





دانشکده: علوم پایه

گروه: شیمی

عنوان:

کاربرد مطالعات رابطه کمی ساختار – فعالیت برای پیش‌بینی و
مدل‌سازی فعالیت دارویی برخی از داروهای سنتز شده جدید

ارائه‌دهنده: الهه شیرپور

استاد راهنما:

دکتر منصور عرب چم جنگلی

استاد مشاور:

دکتر قدمعلی باقریان دهقی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

تیر ۱۳۸۸



دانشگاه صنعتی شاهرود

دانشکده شیمی

پایان نامه کارشناسی ارشد

الهه شیرپور

تحت عنوان

کاربرد مطالعات رابطه کمی ساختار-فعالیت برای پیش بینی و مدلسازی فعالیت دارویی برخی از داروهای سنتز شده جدید

در تاریخ ۱۳۸۸/۴/۱۷ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مورد ارزیابی و با درجه عالی مورد پذیرش قرار گرفت.

امضاء	مرتبه علمی	نام و نام خانوادگی	اعضای هیات داوران
	دانشیار	منصور عرب چم جنگلی	استاد راهنمای اول
	استادیار	--	استاد راهنمای دوم
	استادیار	قدمعلی باقریان دهقی	استاد مشاور
	استادیار	سیدعلینقی طاهری	نماینده شورای تحصیلات تکمیلی
	استادیار	زهرا کلاترکهدمی	استاد ممتحن
	استادیار	ناصر گودرزی	استاد ممتحن

تقدیم به:

آن که دانش شکوفایی خود را از طلوع

قامت او

به انتظار نشست

و

عدالت آزادی خود را از دستان او طلب

می‌کند.

از خداوند متعال برای توفیق قدم نهادن در راه دانش؛

از مادر و مرحوم پدرم برای فراهم ساختن زمینه‌های مادی و معنوی پیشرفت علمی برای من؛

از همسر مهربانم آقای کامران کمالی برای همیاری و بردباری در تمام سختی‌های دوره‌ی کارشناسی ارشد؛

از خانواده‌ی خود و همسرم برای پشتیبانی من در تمام مراحل تحصیل؛

از استاد عزیزم آقای دکتر منصور عرب چم جنگلی برای راهنمایی‌های بی‌دریغ و فراتر از وظیفه‌ی یک استاد راهنما؛

از آقای دکتر قدمعلی باقریان برای مشاوره‌ی بی‌دریغ در طول تحقیق؛

و از همه‌ی کسانی که در انجام این تحقیق هر نوع کمک مادی یا معنوی را مبذول داشتند؛

کمال تشکر و قدردانی به عمل می‌آورم.

دانشجو تأیید می نماید که مطالب مندرج در این پایان نامه نتیجه تحقیقات خودش می باشد و در صورت استفاده از نتایج دیگران مرجع آن را ذکر نموده است.

کلیه حقوق مادی مترتب از نتایج مطالعات، آزمایشات و نوآوری ناشی از تحقیق موضوع این پایان نامه متعلق به دانشگاه صنعتی شاهرود می باشد.

بهار ۱۳۸۸

چکیده

آنزیم‌های تیروزین کیناز میانجی‌های مهم آبشار پیام‌دهی هستند که نقش کلیدی تعیین‌کننده‌ای در فرایندهای بیولوژیکی گوناگون دارند. تحقیقات نشان داده که کاهش فعالیت این آنزیم در بهبود برخی سرطان‌ها تأثیر به‌سزایی دارد. در بررسی اول رابطه‌ی کمی ساختار - فعالیت ۶۰ آنالوگ ۴- (۳ برمو) -۶ و ۷- دی متوکسی کینازولین که بازدارنده‌های تیروزین کیناز گیرنده‌ی فاکتور رشد اپیدرمی می‌باشند، با استفاده از شبکه‌ی عصبی مصنوعی با الگوریتم آموزشی تنظیم بایزین مورد مطالعه قرار گرفت و با روش رگرسیون خطی چندگانه مقایسه گردید. از میان تعداد زیادی توصیف‌گر محاسبه شده، فقط ۹ توصیف‌گر بوسیله‌ی رگرسیون مرحله به مرحله انتخاب و سپس از آن‌ها به عنوان ورودی شبکه استفاده شد. سری داده‌ها به طور تصادفی به ۴۵ ترکیب سری آموزش، ۱۰ ترکیب سری ارزیابی و ۵ ترکیب سری تست تقسیم گردید. مدل‌های رگرسیون خطی چندگانه و شبکه عصبی مصنوعی ساخته شدند. توانایی پیش‌بینی مدل‌ها با استفاده از سری ارزیابی، سری تست، روش رد تک تک مورد ارزیابی قرار گرفت و خطای مجذور میانگین (MSE) برای سری ارزیابی، سری تست و روش رد تک تک بترتیب ۰/۱۳۲۹، ۰/۱۹۳۷ و ۰/۳۳۶۷ می‌باشد. نتایج حاصله نشان‌دهنده‌ی قدرت پیش‌بینی بالای مدل شبکه‌ی عصبی مصنوعی می‌باشد.

در بررسی دوم QSAR بازدارنده‌های نکروپتوسیس (نکروستاتین-۵) مورد مطالعه قرار گرفتند (نکروپتوسیس نوعی مرگ سلولی می‌باشد که با نکروسیس و آپوپتوسیس متفاوت است). مجموعه‌ی داده‌های مورد بررسی شامل ۱۶۱ ترکیب از مشتقات نکروستاتین-۵ می‌باشد. فعالیت این ترکیبات با استفاده از روش شبکه‌ی عصبی مصنوعی با الگوریتم آموزشی تنظیم بایزین مورد بررسی قرار گرفت. مدل طراحی شده قادر است که با دقت ۱۰۰٪ ترکیبات مورد نظر را از نظر بازدارندگی به دو دسته‌ی فعال و غیرفعال تقسیم کند. در قسمت دوم با ۵۲ مورد از ترکیبات فعال و ایجاد ۸ توصیف‌گر مدل‌هایی با قدرت پیش‌گویی مستقیم PEC_{50} بر حسب مولار مربوط به هر ترکیب ایجاد شد. خطای

مجذور میانگین (MSE) برای سری ارزیابی، سری تست و روش رد تک تک به ترتیب ۰/۰۶۵۱، ۰/۰۷۶۷ و ۰/۰۵۸۹ می باشد.

کلید واژه‌ها: ارتباط کمی ساختار- فعالیت، توصیف گر مولکولی، فعالیت دارویی، بازدارنده، شبکه‌ی عصبی مصنوعی

فصل اول: مقدمه

۱-۱- ارتباط کمی ساختار - فعالیت.....	۱
۱-۱-۱- جمع آوری لیستی از ترکیبات.....	۲
۱-۱-۲- بهینه سازی ساختار هندسی.....	۳
۱-۲-۱- روش های آغازین.....	۴
۱-۲-۲- روش های نیمه تجربی.....	۵
۱-۲-۳- روش های تابعی چگالی.....	۶
۱-۲-۴- روش های مکانیک مولکولی.....	۶
۱-۳- استخراج توصیفگرها.....	۶
۱-۴- انتخاب بهترین توصیف کننده ها.....	۷
۱-۵- مدل سازی.....	۸
۱-۶- رگرسیون خطی چندگانه (MLR).....	۹
۱-۷- شبکه عصبی مصنوعی.....	۱۴
۱-۷-۱- اساس بیولوژیکی شبکه های عصبی مصنوعی.....	۱۵
۱-۷-۲- ساختار و عملکرد شبکه های عصبی مصنوعی.....	۱۷
۱-۷-۳- توابع محرک.....	۲۰
۱-۷-۴- آموزش شبکه های عصبی مصنوعی.....	۲۲
۱-۷-۵- الگوریتم آموزشی پس انتشار	
خطا.....	۲۲
۱-۸- ارزیابی مدل.....	۲۸
۱-۳- سرطان و اهمیت مطالعه ی آن.....	۳۳
۱-۳-۱- تیروزین کیناز- نقش و اهمیت آن در سرطان.....	۳۵
۱-۳-۲- گیرنده ی فاکتور رشد اپیدرمی و رابطه ی آن با سرطان.....	۳۷
۱-۴- مرگ سلول.....	۳۹
۱-۵- مروری بر کارهای انجام شده در QSAR ترکیبات دارویی.....	۴۴

فصل دوم: محاسبات و مدل سازی

۱-۲- محاسبات و مدلسازی بازدارنده ی تیروزین کیناز گیرنده ی فاکتور رشد اپیدرمی.....	۵۱
۱-۱-۲- سری داده ها.....	۵۱
۱-۲- نرم افزارهای مورد استفاده.....	۵۲
۱-۳- انجام محاسبات و مدل سازی.....	۵۳
۱-۳-۱- رسم و بهینه سازی ساختار	
مولکول ها.....	۵۳

