۲	CH ₃	H	3-Cl	3-OCOCH ₃	۱/۴۸۰

۳-۲- انجام محاسبات و مدلسازی

۳-۳-۱- 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 ۵- فینیل -۱- فینیل آمینو-۱-H-ایمیدازول 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 قرارگرفت و بهینه سازی تا زمانی ادامه یافت تا اینکه گرادیان به ۰/۰۰۱ کیلوگالری بر مول برسد.

۲-۳-۲- 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 برگزیده 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 برگزیده 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 بایستی 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 ۴۴ توصیف کننده دارای ارتباط معناداری با داده‌های CC_{50} مولکولهای مورد بررسی می‌باشند. همانطور که بعداً خواهیم گفت، از میان این ۴۴ توصیف کننده تعداد ۲۸ توصیف کننده در مدلسازی مورد استفاده قرار گرفت. این توصیفگرها به‌مراه کلاس مربوطه و نام کامل آنها در جدول (۲-۲) آمده است.

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

جدول (۲-۲) - توصیف‌گرهای بکار رفته در مدل‌سازی

شماره	Symbol	Class	Meaning
۱	H-050	Atom-centred fragments	Hydrogen attached to heteroatom
۲	RGyr	Geometrical	Radius of gyration (mass weighted)
۳	Mor22m	3D-MoRSE	3D-MoRSE-signal 22
۴	GATS7v	2D-Autocorrelation	Geary autocorrelation-lag 7
۵	RDF035m	RDF	Radial distribution function-3.5
۶	Mor13m	3D-MoRSE	3D-MoRSE-signal 13
۷	E3m	WHIM	3 rd component accessibility directional WHIM index
۸	Mor04e	3D-MoRSE	3D-MoRSE-signal 04
۹	MATS4e	2D-Autocorrelation	Moran autocorrelation-lag 4
۱۰	Mor09p	3D-MoRSE	3D-MoRSE-signal 13
۱۱	H3p	GETAWAY	H autocorrelation of lag 3
۱۲	MATS1v	2D-Autocorrelation	Moran autocorrelation-lag 1
۱۳	Mor30p	3D-MoRSE	3D-MoRSE-signal 30
۱۴	RDF110p	RDF	Radial distribution function-11.0
۱۵	G2u	WHIM	2st component symmetry directional WHIM index
۱۶	RDF085m	RDF	Radial distribution function-8.5
۱۷	Mor04p	3D-MoRSE	3D-MoRSE-signal 04
۱۸	Mor16v	3D-MoRSE	3D-MoRSE-signal 16
۱۹	BEHe7	BCUT	Highest eigenvalue n.7 of Burden index
۲۰	JGI3	Galvez Topological Charge Indices	Mean topological charge index of order 3
۲۱	RDF040u	RDF	Radial distribution function-4.0
۲۲	nO	Constitutional	Number of oxygen atoms
۲۳	qneg	Charge	Maximum negative charge
۲۴	RDF100e	RDF	Radial distribution function-10.0
۲۵	Mor10u	3D-MoRSE	3D-MoRSE-signal 10
۲۶	RDF060m	RDF	Radial distribution function-6.0
۲۷	Mor16u	3D-MoRSE	3D-MoRSE-signal 16
۲۸	RDF065e	RDF	Radial distribution function-6.5

گردید. شبکه انتخاب شده یک شبکه عصبی با جریان رو به جلو^۱ با الگوی آموزش پس انتشار^۲ 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، ۴۲ سری داده مشتقات ۵- فیئل ۱- فیئل آمینو-۱-H-ایمیدازول، به صورت تصادفی به دو زیر گروه تقسیم شدند:

الف (۳۱ ترکیب به عنوان سری آموزش انتخاب شدند که این داده‌ها برای آموزش شبکه مورد استفاده قرار گرفتند.

ب) ۱۱ 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 جذر میانگین^۳ (RMSE) حاصله در سری پیش‌بینی می‌باشد.