مولکولی.....	۲-۳-۱-۲- محاسبه‌ی توصیف‌گرهای
مهم.....	۳-۳-۱-۲- انتخاب توصیف‌گرهای
.....	۴-۳-۱-۲- روند ایجاد مدل‌های
.....	۵-۳-۱-۲- مدل رگرسیون خطی چندگانه
.....	۵۸- (MLR)
.....	۱-۵-۳-۱-۲- ارزیابی مدل رگرسیون خطی چندگانه‌ی ارائه شده با استفاده از پارامترهای
.....	آماري.....
.....	۶۴-.....
.....	۶-۳-۱-۲- مدل شبکه عصبی مصنوعی و بهینه‌سازی
.....	آن.....
.....	۱-۶-۳-۱-۲- انتخاب تعداد لایه‌های شبکه
.....	عصبی.....
.....	۲-۶-۳-۱-۲- انتخاب الگوریتم آموزشی، تعداد متغیرهای ورودی شبکه، انتخاب نوع تابع تبدیل و تعداد
.....	گره‌های لایه
.....	پنهان.....
.....	۳-۶-۳-۱-۲- انتخاب تعداد دوره‌های
.....	آموزش.....
.....	۴-۶-۳-۱-۲- هندسه‌ی شبکه عصبی مصنوعی بهینه
.....	شده.....
.....	۷-۳-۱-۲- ارزیابی مدل شبکه عصبی مصنوعی ارائه
.....	شده.....
.....	۱-۷-۳-۱-۲- ارزیابی شبکه با استفاده از سری
.....	ارزیابی.....
.....	۲-۷-۳-۱-۲- ارزیابی شبکه با استفاده از سری
.....	تست.....
.....	۳-۷-۳-۱-۲- ارزیابی شبکه به روش رد مرحله‌ای تک
.....	تک.....
.....	۴-۷-۳-۱-۲- آزمون Y -
.....	تصادفی.....
.....	۵-۷-۳-۱-۲- ارزیابی مدل شبکه عصبی ارائه شده با استفاده از پارامترهای آماری.....
.....	۲-۲- محاسبات و مدل‌سازی بازدارنده‌ی نکرپتوسیس (نکروستاتین-)
.....	۸۲-.....
.....	۱-۲-۲- سری
.....	داده‌ها.....
.....	۸۲-.....
.....	۲-۲-۲- نرم افزارهای مورد
.....	استفاده.....
.....	۳-۲-۲- انجام محاسبات و
.....	مدل‌سازی.....
.....	۱-۳-۲-۲- رسم و بهینه‌سازی ساختار
.....	مولکول‌ها.....
.....	۲-۳-۲-۲- طبقه‌بندی ترکیبات نکروستاتین-
.....	۵.....
.....	۱-۲-۳-۲-۲- محاسبه و انتخاب توصیف‌گرهای
.....	مهم.....

آن.....	۲-۲-۳-۲-۲- مدل شبکه عصبی مصنوعی برای طبقه‌بندی ترکیبات نکروستاتین-۵ و بهینه‌سازی
۹۳.....	
آن.....	۲-۲-۳-۲-۲- ارزیابی شبکه برای طبقه‌بندی ترکیبات نکروستاتین-۵ با استفاده از سری
تست.....	
۹۸.....	
فعال.....	۲-۲-۳-۳- مجموعه داده‌های مولکول‌های
۹۸.....	
مولکولی.....	۲-۲-۳-۴- محاسبه‌ی توصیف‌گرهای
۹۸.....	
مهم.....	۲-۲-۳-۵- انتخاب توصیف‌گرهای
۹۹.....	
۲-۲-۳-۶- روند ایجاد مدل‌های QSAR	
۹۹.....	
آن.....	۲-۲-۳-۷- مدل شبکه عصبی مصنوعی و بهینه‌سازی
۱۰۰.....	
عصبی.....	۲-۲-۳-۱-۷- انتخاب تعداد لایه‌های شبکه
۱۰۰.....	
پنهان.....	۲-۲-۳-۲-۷- انتخاب تعداد متغیرهای ورودی شبکه، انتخاب نوع تابع تبدیل و تعداد گره‌های لایه
۱۰۱.....	
آموزش.....	۲-۲-۳-۷-۳- انتخاب تعداد دوره‌های
۱۰۴.....	
شده.....	۲-۲-۳-۷-۴- هندسه‌ی شبکه عصبی مصنوعی بهینه
۱۰۶.....	
شده.....	۲-۲-۳-۸- ارزیابی مدل ارائه
۱۰۷.....	
ارزیابی.....	۲-۲-۳-۸-۱- ارزیابی شبکه با استفاده از سری
۱۰۷.....	
تست.....	۲-۲-۳-۸-۲- ارزیابی شبکه با استفاده از سری
۱۰۸.....	
تک.....	۲-۲-۳-۸-۳- ارزیابی شبکه به روش رد مرحله‌ای تک
۱۰۹.....	
تصادفی.....	۲-۲-۳-۸-۴- آزمون Y –
۱۱۲.....	
آمارى.....	۲-۲-۳-۸-۵- ارزیابی مدل شبکه عصبی ارائه شده با استفاده از پارامترهای
۱۱۳.....	

فصل سوم: بحث و نتیجه گیری

۱-۳.....	۳-۱- بررسی نتایج مدل‌سازی بازدارنده‌ی تیروزین کیناز.....
۱۱۴.....	
۱-۱-۳.....	۳-۱-۱- توصیف‌گرهای انتخاب شده.....
۱۱۴.....	
۲-۱-۳.....	۳-۱-۲- نتیجه‌گیری نهایی.....
۱۲۱.....	
۲-۳.....	۳-۲- بررسی نتایج مدل‌سازی بازدارنده‌های نکرپتوزیس، نکروستاتین – ۵.....
۱۲۳.....	
۱-۲-۳.....	۳-۲-۱- توصیف‌گرهای انتخاب شده.....
۱۲۳.....	
۱-۲-۳.....	۳-۲-۱- نتیجه‌گیری نهایی.....
۱۲۶.....	

۱۲۷.....	۳-۳- آینده‌نگری.....
۱۲۸.....	فهرست منابع و مراجع.....

فهرست اشکال

۱۷.....	شکل (۱-۱)- ساختار نرون طبیعی.....
۱۸.....	شکل (۲-۱)- مدل نرون مصنوعی.....
۲۰.....	شکل (۳-۱)- شبکه های عصبی مصنوعی چندلایه.....
۲۱.....	شکل (۴-۱)- دو نمونه از توابع فعال ساز.....
۲۴.....	شکل (۵-۱)- شمای یک شبکه عصبی مصنوعی سه لایه و الگوریتم پس انتشار خطا در این الگوریتم.....
۳۶.....	شکل (۶-۱)- مکانیسم عمل تیروزین کیناز.....
۳۷.....	شکل (۷-۱)- طرح کلی مکانیسم های مختلف که منجر به فعالیت تومورزایی تیروزین کیناز می شوند.....
۴۱.....	شکل (۸-۱)- ویژگی های مورفولوژی آپوپتوسیس.....
۴۲.....	شکل (۹-۱)- واکنش های آبشاری فعال شدن کاسپازها توسط گیرنده های مرگ.....
۴۳.....	شکل (۱۰-۱)- ساختار نکروستاتین ها.....
۵۲.....	شکل (۱-۲)- اسکلت مولکولی ترکیبات مورد بررسی.....
	شکل (۲-۲)- نمودار پراکندگی مقادیر پیش بینی شده ی PIC ₅₀ با مدل رگرسیون خطی چند گانه بر حسب مقادیر واقعی در سری ارزیابی.....
۶۱.....	شکل (۳-۲)- نمودار پراکندگی مقادیر پیش بینی شده ی PIC ₅₀ با مدل رگرسیون خطی چند گانه بر حسب مقادیر واقعی در سری تست.....
	شکل (۴-۲)- نمودار پراکندگی مقادیر پیش بینی شده ی PIC ₅₀ با مدل رگرسیون خطی چند گانه بر حسب مقادیر واقعی برای کل داده ها.....
۶۳.....	شکل (۵-۲)- نمودار باقیمانده ها برای کل داده ها (باقیمانده ی مقادیر پیش بینی شده ی PIC ₅₀ توسط مدل رگرسیون خطی چند گانه از مقادیر واقعی بر حسب مقادیر تجربی برای کل ترکیبات).....

- شکل (۶-۲) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم زیگموئید..... ۶۸
- شکل (۷-۲) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت هیپربولیک..... ۶۹
- شکل (۸-۲) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکورت و تابع تبدیل لگاریتم زیگموئید..... ۷۰
- شکل (۹-۲) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکورت و تابع تبدیل تانژانت هیپربولیک..... ۷۱
- شکل (۱۰-۲) - تغییرات مقادیر خطای مجذور میانگین بر حسب تعداد گره‌های مختلف در لایه پنهان..... ۷۴
- شکل (۱۱-۲) - ساختار هندسی شبکه عصبی استفاده شده..... ۷۴
- شکل (۱۲-۲) - ترسیمی از مقادیر پیش‌بینی شده‌ی PIC₅₀ با شبکه عصبی مصنوعی بر حسب مقدار واقعی در سری ارزیابی..... ۷۶
- شکل (۱۳-۲) - ترسیمی از مقادیر پیش‌بینی شده‌ی PIC₅₀ با شبکه عصبی مصنوعی بر حسب مقدار واقعی در سری تست..... ۷۷
- شکل (۱۴-۲) - ترسیمی از مقادیر پیش‌بینی شده‌ی PIC₅₀ با شبکه عصبی مصنوعی بر حسب مقدار واقعی برای کل داده‌ها..... ۷۹
- شکل (۱۵-۲) - نمودار باقیمانده‌ها برای کل داده‌ها (باقیمانده‌ی مقادیر پیش‌بینی شده‌ی PIC₅₀ توسط شبکه عصبی مصنوعی از مقادیر واقعی بر حسب مقادیر تجربی برای کل ترکیبات)..... ۷۹
- شکل (۱۶-۲) - روش سنتز Nec-5 شرایط و واکنش گرها..... ۸۳
- شکل (۱۷-۲) - اسکلت مولکولی ترکیبات مورد بررسی ۸۳
- شکل (۱۸-۲) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم زیگموئید..... ۱۰۲
- شکل (۱۹-۲) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت هیپربولیک..... ۱۰۳
- شکل (۲۰-۲) - تغییرات مقادیر MSE بر حسب تعداد گره‌های مختلف در لایه پنهان..... ۱۰۵
- شکل (۲۱-۲) - ساختار هندسی شبکه عصبی استفاده شده..... ۱۰۶
- شکل (۲۲-۲) - ترسیمی از مقادیر پیش‌بینی شده‌ی PEC₅₀ با شبکه عصبی مصنوعی بر حسب مقدار واقعی در سری ارزیابی..... ۱۰۸
- شکل (۲۳-۲) - ترسیمی از مقادیر پیش‌بینی شده‌ی PEC₅₀ بر حسب مقدار واقعی در سری تست..... ۱۰۹
- شکل (۲۴-۲) - ترسیمی از مقادیر پیش‌بینی شده‌ی PEC₅₀ با شبکه عصبی مصنوعی بر حسب مقدار واقعی برای کل داده‌ها..... ۱۱۱
- شکل (۲۵-۲) - نمودار باقیمانده‌ها برای کل داده‌ها (باقیمانده‌ی مقادیر پیش‌بینی شده‌ی PEC₅₀ توسط شبکه عصبی مصنوعی از مقادیر واقعی بر حسب مقادیر تجربی برای کل ترکیبات)..... ۱۱۱

فهرست جداول

جدول (۱-۱) - معمول ترین توابع فعال ساز.....	۲۱
جدول (۱-۲) - جزئیات ساختار مولکولی ترکیبات مور _ _ _	۵۴
جدول (۲-۲) - تعداد کل توصیف گرهای انتخاب شده بوسیله ی روش رگرسیون خطی چندگانه.....	۵۵
جدول (۳-۲) - مقادیر ضریب همبستگی (R^2)، آماره ی F و خطای استاندارد برای مدل های ۲ تا ۷.....	۵۸
جدول (۴-۲) - مقادیر خطای مجذور میانگین و ضریب همبستگی (R^2) محاسبه شده برای سری ارزیابی برای مدل های (توصیف گرهای) ۲ تا ۷.....	۶۰
جدول (۵-۲) - مقادیر مشاهده شده و پیش بینی شده ی PIC ₅₀ برای ترکیبات مجموعه ی سری ارزیابی با معادله ی رگرسیون خطی چند گانه (۷ توصیف گر).....	۶۱
جدول (۶-۲) - نتایج حاصل از ارزیابی مدل رگرسیون خطی چندگانه با استفاده از سری تست	۶۱
جدول (۷-۲) - نتایج حاصل از ارزیابی مدل رگرسیون خطی چندگانه با استفاده از روش رد تک تک.....	۶۲
جدول (۸-۲) - پارامترهای آماری محاسبه شده برای مدل رگرسیون خطی چندگانه.....	۶۴
جدول (۹-۲) - مقادیر خطای مجذور میانگین برای شبکه هایی با الگوریتم آموزشی تنظیم بایزین (trainbr) و تابع تبدیل لگاریتم زیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گرهای مختلف در لایه پنهان.....	۶۸
جدول (۱۰-۲) - مقادیر خطای مجذور میانگین برای شبکه هایی با الگوریتم آموزشی تنظیم بایزین (trainbr) و تابع تبدیل تانژانت هیپربولیک (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گرهای مختلف در لایه پنهان.....	۶۹
جدول (۱۱-۲) - مقادیر خطای مجذور میانگین برای شبکه هایی با الگوریتم آموزشی لونیبرگ - مارکورت (trainlm) و تابع تبدیل لگاریتم زیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گرهای مختلف در لایه پنهان.....	۷۰
جدول (۱۲-۲) - مقادیر خطای مجذور میانگین برای شبکه هایی با الگوریتم آموزشی لونیبرگ - مارکورت (trainlm) و تابع تبدیل تانژانت هیپربولیک (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گرهای مختلف در لایه پنهان.....	۷۱

جدول (۲-۱۳) - توصیف گره‌های بکار رفته در مدلسازی.....	۷۲
جدول (۲-۱۴) - ماتریس همبستگی برای توصیف گره‌های انتخاب شده.....	۷۲
جدول (۲-۱۵) - مقادیر خطای مجذور میانگین برای تعداد گره‌های مختلف در لایه پنهان.....	۷۳
جدول (۲-۱۶) - پارامترهای شبکه‌ی بهینه.....	۷۵
جدول (۲-۱۷) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری ارزیابی.....	۷۶
جدول (۲-۱۸) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری تست.....	۷۷
جدول (۲-۱۹) - نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از روش رد تک تک.....	۷۸
جدول (۲-۲۰) - مقادیر R^2 برای سری ارزیابی و سری تست بعد از چندین تست Y - تصادفی.....	۸۰
جدول (۲-۲۱) - پارامترهای آماری محاسبه شده برای مدل شبکه عصبی.....	۸۱
جدول (۲-۲۲) - جزئیات ساختار مولکولی مشتقات Nec-5.....	۸۴
جدول (۲-۲۳) - نتایج ارزیابی شبکه‌های ایجاد شده با الگوریتم آموزشی تنظیم بایزین (trainbr) و تابع تبدیل لگاریتم زیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان.....	۹۴
جدول (۲-۲۴) - توصیف گره‌های بکار رفته در مدلسازی.....	۹۶
جدول (۲-۲۵) - ماتریس همبستگی برای توصیف گره‌های انتخاب شده.....	۹۶
جدول (۲-۲۶) - نتایج ارزیابی شبکه‌هایی با ۲ نرون در لایه‌ی پنهان و تعداد دوره‌های آموزشی مختلف.....	۹۷
جدول (۲-۲۷) - نتایج حاصل از ارزیابی مدل با استفاده از سری تست.....	۹۸
جدول (۲-۲۸) - مقادیر خطای مجذور میانگین برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم زیگموئید با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان.....	۱۰۲
جدول (۲-۲۹) - مقادیر خطای مجذور میانگین برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت هیپربولیک با تعداد متغیرهای ورودی متفاوت و تعداد گم.....	۱۰۳
جدول (۲-۳۰) - توصیف گره‌های بکار رفته در مدل سازی.....	۱۰۴
جدول (۲-۳۱) - ماتریس همبستگی برای توصیف گره‌های انتخاب شده.....	۱۰۴
جدول (۲-۳۲) - مقادیر خطای مجذور میانگین برای تعداد گره‌های مختلف در لایه پنهان.....	۱۰۵
جدول (۲-۳۳) - پارامترهای شبکه‌ی بهینه.....	۱۰۶
جدول (۲-۳۴) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری ارزیابی.....	۱۰۷
جدول (۲-۳۵) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری تست.....	۱۰۸
جدول (۲-۳۶) - نتایج حاصل از ارزیابی مدل با استفاده از روش رد تک تک.....	۱۱۰
جدول (۲-۳۷) - مقادیر R^2 برای سری ارزیابی و سری تست بعد از چندین تست Y - تصادفی.....	۱۱۲
جدول (۲-۳۸) - پارامترهای آماری محاسبه شده.....	۱۱۳