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

^۱ - Feed-forward ANNs

^۲ - Back propagation algorithm

^۳ - Root mean squares errors

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 در بکارگیری شبکه عصبی در شیمی، 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 ممنتوم که تا حدودی مشکل عدم همگرایی را حل می‌کند. ب) بکارگیری روشهای مرتبه دوم برای یادگیری شبکه می‌باشد. از میان روشهای مرتبه دوم الگوریتم لوبنبرگ - مارکوت^۱ (LM) یکی از رایجترین شیوه‌ها می‌باشد که هنگامی که پارامترها کارایی شبکه RMSE می‌باشد بسیار مناسب است [۳۶-۳۸].

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 پس انتشار مورد بررسی قرار گرفتند و نتایج حاصله نشان داد که از میان تمام الگوریتم‌های مورد بررسی

^۱ - Levenberg-Marquardt back propagation algorithm

قرار گرفته، الگوریتم LM دارای کمترین خطای پیش‌بینی و بعلاوه سرعت آموزش بسیار بالا می‌باشد. بنابراین الگوریتم یادگیری پس انتشار لونیبرگ - مارکوت (تابع Trainlm در جعبه ابزار نرم افزار مطلب^۱) 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_{50} 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 آموزش داده شد و سپس شبکه آموزش داده شده برای پیش‌بینی داده‌های CC_{50} در سری پیش‌بینی مورد استفاده قرار گرفت. نتایج حاصله نشان داد که با بکارگیری تابع تبدیل خطی در لایه پنهان کمترین خطای پیش‌بینی مشاهده می‌شود. بنابراین تابع تبدیل خطی (purelin در جعبه ابزار نرم افزار مطلب) برای هر دو لایه پنهان و خروجی در نظر گرفته شد.

۲-۳-۴-۴- انتخاب تعداد ورودیها (تعداد توصیفگرها)

برای یافتن بهترین تعداد توصیفگرها در مدلسازی، تعداد ورودیهای مورد استفاده در سری آموزش شبکه عصبی بررسی و بهینه شد. روش کار به اینصورت بود که شبکه‌های عصبی مختلفی با تعداد متفاوتی گره در لایه پنهان (بین ۲ تا ۱۰) و تعداد متفاوت ورودیها (بین ۱ تا

^۱ - Matlab toolbox

۴۴ ورودی) طراحی و آموزش داده شده. سپس این شبکه‌ها برای پیش‌بینی داده‌های CC₅₀ در سری پیش‌بینی بکار گرفته شدند. نتایج حاصله نشان داد که در تعداد گره‌های لایه پنهان متفاوت، تعداد ۲۸ ورودی حداقل خطای پیش‌بینی را ایجاد می‌کند. بنابراین، از میان ۴۴ توصیفگر انتخاب شده با روش رگرسیون خطی مرحله‌ای، تعداد ۲۸ توصیف گر به عنوان ورودی شبکه عصبی مورد استفاده قرار گرفت. ۲۸ توصیفگر انتخاب شده برای تمامی دسته داده ۴۲ تایی در جدول (۲-۲) نشان داده شده است.

۲-۳-۴-۵- انتخاب سرعت آموزش، تعداد گره‌های لایه پنهان و تعداد دوره‌های

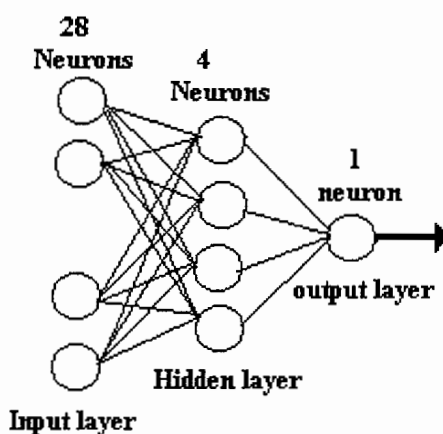
آموزش

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 RMSE سری داده‌های پیش‌بینی ندارد. بنابراین سرعت آموزش ۰/۰۱۰ برای آموزش شبکه عصبی مورد استفاده قرار گرفت.