فصل اول

مقدمه

۱-۱- ارتباط کمی ساختار - فعالیت

از نظر شیمی دانان فعالیت‌ها و خواص یب ناشی از ویژگی‌های ساختاری آن است. هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده‌شده‌ی مولکول انجام گیرد، به

آن ارتباط کمی ساختار - خاصیت (QSPR)^۱ می‌گویند مانند یافتن رابطه‌ای بین خواصی نظیر نقطه‌ی جوش و ساختار مولکولی. اما وقتی خواصی از نوع بیولوژیکی (مانند فعالیت دارو) توصیف می‌شود، اشاره به ارتباط کمی ساختار - فعالیت (QSAR)^۲ خواهد داشت [۱]. QSAR سعی در پیدا کردن رابطه‌ی هماهنگ میان فعالیت‌های شیمیایی و فیزیکی با ویژگی‌های مولکولی دارد، به گونه‌ای که بتوان این قواعد را برای ارزیابی فعالیت ترکیبات جدید به کار برد. در واقع نتایج این نوع مطالعات علاوه بر شفاف‌سازی نحوه‌ی ارتباط بین خواص مولکول‌ها و ویژگی‌های ساختمانی آن‌ها، به پژوهش‌گران در پیش‌بینی رفتار مولکول‌های جدید بر اساس رفتار مولکول‌های مشابه کمک می‌کند [۲]. مدل‌های QSAR را می‌توان بر اساس پارامترهایی که بصورت تجربی بدست آمده و یا از طریق تئوری محاسبه شده‌اند، به زبان ریاضی بیان نمود [۳].

کورن هانش^۳ برای اولین بار در سال ۱۹۶۴ مفهوم رابطه‌ی کمی ساختار - فعالیت یا همان QSAR را مطرح کرد. او نشان داد که توصیف‌گرهای کمی^۴ مولکولی در ترکیبات دارویی با ویژگی‌های بیولوژیک این ترکیبات ارتباط تنگاتنگی دارند. او و دانشمندان پس از وی، از توصیف‌گرهای ساده‌ای برای توصیف ساختار ترکیبات استفاده کردند که توانایی توصیف ویژگی‌های الکترونی، هندسی، فضایی و هیدروفوبیک ترکیبات دارویی را داشتند. هانش و فوجیتا^۵ از روش‌های رگرسیون خطی چندتایی^۶ در مطالعات QSAR استفاده کردند [۴]. از سال ۱۹۷۰ به بعد تعداد مقالات منتشرشده در QSAR و QSPR سریعاً افزایش یافت. در QSAR به روش نوین از توصیف‌گرهای پیچیده‌تری که ویژگی‌های کوانتومی ترکیبات را نیز مشخص می‌کنند، استفاده می‌شود.

^۱ - Quantitative structure – property relationship

^۲ - Quantitative structure – activity relationship

^۳ - Corren Hansch

^۴ - Quantitative Descriptors

^۵ - Fujita

^۶ - Multiple Linear Rrgression

در هر روند (QSAR) می‌توان شش مرحله‌ی مشخص را از هم تفکیک کرد. این شش مرحله

عبارتند از:

۱- جمع‌آوری فهرستی از ترکیبات

۲- فراهم کردن ساختار هندسی

۳- استخراج توصیف‌کننده

۴- انتخاب بهترین توصیف‌کننده‌ها برای ورود به مدل

۵- مدل‌سازی

۶- ارزیابی مدل

۱-۱-۱- جمع‌آوری لیستی از ترکیبات

اولین مرحله در توسعه‌ی مطالعات QSAR و QSPR جمع‌آوری فهرستی از ترکیبات است که کمیت‌های تجربی خاصی برای آنها توسط محققین تجربی گزارش شده‌اند. عموماً جمع‌آوری مولکول‌های سری داده‌ها از طریق جستجو در مقالات و مجلات علمی صورت می‌گیرد. از میان مجموعه مولکول‌های جمع‌آوری شده تعدادی که داده‌های گزارش‌شده در مورد آنها در شرایط عملی یکسانی بدست آمده باشند، به عنوان سری داده‌ها انتخاب می‌شوند. برای اینکه QSAR یک روش مطلوب و قابل قبول محسوب شود، بهتر است تنوع گونه‌ها زیاد باشد و ترکیبات بصورتی انتخاب شوند که از نظر ساختاری تا حدود زیادی به مولکول اصلی شباهت داشته باشند. هرچه سری داده‌ها بزرگ‌تر باشد، مدل حاصله از قدرت پیش‌بینی بالاتری برخوردار خواهد بود.

۱-۱-۲- بهینه‌سازی ساختار هندسی

فراهم کردن ساختار هندسی مولکول‌ها و بهینه کردن ساختار آن‌ها مرحله‌ی بعدی در مطالعات QSAR می‌باشد. به این منظور از شیمی محاسباتی^۱ استفاده می‌شود. شیمی محاسباتی ساختارهای مولکولی را به صورت پارامترهای عددی معرفی و رفتار آن‌ها را با معادلات کوانتومی و فیزیک کلاسیک شبیه‌سازی می‌نماید. این امر به دانشمندان امکان می‌دهد که بتوانند از این طریق به اطلاعات مولکول از جمله ساختار هندسی، انرژی‌ها و خواص مربوطه از جمله خواص الکترونیک و اسپکتروسکوپی و اثرات حلال دسترسی پیدا کنند. در سال‌های اخیر روش‌های محاسباتی در بین شیمی‌دانان بسیار رواج یافته است. در این روش‌ها به راحتی می‌توان محاسبات را انجام داد، بدون این‌که از اصول اولیه و طریقه‌ی محاسبه آگاهی و دانش لازم را داشت. شیمی محاسباتی ساختارها و واکنش‌های شیمیایی را شبیه‌سازی می‌کند و این امر به شیمی‌دانان اجازه می‌دهد که به مطالعه‌ی روش‌های شیمیایی با اجرای محاسبات توسط کامپیوتر بپردازند بدون این‌که از روش‌های تجربی و آزمایشگاهی استفاده کنند [۱ و ۵]. روش‌های محاسباتی در شیمی به چند دسته تقسیم می‌شوند:

۱- روش‌های آغازین^۲

۲- روش‌های نیمه تجربی^۳

۳- روش‌های تابعی چگالی^۴

۴- روش‌های مکانیک مولکولی^۵

^۱ - Computational Chemistry

^۲ - Ab initio methods

^۳ - Semi-empirical methods

^۴ - Density functional theory (DFT)

^۵ - Molecular mechanics

۱-۲-۱-۱- روش‌های آغازین

در روش‌های آغازین از هیچ داده‌ی تجربی استفاده نمی‌شود فقط ثابت‌های فیزیکی مانند جرم الکترون، بار الکترون و ثابت پلانک استفاده می‌شود به همین دلیل به آن روش آغازین گفته می‌شود (چون از هیچ داده‌ی تجربی استفاده نمی‌شود). تمام پارامترهای ساختاری از حل دقیق معادلات شرودینگر به دست می‌آیند. اگر تابع موج ψ را برای یک سیستم داشته باشیم، داریم:

$$\hat{H}\psi = E\psi$$

که $\hat{H}$ معرف عمل گر هامیلتونی، ψ بیانگر تابع موج و E انرژی سیستم مورد نظر می‌باشد که شامل انرژی جنبشی و انرژی پتانسیل است. مشکل مهم و اساسی در این معادله بدست آوردن ψ است. برای سیستم ساده مانند هیدروژن ψ به طور دقیق بدست می‌آید. وقتی یک الکترون و یک هسته داشته باشیم در تابع پتانسیل $1/r$ ظاهر می‌شود. بنابراین انرژی پتانسیل به فاصله‌ی الکترون از هسته و بار آن‌ها بستگی دارد. اما اگر تعداد الکترون‌ها زیاد شود در تابع پتانسیل دافعه‌ی الکترون – الکترون نیز به صورت $1/r_{1,2}$ ظاهر می‌شود که حل معادلات شرودینگر را مشکل می‌سازد. برای حل این مشکل از روش هارتری – فاک^۱ استفاده می‌شود، در اینجا برهمکنش دو الکترون در نظر گرفته نمی‌شود، بلکه از عملگر فاک استفاده می‌شود. یعنی الکترون دوم به صورت ابر الکترونی در نظر گرفته می‌شود و برهمکنش الکترون اول با ابری از الکترون دوم در نظر گرفته می‌شود، اما مشکلی که در این روش وجود دارد این است که برهمکنش‌های لحظه‌ای بین الکترون اول و الکترون دوم در نظر گرفته نمی‌شود و همچنین برای حل دقیق معادلات شرودینگر ناگزیر از حل تعداد زیادی انتگرال‌های پیچیده و دشوار هستیم، که این کار علاوه بر پیچیدگی وقت زیادی نیز می‌گیرد.

^۱ - Hartree-fock

در روش هاتری - فاک باید نوع هارتتری - فاک به کار برده شده را مشخص کرد یعنی باید مشخص شود که هارتتری - فاک محدود شده^۱ یا هارتتری - فاک نامحدود^۲ 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 هارتتری - فاک محدود شده گفته می‌شود و یا این که برای هر کدام از الکترون‌ها یک تابع موج خاص تعریف کرد، که به این حالت هارتتری - فاک نامحدود گفته می‌شود. در هارتتری - فاک محدود شده حجم محاسبات به نصف کاهش می‌یابد، ولی مزیت هارتتری - فاک نامحدود این است که اگر اختلافی بین تابع موج دو الکترون وجود داشته باشد، را می‌توان مشاهده نمود. معمولاً اگر لایه‌ی آخر پر باشد تفاوت چندانی بین دو الکترون وجود ندارد، لذا از هارتتری - فاک محدود شده استفاده می‌شود [۱].

۱-۲-۲- روش‌های نیمه تجربی

روش‌های نیمه تجربی مانند روش‌هایی که در برنامه‌هایی نظیر MOPAC و Hyperchem وارد شده‌اند و محاسبات در این روش‌ها بر پایه‌ی مکانیک کوانتومی است. در این روش‌ها یک سری از مقادیر تجربی نیز وارد شده‌اند و فقط الکترون‌های لایه‌ی ظرفیت در محاسبه در نظر گرفته می‌شوند. روش‌های نیمه تجربی مختلفی مانند صرف نظر کردن کامل از انتگرال‌های همپوشانی جزئی^۳ (CNDO)، صرف نظر از برخی انتگرال‌های همپوشانی در الکترون‌ها^۴ (INDO)، صرف نظر دو اتمی از انتگرال همپوشانی^۵ (MNDO) ارائه شده است [۶]. روش AM1^۶ همان روش MNDO اصلاح شده می‌باشد که روی مولکول‌های قطبی نتایج خوبی می‌دهد ولی پیوند هیدروژنی را نمی‌تواند در نظر

^۱ - Restricted hartree-fock (RHF)

^۲ - Unrestricted hartree-fock (UHF)

^۳ - Complet neglected of differential overlap

^۴ - Intermediate neglected of differential overlap

^۵ - Modified neglected of diatomic overlap

^۶ - Austin methods

بگیرد. در این روش هسته و لایه‌های داخلی را به شکل یک هسته‌ی مرکزی^۱ در نظر می‌گیریم و محاسبات روی الکترون‌های ظرفیت انجام می‌گردد. سایر پارامترهای مورد نیاز در این روش از اطلاعات تجربی مانند اسپکتروسکوپی استخراج می‌گردند [۱ و ۵].

۱-۲-۳- روش‌های تابعی چگالی

در این روش‌ها به جای استفاده از تابع موج، از دانسیته الکترونی استفاده می‌شود و مزیتی که دارند 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].

۱-۲-۴- 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. نتایج این روش دقیق نیستند و لذا برای مولکول‌هایی که تعداد اتم‌های آن‌ها زیاد نیست، به کار می‌رود [۱ و ۵].

۱-۳-۱- استخراج توصیف‌گرها

پس از بهینه‌سازی ساختار مولکول‌ها بایستی توصیف‌کننده‌های وابسته به ساختار مولکول استخراج شوند. توصیف‌کننده‌ها مقادیری عددی هستند که ویژگی‌های مختلف ساختاری و الکترونی مولکول‌ها را به طور کمی نشان می‌دهند. توصیف‌گرهایی که در مطالعات QSAR مورد استفاده قرار می‌گیرند شامل دو دسته توصیف‌گرهای تئوری و توصیف‌گرهای تجربی می‌باشند. توصیف‌گرهای

^۱ - Core

تجربی از طریق آزمایش اندازه‌گیری می‌شوند و برای دسته ترکیبات محدودی در دسترس هستند. در حالی که توصیف‌گرهای تئوری بدون نیاز به داده‌های تجربی به کمک نرم‌افزارهای مناسب برای طیف وسیعی از ترکیبات قابل محاسبه هستند. بنابراین توصیف‌گرهای تئوری در مطالعات QSAR از اهمیت زیادی برخوردار هستند [۱].

۱-۴- انتخاب بهترین توصیف‌کننده‌ها

یکی از مهمترین مراحل QSAR تعیین و انتخاب توصیف‌گرهای مناسب می‌باشد. زیرا توصیف‌کننده‌های زائد کار برآزش و مدل‌سازی را طولانی می‌کنند و تأثیری در بهبود نتایج نخواهند داشت. به این منظور لازم است در انتخاب توصیف‌کننده‌ها دقت لازم به عمل آید و توصیف‌گرهایی انتخاب شوند که با پارامتر مورد نظر ارتباط نزدیکی داشته و بتوانند تغییرات آن را به خوبی توجیه نمایند. یکی از روش‌های انتخاب توصیف‌گرهای مناسب روش برآزش مرحله‌ای^۱ می‌باشد که با محاسبه‌ی ضرایب همبستگی^۲ مشخص می‌شوند. ضریب همبستگی میزان نزدیکی اطلاعات بدست آمده به خط رگرسیون را نشان می‌دهد. مقدار ضریب همبستگی بین ۰ و ۱ تغییر می‌کند. صفر یعنی هیچ ارتباطی بین فعالیت و توصیف‌کننده وجود ندارد و یک یعنی ارتباط کامل است. پس توصیف‌گرهایی با ضرایب همبستگی بزرگ در برآزش منحنی برای ساختن مدل جهت پیش‌گویی خواص استفاده می‌شوند [۱].

^۱ - stepwise

^۲ - Interrelation coefficient

۱-۱-۵- مدل سازی

پس از ارزیابی توصیف کننده ها و انتخاب مناسب ترین آن ها، هدف ساختن مدلی است که قادر باشد با استفاده از توصیف گرهای مولکولی خاصیت یا فعالیت مورد نظر را پیش گویی نماید. روش های متفاوتی در مدل سازی به کار گرفته می شود که می توان آن ها را در دو دسته ی کلی زیر قرارداد:

الف - استفاده از روش های آماری چند متغیره: در این روش از روش های کالیبراسیون چند متغیره نظیر رگرسیون خطی چندگانه و روش حداقل مربعات جزئی^۱ استفاده می شود. این روش ها خطی می باشند و وقتی کارایی مناسبی دارند که داده ها دارای رفتار خطی باشند. یک رگرسیون با ایجاد معادله خطی انجام می شود.

$$\text{Property (activity)} = \text{constant} + (c_1 \cdot p_1) + (c_2 \cdot p_2) + \dots + (c_n \cdot p_n)$$

که پارامترهای p_1 تا p_n توصیف کننده هایی هستند که به روش محاسباتی برای هر مولکول به دست می آیند و ضرایب c_1 تا c_n از طریق برازش توصیف کننده ها با فعالیت در مجموعه ای از مولکول ها که دارای خواص مشترک هستند، محاسبه می شوند.

ب) استفاده از روش های غیرخطی: جهت کم کردن انحراف بین مقادیر برازش شده و مقادیر صحیح بهتر است از یک برازش غیرخطی در مدل سازی استفاده شود. از میان روش های غیرخطی روش شبکه ی عصبی مصنوعی بطور گسترده در مطالعات QSAR و QSPR به کار گرفته شده است [۱].

در دو بخش بعد به بررسی مختصر مدل های استفاده شده برای بررسی رابطه ی ساختار مولکولی و فعالیت بازدارندگی ترکیبات موجود می پردازیم.

^۱ - Partial Least Squares

۱-۶- رگرسیون خطی چندگانه (MLR)

رگرسیون خطی چندگانه راهی برای یافتن یک رابطه‌ی خطی میان چند متغیر مستقل و یک متغیر وابسته است. برای درک بهتر، ابتدا نگاهی به مفهوم رگرسیون و ساده‌ترین نوع آن یعنی مدل رگرسیون خطی ساده می‌اندازیم.

برای اولین بار، فرنسیس گالتون^۱ (۱۹۱۱) ایده‌ی رگرسیون را مطرح و از آن در حل یک مسأله‌ی وراثتی استفاده کرد. از رگرسیون برای پیش‌گویی مقدار یک متغیر مجهول – که از آن به متغیر وابسته تعبیر می‌شود – در رابطه با یک متغیر معلوم – که به عنوان متغیر مستقل شناخته می‌شود – استفاده می‌شود.