برای انتخاب بهترین زمان آموزش (تعداد دوره‌های آموزش) و تعداد گره‌های لایه پنهان این دو پارامتر به طور همزمان بهینه شدند. به این منظور تمام ترکیبات دوتایی دو پارامتر فوق در ناحیه ۱۰-۲ با فواصل یک برای تعداد گره‌ها و در ناحیه ۱۰-۱۰۰ با فواصل ۱۰ برای تعداد دوره‌های آموزش انتخاب شده و شبکه با آنها آموزش داده شد. شبکه‌های آموزش داده شده در پیش‌بینی داده‌های CC₅₀ سری پیش‌بینی بکارگرفته شدند. نتایج حاصله نشان داد که مقدار RMSE در ناحیه ۰/۴۲ تا ۰/۱۰ برای تمام ترکیب حالات از دو پارامتر فوق، تغییر می‌کند و یک رابطه عکس بین تعداد دوره‌های آموزش و تعداد گره‌ها وجود دارد بنابراین تعداد گره‌های برابر با ۴ و تعداد دور آموزش ۱۰۰ که حداقل RMSE را ایجاد می‌کند انتخاب گردید.

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 لوبنبرگ-مارکوت انتخاب گردید. ساختار هندسی شبکه بهینه شده و پارامترهای بهینه آن شکل (۲-۲) آمده است. در شرایط بهینه حاصل حداکثر زمان لازم برای آموزش و پیش‌بینی در حدود ۷۰ ثانیه است که نشان دهنده سرعت بسیار بالای آموزش و پیش‌بینی مدل ارائه شده است.



شکل (۲-۲) - 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 ۳۱ سری از مولکولها آموزش داده شد و سپس شبکه آموزش داده شده برای پیش‌بینی داده‌های CC_{50} ، ۱۱ سری از مولکولها که در سری آموزش وجود نداشتند بکار گرفته شد. نتایج حاصله در جدول (۲-۳) نشان داده شده است. نتایج حاصله و مقایسه آنها با روش رگرسیون خطی چندگانه که برای این سری از ترکیبات در مراجع گزارش شده است [۳۹]، بیانگر این است که مدل ارائه شده در این پروژه خطاهای پیش‌بینی بسیار کمتری از روش رگرسیون خطی داراست.

جدول (۲-۳) - نتایج حاصله از ارزیابی مدل با استفاده از سری پیش‌بینی

شماره	داده های CC_{50}				
	مقادیر واقعی	مقادیر پیش‌بینی شده		درصد خطا	
		شبکه عصبی	روش MLR [۳۹]	شبکه عصبی	روش MLR
۳۲	۰/۷۹۱	۰/۷۸۹	۰/۷۹۶	۰/۲	-۰/۶۳
۳۳	۱/۳۵۷	۱/۳۶	۱/۵۱۸	-۰/۲۲	-۱۱/۹
۳۴	۰/۸۳۲	۰/۹۲۸	۰/۹۳۵	-۱۱/۵	-۱۲/۴
۳۵	۱/۴۰۹	۱/۴۳۶	۱/۳۰۱	-۱/۹	۷/۷
۳۶	۱/۳۴۱	۱/۳۱۱	۱/۱۴۷	۲/۳	۱۴/۵
۳۷	۱/۷۸۵	۱/۷۷۷	۱/۸۱۷	۰/۴۵	-۱/۸
۳۸	۱/۲۸۲	۱/۲۶۵	۱/۰۹۶	۱/۳	۱۴/۵
۳۹	۱/۵۱۱	۱/۶۰۰	۱/۳۶۱	-۵/۶	۹/۹
۴۰	۱/۶۹۰	۱/۶۹۰	۱/۷۱۹	۰/۰	-۱/۷
۴۱	۱/۳۵۰	۱/۳۵۵	۱/۴۴۷	-۰/۳۷	-۷/۲
۴۲	۱/۴۸۰	۱/۴۵۰	۱/۴۴۰	۲/۰	۲/۷

۲-۳-۵-۲- ارزیابی شبکه عصبی به روش رد مرحله ای تک تک^۱

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 ۴۱ ترکیب باقیمانده آموزش داده شد، سپس شبکه آموزش داده شده برای پیش‌بینی داده CC₅₀ ترکیب کنار گذاشته شده بکار گرفته شد. این عمل به ترتیب برای تمام ۴۲ مولکول مورد استفاده انجام گرفت. نتایج حاصله در جدول (۲-۴) خلاصه شده‌اند. نتایج بیانگر قدرت پیش‌بینی بالای مدل و قدرت تعمیم شبکه عصبی ارائه شده می‌باشد. لذا با این مدل می‌توان داده‌های CC₅₀ را برای هر ترکیب جدیدی با خطای بسیار کم و قابل قبول، پیش‌بینی کرد.

^۱- One-leave out method

جدول (۲-۴) - نتایج حاصله از ارزیابی مدل با استفاده از روش رد تک تک

شماره	داده های CC ₅₀				
	مقادیر واقعی	مقادیر پیش بینی شده		درصد خطا	
		شبکه عصبی	روش MLR [۳۹]	شبکه عصبی	روش MLR
۱	۱/۱۷۸	۱/۱۵۸	۱/۳۰۱	۱/۵	-۱۰/۶
۲	۰/۶۵۷	۰/۶۴۷	۱/۰۶۱	۱/۵	-۶۱/۵
۳	۰/۸۷۲	۰/۸۸۳	۰/۷۷۴	-۱/۳	۱۱/۲
۴	۱/۴۷۸	۱/۴۹۲	۱/۳۰۱	-۰/۹۵	۱۲/۰
۵	۱/۵۰۹	۱/۵۵۷	۱/۴۸۰	-۳/۲	۱/۹
۶	۱/۳۸۹	۱/۴۷۰	۱/۳۰۱	-۵/۸	۶/۳
۷	۱/۴۱۲	۱/۳۷۲	۱/۳۰۱	۲/۸	۷/۹
۸	۰/۷۱۸	۰/۷۱۸	۰/۷۹۶	۰/۰	-۱۰/۹
۹	۱/۳۵۴	۱/۳۳۸	۱/۳۸۱	۱/۲	-۲/۰
۱۰	۱/۳۱۶	۱/۲۸۰	۱/۳۳۸	۲/۷	-۱/۸
۱۱	۰/۹۶۶	۰/۹۵۸	۱/۱۵۹	۰/۸۳	-۲۰/۰
۱۲	۱/۰۵۴	۱/۰۸۸	۱/۲۹۱	۳/۲	-۲۲/۵
۱۳	۱/۳۷	۱/۳۵۵	۱/۲۹۱	۱/۱	۵/۸
۱۴	۱/۲۷۱	۱/۲۷۵	۱/۲۹۱	-۰/۳۱	-۱/۶
۱۵	۱/۴۴	۱/۳۹۳	۱/۳۰۱	۳/۳	۹/۷
۱۶	۱/۴۲	۱/۴۴۷	۱/۳۵۳	-۱/۹	۴/۷
۱۷	۱/۰۰۳	۱/۰۳۲	۱/۱۵۸	-۲/۹	-۱۵/۴
۱۸	۱/۱۷۲	۱/۱۷۶	۱/۱۴۰	-۰/۳۴	۲/۷
۱۹	۰/۹۲۱	۰/۸۹۰	۰/۷۰۶	۳/۴	۲۳/۳
۲۰	۰/۷۲۸	۰/۷۵۳	۰/۶۹۵	-۳/۴	۴/۵
۲۱	۱/۴۶۳	۱/۴۵۶	۱/۴۱۵	۰/۴۸	۳/۳
۲۲	۱/۲۷۵	۱/۲۴۷	1.162	۲/۲	۸/۹
۲۳	۱/۷۵۷	۱/۷۲۷	۱/۶۰۱	۱/۷	۸/۹
۲۴	۱/۹۲۳	۱/۹۸۸	۱/۷۸۰	-۳/۴	۷/۴
۲۵	۱/۵۶۸	۱/۵۳۱	۱/۵۹۰	۲/۴	-۱/۴