متغیر مستقل متغیری است که مقدار آن توسط محقق مشخص می‌شود و طبیعت آن طوری است که می‌توان آن را به طور مستقل تغییر داد. از طرفی مقدار متغیر وابسته با آهنگ متغیر مستقل تغییر می‌کند؛ به طوری که به ازای هر مقدار از متغیر مستقل، یک یا چند مقدار برای متغیر وابسته وجود دارد. یک مدل رگرسیون خطی بر اساس فرض‌های زیر استوار است:

۱- مقادیر متغیر مستقل از قبل توسط محقق معین شده است و در طول رگرسیون تغییر نخواهد کرد.

۲- خطای اندازه‌گیری متغیر مستقل ناچیز و قابل چشم‌پوشی است.

۳- به ازای هر مقدار متغیر مستقل زیرمجموعه‌ای از مقادیر متغیر وابسته با پراکندگی نرمال وجود دارد.

۴- واریانس در زیرمجموعه‌های متغیر وابسته یکسان است.

۵- میانگین‌های زیرمجموعه‌های متغیر وابسته روی یک خط واقعند. این فرض که فرض خطی بودن نیز نام دارد به شکل زیر بیان می‌شود:

^۱ - Francis Galton

$$\mu_{y|x} = \alpha + \beta x \quad \text{رابطه‌ی (۱-۱)}$$

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

۶- مقادیر متغیر وابسته از نظر آماری مستقل‌اند.

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

$$y = \alpha + \beta x + e \quad \text{رابطه‌ی (۲-۱)}$$

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

روشی که غالباً برای بدست آوردن معادله‌ی خط بین متغیرهای مستقل و وابسته به کار می‌رود، به روش کمترین مربعات^۱ موسوم است. اصولاً برای رسم یک خط دانستن مقادیر α و β در معادله‌ی مربوط به آن کافی است. این مقادیر را می‌توان از دستگاه معادلات زیر – که به معادلات نرمال معروف است – بدست آورد:

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

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

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

^۱ - Least Squares Method

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

خطی که با این روش به دست می‌آید دارای بهترین معادله برای مرتبط کردن متغیرهای مستقل و وابسته است؛ هر چند همه‌ی مشاهدات تجربی بر روی این خط واقع نمی‌شوند. ویژگی چنین خطی این است که مجموع مربعات انحراف عمودی نقاط مشاهده شده (y_i) از این خط، همواره کمتر از مجموع مربعات انحراف عمودی نقاط مشاهده شده (y_i) از هر خط مفروض دیگر است. پس از این مقدمه، به بررسی رگرسیون خطی با چند متغیر مستقل می‌پردازیم که مدل مورد استفاده‌ی ما در این تحقیق است. در این نوع از رگرسیون فرض‌هایی مشابه آن چه در حالت قبل دیدیم، در نظر گرفته می‌شود. معادله‌ی این نوع رگرسیون به این صورت است:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + e_i \quad \text{رابطه‌ی (۶-۱)}$$

که y_i یکی از مقادیر زیرمجموعه‌ی متغیر وابسته است. مقادیر β ضرایب متغیرهای مستقل اند. e_i هم یک متغیر تصادفی است که میانگین آن صفر و واریانس آن برابر واریانس زیرمجموعه‌های متغیر وابسته است. مقدار این متغیر بیانگر میزان خطای رابطه‌ی رگرسیون است. در حالتی که دو متغیر مستقل داشته باشیم، رابطه‌ای که ذکر شد معادله‌ی صفحه‌ای است که نقاط روی آن مجموعه‌ی جواب رگرسیون خطی چندگانه است.

برای به دست آوردن معادله‌ی رگرسیون خطی چندگانه لازم است مجموع مربعات انحرافات مقادیر پیش‌گویی شده برای متغیر وابسته از مقادیر واقعی آن به کمترین مقدار ممکن برسد. به عبارت دیگر باید مقدار $\sum e_i^2$ در این معادله به کمترین حد ممکن کاهش یابد. مقدار $\sum e_i^2$ طبق رابطه‌ی زیر تعریف می‌شود:

$$\sum e_j^2 = \sum (y_i - \hat{y})^2 \quad \text{رابطه‌ی (۷-۱)}$$

که در آن y_i مقدار مشاهده شده‌ی متغیر وابسته و $\hat{y}$ مقدار محاسبه شده‌ی آن به وسیله‌ی معادله‌ی رگرسیون است. در این جا نیز همانند رگرسیون خطی ساده برای محاسبه ضرایب، از روش کمترین مجموع مربعات استفاده می‌شود.

با حل معادلات نرمال می‌توان به رابطه‌ی رگرسیون خطی چندگانه رسید. این معادلات عبارتند از:

$$\begin{cases} \sum y_i = nb_0 + b_1 \sum x_{1i} + b_2 \sum x_{2i} + \dots + b_k \sum x_{ki} \\ \sum x_{1i} y_i = b_0 \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} + \dots + b_k \sum x_{1i} x_{ki} \\ \sum x_{2i} y_i = b_0 \sum x_{2i} + b_1 \sum x_{2i} x_{1i} + b_2 \sum x_{2i}^2 + \dots + b_k \sum x_{2i} x_{ki} \\ \dots \\ \sum x_{ki} y_i = b_0 \sum x_{ki} + b_1 \sum x_{ki} x_{1i} + b_2 \sum x_{ki} x_{2i} + \dots + b_k \sum x_{ki}^2 \end{cases} \quad \text{رابطه‌ی (۸-۱)}$$

مساوی بودن تعداد معادلات و مجهولات به ما امکان به دست آوردن مقادیر صحیح ضرایب را می‌دهد. مجموع انحراف موجود در پیش‌گویی‌های مربوط به متغیر وابسته را می‌توان دو قسمت کرد. قسمت اول شامل مقداری از مجموع انحرافات است که توسط رگرسیون توضیح داده شده است. از این قسم به انحرافات توضیح داده شده یاد می‌شود. قسمت دوم شامل مقداری از مجموع انحرافات است که توسط رگرسیون توضیح داده نشده است. به زبان ریاضی، انحراف توضیح داده شده را می‌توان به صورت $\sum (\hat{y} - \bar{y})^2$ بیان کرد. این عبارت بیان‌گر مجموع مربعات انحراف مقادیر پیش‌گویی شده‌ی متغیر وابسته از میانگین مقادیر آن است. این جزء از انحرافات را مجموع مربعات انحرافات ناشی از رگرسیون هم می‌خوانند؛ که با SSR نشان داده می‌شود. از طرفی، انحراف توضیح داده نشده را با $\sum (\hat{y}_i - \hat{y})^2$ بیان می‌کنند که نشان‌گر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌گویی شده برای آن است. این کمیت را به نام مجموع مربعات خطا نیز می‌خوانند و به SSE نمایش می‌دهند. هم چنین حاصل جمع این دو جزء، مجموع مربعات یا SST نامیده می‌شود:

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y_i - \hat{y})^2 \quad \text{رابطه‌ی (۹-۱)}$$

یا

$$SST = SSR + SSE$$

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

$$r^2 = \frac{SSR}{SST} \quad \text{رابطه‌ی (۱۰-۱)}$$

از این رو هر چه مقدار این شاخص بزرگتر باشد، مدل صحیح‌تر خواهد بود [۷]. ما برای ارزیابی مدل‌های خود از شاخص دیگری بنام MSE بهره جستیم که در قسمت (۱-۱-۸) به معرفی آن خواهیم پرداخت.

۱-۱-۷- شبکه عصبی مصنوعی^۱

نحوه یادگیری و ایجاد ارتباط میان مفاهیم و اشیاء یکی از رموز پیچیده‌ی دستگاه آفرینش است. مغز به عنوان یک سیستم پردازش اطلاعات با ساختار کاملاً پیچیده حدود دو درصد وزن بدن را تشکیل می‌دهد و کلیه اعمال آگاهانه و بسیاری از رفتارهای ناآگاهانه را کنترل می‌کند. این دستگاه که از تجمع و آرایش ویژه‌ی سلول‌های تخصص یافته‌ای به نام نرون^۲ تشکیل شده است، با تأثیرپذیری و تأثیرگذاری متقابل بر دستگاه‌های دیگر بدن، ارتباط هدفمند جاندار را با محیط اطرافش ممکن می‌سازد. دریافت پیام‌های بیرونی، تحلیل و پاسخ‌دهی به آن‌ها منوط به چینش صحیح و عملکرد سریع نورونهاست [۸].