۲۶	۱/۵۵۵	۱/۵۲۰	۱/۵۹۰	۲/۲	-۲/۲
۲۷	۱/۴۶۳	۱/۴۸۲	۱/۶۵۲	-۱/۳	-۱۲/۹
۲۸	۱/۴۳۴	۱/۴۵۲	۱/۴۵۷	-۱/۳	-۱/۶
۲۹	۰/۵۸۴	۰/۵۴۰	۰/۷۱۱	-۷/۵	-۲۱/۸
۳۰	۰/۷۹۰	۰/۸۵۷	۰/۷۰۰	-۸/۵	۱۱/۴
۳۱	۰/۷۹۴	۰/۸۴۰	۱/۰۰۵	-۵/۸	-۲۶/۶
۳۲	۰/۷۹۱	۰/۷۷۲	۰/۷۹۶	۲/۴	-۰/۶۳
۳۳	۱/۳۵۷	۱/۳۶۹	۱/۵۱۸	-۰/۸۸	-۱۱/۹
۳۴	۰/۸۳۲	۰/۹۰۰	۰/۹۳۵	-۸/۲	-۱۲/۴
۳۵	۱/۴۰۹	۱/۴۳۷	۱/۳۰۱	-۲/۰	۷/۷
۳۶	۱/۳۴۱	۱/۲۸۲	۱/۱۴۷	۴/۴	۱۴/۵
۳۷	۱/۷۸۵	۱/۷۷۷	۱/۸۱۷	۰/۴۵	-۱/۸
۳۸	۱/۲۸۲	۱/۲۳۰	۱/۰۹۶	۴/۱	۱۴/۵
۳۹	۱/۵۱۱	۱/۵۸۸	۱/۳۶۱	-۵/۱	۹/۹
۴۰	۱/۶۹۰	۱/۶۸۶	۱/۷۱۹	۰/۲۴	-۱/۷
۴۱	۱/۳۵۰	۱/۳۶۳	۱/۴۴۷	۰/۹۶	-۷/۲
۴۲	۱/۴۸۰	۱/۴۵۰	۱/۴۴۰	۲/۰۳	۲/۷

فصل سوم

بحث و نتیجه گیری

با توجه به اینکه تا به حال هیچ گزارشی دربارهٔ بکارگیری شبکه‌های عصبی مصنوعی در مطالعات QSAR مشتقات ۵- فینل -۱- فینل آمینو-۱-H-ایمیدازول. 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 پیش‌انجام گرفت. و در نهایت ۲۸ 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.

توصیف‌گر H-050 نشان دهنده تعداد اتم هیدروژن متصل به هترواتم می‌باشد و توصیف‌گر n_0 بیانگر تعداد اتم‌های اکسیژن است. علت استفاده از این توصیف‌گرها در مدلسازی این است که داده‌های تجربی نشان می‌دهند که حضور گروه‌های دهنده پیوند هیدروژنی در

موقعیت متای حلقه فنیل که در موقعیت شماره ۵ هسته ایمدازول قرار گرفته است، میزان سیتوتوکسی‌سی تی را کاهش می‌دهد.

علت ظاهر شدن شعاع ژیرو (Rgyr) در مدل این است که این پارامتر بیانی از اندازه متوسط مولکول می‌باشد. حال با توجه به اینکه اندازه متوسط یک خاصیت ساده و مهم مولکول می‌باشد که اطلاعاتی را درباره بر هم کنش مولکول با محیط پیرامون آن یا مولکولهای هدف می‌دهد [۴۰]. لذا توصیفگر Rgyr به عنوان یک توصیفگر مهم در مدل ظاهر شده است.