پیاده‌سازی ویژگی‌های شگفت‌انگیز مغز انسان در یک سیستم مصنوعی از دیرباز مورد توجه بوده است، در سال ۱۹۴۳، مک کلوج^۳ یک نوروفیزیولوژیست و یک ریاضیدان به نام پیتز^۴ برای نخستین بار یک مدل ریاضی برای شبیه‌سازی عملکرد یک سلول عصبی پیشنهاد کردند [۹]. کار این افراد را می‌توان نقطه شروع حوزه‌ی علمی شبکه عصبی مصنوعی نامید. در دهه‌های ۵۰ و ۶۰ میلادی، یک گروه از پژوهشگران اولین شبکه‌های عصبی مصنوعی را ساختند. ابتدا آنان مدل‌های خود را به شکل سخت‌افزارهای الکتریکی و سپس به صورت نرم‌افزارهای انعطاف‌پذیر رایانه‌ای طراحی کردند. شبکه‌های عصبی مصنوعی ابتدایی از یک لایه از نرون‌ها به نام پرسپترون^۵ تشکیل شده بودند. هر چند این شبکه‌های ابتدایی در حل برخی مسائل مفید بودند، بسیاری از مسائل به وسیله‌ی این مدل‌های ساده حل نشد. ماروین مینسکی^۶ و سیمور پاپرت^۷ [۱۰] در سال ۱۹۶۹ در کتابی به نام پرسپترون‌ها^۸ اثبات کردند که شبکه‌های عصبی تک لایه که در آن روزگار معمول بودند، قابلیت حل بسیاری از

^۱ - Artificial Neural Network

^۲ - Neuron

^۳ - McCulloch

^۴ - Pitts

^۵ - Perceptron

^۶ - Marvin Minsky

^۷ - Symore Papert

^۸ - Perceptrons

مسائل ساده را هم ندارند. عدم موفقیت شبکه‌های عصبی مصنوعی در حل این مسائل و به دنبال آن نگرش این کتاب تأثیر عمیقی در جامعه محققان گذاشت و موجب بروز یک دوره‌ی خمودی در تاریخ شبکه‌های عصبی مصنوعی گردید.

علی رغم ناامیدی عمومی ایجاد شده درباره‌ی شبکه‌های عصبی، برخی از دانشمندان همچون توو کوهونن^۱، استفن گراسبرگ^۲ و جیمز اندرسون^۳ همچنان به تلاش‌های خود ادامه دادند. در دهه‌های ۷۰ و ۸۰ این دانشمندان با ایجاد شبکه‌های چند لایه موفق به حل بسیاری از مسائل شدند. هم‌اکنون شبکه‌های عصبی مصنوعی به طور گسترده در شاخه‌های مختلف علمی و تحقیقاتی مورد استفاده است. توانایی این ابزار برای یافتن روابط غیرخطی پیچیده میان مفاهیم مختلف، آن را به روشی ارزشمند در دنیای علم بدل کرده است. برخی از توانایی‌های شبکه‌های عصبی مصنوعی را می‌توان در موارد زیر خلاصه کرد [۸]:

۱- یافتن و آموزش پذیری الگوهای خطی و غیرخطی

۲- پیش‌گویی و کلاسه‌بندی

۳- بهینه‌سازی

۱-۷-۱-۱- اساس بیولوژیکی شبکه‌های عصبی مصنوعی

قبل از پرداختن به شبکه‌های عصبی مصنوعی ضروری است نگاهی اجمالی به الگوی طبیعی آن بیاندازیم. مغز به عنوان یک سیستم پردازش اطلاعاتی با ساختار موازی از ۱۰۰ میلیارد (۱۰^{۱۱}) نرون تشکیل شده است. شکل (۱-۱) نمای کلی یک نرون را نشان می‌دهد. هر نرون دارای سه قسمت دندریت^۴، جسم سلولی^۵ و اکسون^۱ است. پیام عصبی به صورت یک سیگنال الکتریکی به طور معمول

^۱ - Teo Kohonen

^۲ - Stefan Grossberg

^۳ - James Anderson

^۴ - Dendrite

^۵ - Cell Body, Soma

از دندريت به جسم سلولى و اكسون منتقل مى‌شود. بين اكسون يك نرون و دندريت نرون بعدى يك فضاى خالى وجود دارد كه سيناپس^۲ ناميده مى‌شود. سيناپس‌ها با كنترل انتقال پيام عصبى از يك نرون به نرون ديگر نقش مهمى در فعاليت دستگاه عصبى دارند. انتقال‌دهنده‌هاى عصبى^۳ مواد شيميايى هستند كه عمل انتقال پيام را در سيناپس‌ها انجام مى‌دهند. اكسون يك نرون با آزاد كردن مقادير مشخصى از يك انتقال‌دهنده عصبى در فضاى سيناپسى، باعث تحريك دندريت نرون مجاور شده، پيام عصبى را به آن منتقل مى‌كند [۱۱].

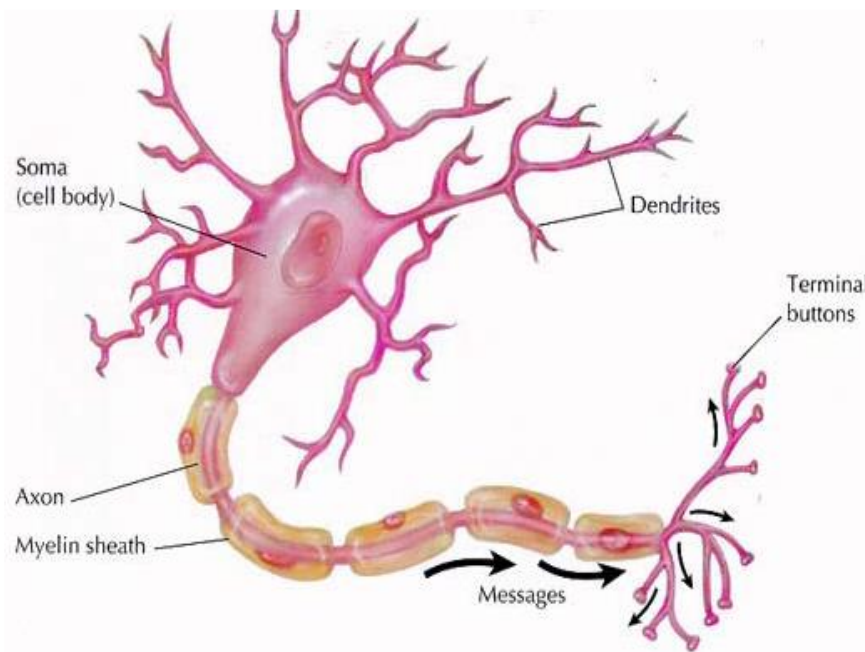
غشاي سلولى نقش اصلى را در ايجاد و انتقال پيام عصبى در طول يك نرون به عهده دارد. در حالت عادى يا استراحت، غشاي نرون‌ها دارى پتانسيل الكترىكى حدود -65mv مى‌باشند. وجود و دوام اين مقدار از پتانسيل الكترىكى مربوط به اختلاف غلظت يون‌ها در دو طرف غشاست. غلظت يون سدیم و كلر در طرف خارجى غشاء بسيار بيشتر از طرف داخلى آن است. از طرف ديگر غلظت يون پتاسيم در طرف داخلى غشاء بيشتر است. پمپ سدیم - پتاسيم [۱۲] با مصرف انرژى مسئول نگه‌داشتن اين اختلاف غلظت است. تحريك غشاي يك نرون بوسيله انتقال‌دهنده‌هاى عصبى باعث نفوذپذيرى غشاي آن نسبت به يون سدیم مى‌شود. تحريك نرون از ناحيه‌ى يك سيناپس موجب باز شدن كانال‌هاى سدیمی و شارش يون‌هاى سدیم به درون نرون و افزايش پتانسيل غشاء به حدود -45mv مى‌شود. تحريك سيناپس‌هاى متعدد در يك نرون اثر تجمعى روى آن داشته و پتانسيل غشاء را تا حدود $+20\text{mv}$ افزايش مى‌دهد. اين حد از پتانسيل را پتانسيل عمل^۴ گویند كه موجب تحريك كامل نرون و انتقال پيام عصبى مى‌شود. پس از اين تحريك و انتقال پيام، پتانسيل غشاء دوباره به حالت استراحت باز مى‌گردد. چون بيان جزئيات فزيولوژى نرون‌ها از هدف اين نوشتار خارج است، به بيان همين كليات ضرورى اكتفا مى‌كنيم.