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 در مدل‌سازی مورد استفاده قرار گرفته‌اند زیرا بیانی از آرایش سه‌بعدی اتمها زیادی از مولکولها با تفاوت‌های ساختاری زیاد را دارند.

توصیف‌گرهای Whim حاوی اطلاعات سه بعدی نظیر نحوه توزیع اتمی با توجه یک مرجع در مولکول می‌باشند [۴۲]. برخی از این توصیف‌گرها (f_{3m}, G_{2u}) در مدل ظاهر شده‌اند و بیانگر تأثیر کنفورماسیون دارو در حین بر هم کنش با ویروس می‌باشند.

توصیف‌گرهای اندیسه‌های بار گالوز تحت تأثیر حضور هترواتمها با توجه به مقادیر الکترونگاتیوی آنها می‌باشند. این توصیف‌گرها حاوی اطلاعات مهمی در رابطه با ساختار ترکیب و فعالیت آن با توجه به توپولوژی مولکول و انتقال بار در مولکول می‌باشند [۴۳ و ۴۴] به ویژه در برگیرنده اطلاعاتی در مورد تغییر نسبی موقعیت گروههای الکترون‌کشنده روی حلقه فنیل می‌باشند. توصیف‌گر Jgi3 از میان این توصیف‌گرها در مدل ظاهر شده است و علت حضور و اهمیت این توصیف‌گر این است که حضور استخدا‌فهای الکترون‌کشنده در موقعیت پارای حلقه فنیل ۱- فنیل آمینو عامل مناسبی برای خاصیت دارویی این مشتقات نمی‌باشد.

توصیف‌گرهای تابع توزیع شعاعی (RDF) [۴۵] به صورت احتمال یافتن یک اتم در یک حجم کروی با شعاع معینی تفسیر می‌شوند. علت حضور توصیف‌گرهای RDF در مدل‌سازی به این علت است که عدم حضور هر استخلافی در موقعیت ۳ و ۲ حلقه فنیل جزء ۱- فنیل میزان CC_{50} را کاهش می‌دهد.

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 (۲-۳) و (۲-۴) آمده است. برای بررسی میزان کارایی مدل ارائه شده و مقایسه آن با روش رگرسیون خطی چندگانه چهار پارامتر آماری زیر محاسبه شدند.

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

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

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

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

نتایج حاصل در جدول (۳-۱) آمده است. نتایج حاصله نشان دهنده قدرت پیش‌بینی بالای مدل شبکه عصبی نسبت به مدل MLR می‌باشد. این نتایج آشکار بیانگر این هستند که روش QSAR ارائه شده در این پروژه قادر است خاصیت داروی ضد ایدز هر مشتقی از ۵- فنیل ۱- فنیل آمینو-۱-H-ایمیدازول با هر استخلافی را با قدرت بالا و خطای بسیار کم پیش‌بینی کند.

جدول (۳-۱) - پارامترهای آماری محاسبه شده

پارامتر	مدل پیشنهادی		روش MLR	
	سری پیش بینی	کل داده ها	سری پیش بینی	کل داده ها
MAE	۰/۰۲۸	۰/۰۲۶	۰/۰۷۷	۰/۱۲۰
REP	۳/۶۰	۳/۲۷	۱۰/۲۵	۱۲/۰۷
RMSE	۰/۰۴۱۸	۰/۰۳۷	۰/۱۱۹	۰/۱۳۵
R ²	۰/۹۸۰۸	۰/۹۸۷۷	۰/۸۵۰۰	۰/۸۳۱۳

۳-۳- نتیجه گیری نهایی

در این پروژه یک مدل شبکه عصبی با آزمایش به کمک الگوریتم لونیگ مارکوت برای مطالعات QSAR مشتقات ۵- فینل -۱- فینل آمینو-۱-H-ایمیدازول می‌باشد. بر اساس اطلاعات موجود، این اولین گزارش در مورد بکارگیری شبکه عصبی با الگوریتم آموزش لونیگ 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 ۵- فینل -۱- فینل آمینو-۱-H-ایمیدازول برقرار نماید. این مدل می‌تواند کمک بسیار زیادی در طراحی داروهای ضد ایدز از این مشتقات با استخلافهای مختلف نماید.

مراجع

- [۱]- موسوی جاهد، "ویروس شناسی پزشکی"، انتشارات دانشگاه علوم پزشکی و پیراپزشکی شهید بهشتی، ۱۳۷۸.
- [۲]- مرعشی، م، "ویروس شناسی انسانی"، انتشارات دانشگاه فردوسی مشهد، ۱۳۸۲.
- [3]- R. Garg, S.P. Gupta, H. Gao, M.S. Babu, A.K. Debnath, C. Hansch, Chem. Rev., 99 (1999) 3525.
- [4]- http://www.tthivclinic.com/life_cycle.htm
- [5]- 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.
- [6]- 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.
- [7]- 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.
- [8]- 11. Yasri, A. and Hartsough, D., J. Chem. Inf. Comput. Sci. 41 (2001) 1218.
- [9]- D.C. Young, "Computational Chemistry: A", John Wiley & Sons Inc., New Yourk, 2001.
- [10]- J.B. Foresman, "Exploring Chemistry with Electronic Structure methods", Gaussian Inc., Pittsburgh, 1996.
- [11]- 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.
- [12]- Miri, R., Javidnia, K., Hemmateenejad, B., Azarpira, A. and Amirghofran, Z., Bioorg. Med. Chem. 12 (2004) 2529.

- [13]-. Bajaj, S., Sambhi, S.S. and Madan, A.K., *Bioorg. Med. Chem. Lett.* 15 (2005) 467.
- [14]-. 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.
- [15]- Leonard, J. T. and Roy, K., *Bioorg. Med. Chem.* 14 (2006) 1039.
- [16]-. Lin, L., Lin, W. Q , Jiang, J. H., Shen, G. L. and Yu, R. Q., *Eur. J. Pharmaceut. Sci.* 25 (2005) 245.
- [17]- De Lima Ribeiro, F.A. and Castro Ferreira, M.M., *J. Mol. Struct.: THEOCHEM* 719 (2005) 191.
- [18]- De Paula da Silva, C.H.T., Sanches, S.M. and Taft, C.A., *J. Mol. Graph. Model.* 23 (2004) 89.
- [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]- Lagoja, I. M., Pannecouque, C., Van Aerschot, A., Witvrouw, M., Debyser, Z., Balzarini, J., Hardewijn, P. and De Clercq, E., *J. Med. Chem.* 46 (2003) 1546.
- [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]- Roy, K. and Leonard, J.T., *Bioorg. Med. Chem.* 13 (2005) 2967.
- [40]- Arteca, G.A., *J. Mol. Struct. THEOCHEM* 630 (2003) 113.
- [41]- Schuur, J.H, Selzer, P and Gasteiger, J., *J. Chem. Inf. Comput. Sci.* 36 (1996) 334.
- [42]- Todeschini, R., Lasagni, M. and Marengo, E., *J. Chemom.* 8 (1994) 263.
- [43]- Galvez, J. Garcia, R. Salabert, M.T. and Soler, R., *J. Chem. Inf. Comput. Sci.* 34 (1994) 520.

[44]- Galvez, J., Garcia-Domenech, R., de Julihn-Ortiz, J.V. and Soler, R., J. Chem. Inf. Comput. Sci. 35 (1995) 272.

[45]- Hemmer, M.C., Steinhauer, V. and Gasteiger, J., J. Vib. Spectrosc. 19 (1999) 151.