¹ - Axon

² - Synapse

³ - Neurotransmitters

⁴ - Action Potential



شکل (۱-۱) ساختار نرون طبیعی

۱-۷-۲- ساختار و عملکرد شبکه‌های عصبی مصنوعی

هر نرون مصنوعی نیز مانند یک نرون بیولوژیکی از سه جزء لایه‌ی ورودی، لایه‌ی میانی و لایه‌ی خارجی تشکیل شده است. لایه‌ی ورودی سیگنال‌ها را از محیط خارج دریافت می‌کند. لایه میانی عملیات محاسباتی را روی سیگنال‌های ورودی انجام می‌دهد. لایه‌ی خروجی سیگنال‌ها را به خارج یا نرون‌های دیگر انتقال می‌دهد. عمل لایه‌ی ورودی، لایه‌ی میانی و لایه‌ی خارجی در یک شبکه عصبی شبیه عمل دندریت، جسم سلولی و اکسون در یک نرون بیولوژیکی می‌باشد. ساختار و عملکرد یک نرون مصنوعی ساده در شکل (۲-۱) نمایش داده شده است [۱]. در این نرون، چند ورودی کمی در ضرایب مخصوص که وزن^۱ (w) نام دارد ضرب می‌شوند و وارد نرون می‌گردند.

به طور کلی در شبکه‌های عصبی مصنوعی، هر ورودی به همراه وزن مربوطه به آن معادل یک سیناپس در الگوی طبیعی نرون‌هاست. خواهیم دید که این اوزان نشان دهنده‌ی اهمیت ورودی‌هاست. در بدنه‌ی یک نرون، ورودی‌ها در اوزان مربوط به خود ضرب شده و سپس مشابه آن‌چه

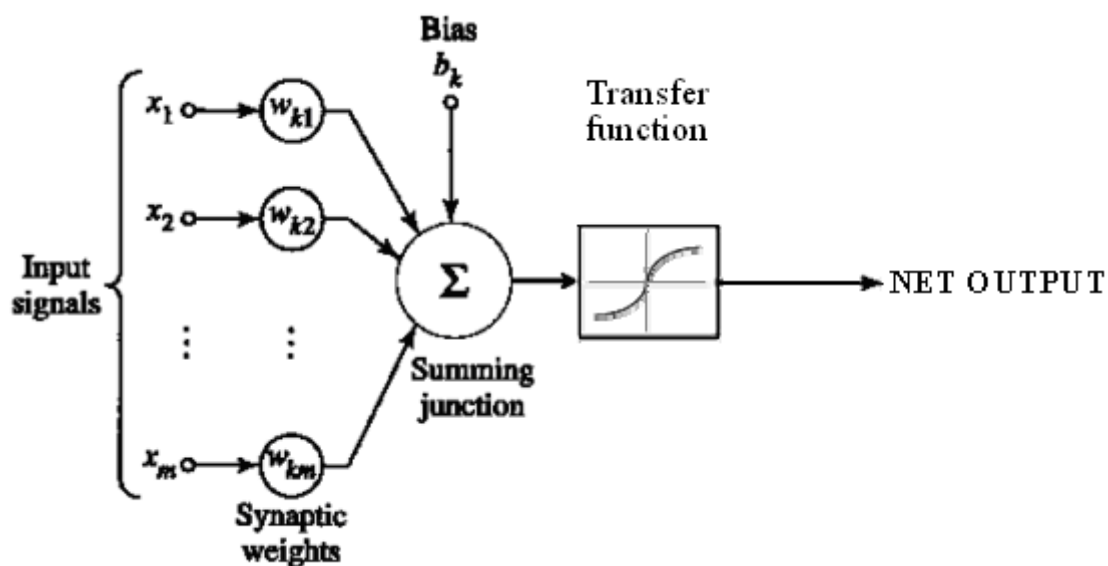
^۱ - Weight

در نرون‌های طبیعی دیدیم، با هم جمع می‌شوند. این حاصل جمع به عنوان ورودی تابع می‌باشد که معادل پتانسیل عمل در یک نرون طبیعی است. نتیجه این تابع که تابع محرک یا انتقال^۱ نام دارد، خروجی نرون است. اگر m ورودی داشته باشیم، OUT خروجی نرون خواهد بود:

$$NET = \sum_{i=1}^m x_i w_{ki} \quad \text{رابطه‌ی (۱-۱)}$$

$$OUT = F(NET + b_k) \quad \text{رابطه‌ی (۱-۲)}$$

که $x_1, x_2, \dots, x_n$ داده‌های ورودی، $w_1, w_2, \dots, w_n$ اوزان اتصال ورودی‌های ۱، ۲ و ... به نرون k ، NET ترکیب خطی ورودی‌ها، b_k بایاس^۲، F تابع تحریک و OUT خروجی نرون است. نقش b_k افزایش یا کاهش مجموع وزن دار است.



شکل (۱-۲) مدل نرون مصنوعی

ورودی متناظر وزن بایاس در واقع برای انتقال مقدار خروجی به میزان مناسب در مجموعه‌ی فضای خروجی می‌باشد. یادگیری برای نرون فوق بدین معنی است که پارامترهای W و b طوری تغییر می‌کنند که رابطه‌ی ورودی - خروجی نرون با هدف خاص مطابقت نماید.

^۱ - Transfer Function

^۲ - Bias

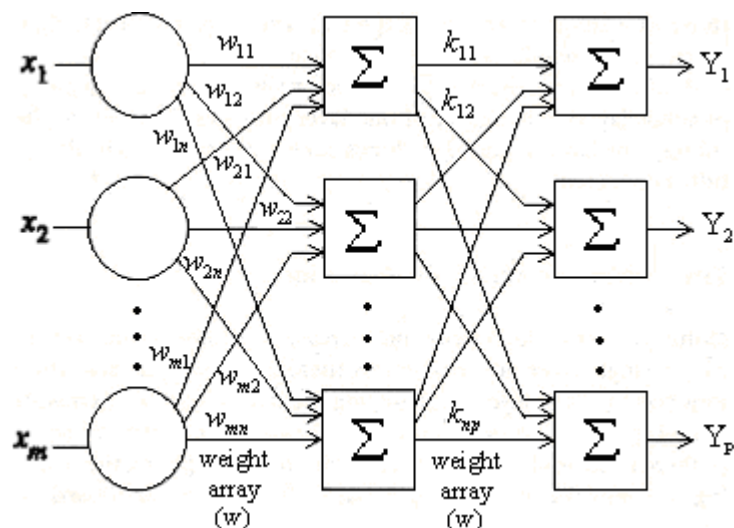
با وجود تنوع نمونه‌های شبکه، تقریباً همه‌ی آن‌ها روی این ساختار بنا نهاده شده‌اند. در این جا ورودی‌ها که با $x_1, x_2, \dots, x_n$ نشان داده شده‌اند و مجموعاً با بردار X نمایش داده می‌شود، متناظر با سیگنال‌ها در نرون بیولوژیکی می‌باشد. هر سیگنال به وزن وابسته‌ی $w_1, w_2, \dots, w_n$ ضرب شده، بعد در بلوک Σ به کار برده می‌شود. هر وزن که با بردار W نمایش داده شده است متناظر با مقاومت اتصال سیناپتیکی بیولوژیکی است. بلوک جمع‌بندی Σ ورودی‌های وزین شده را به صورت جبری جمع می‌کند که با NET نمایش داده شده است.

اگرچه یک نرون تنها توابع محرک با طرح ساده‌ی معینی را نشان می‌دهد ولی توان اصلی محاسبات عصبی بواسطه‌ی اتصالات نرون‌ها در شبکه است. شبکه، گروهی از نرون‌های تنظیم شده در یک لایه است. پرسپترون [۱۳ و ۱۴] - که ساده ترین نوع شبکه عصبی است - از یک لایه از نرون‌ها تشکیل شده است. تابع محرک این نرون‌ها معمولاً تابع خطی است. در شبکه‌های پیشرفته‌تر، تابع فعالیت از نوع لگاریتم زیگموئید و یا تانژانت زیگموئید است.

در شبکه‌های چندلایه، یک خروجی از هر نرون وارد هر کدام از نرون‌های لایه‌ی بعدی می‌شود. در اکثر موارد، هر کدام از ورودی‌های مسأله نیز وارد یک نرون از لایه‌ی اول می‌شود. شکل (۳-۱) یک شبکه‌ی چندلایه را نشان می‌دهد. نرون‌های لایه‌ی اول فاقد تابع فعالیت هستند و هر یک مسؤول وارد کردن یک توصیف‌گر به شبکه‌اند. لایه‌های میانی را لایه‌های مخفی^۱ نیز می‌نامند. شبکه‌های عصبی چندلایه کاملاً به صورت جعبه‌ی سیاه^۲ عمل می‌کنند یعنی پارامترهای داخلی شبکه فقط برای خود شبکه مفید است و هیچ گونه اطلاعاتی نمی‌توان از آن‌ها به دست آورد [۱۵].

^۱ - Hidden layers

^۲ - Black box

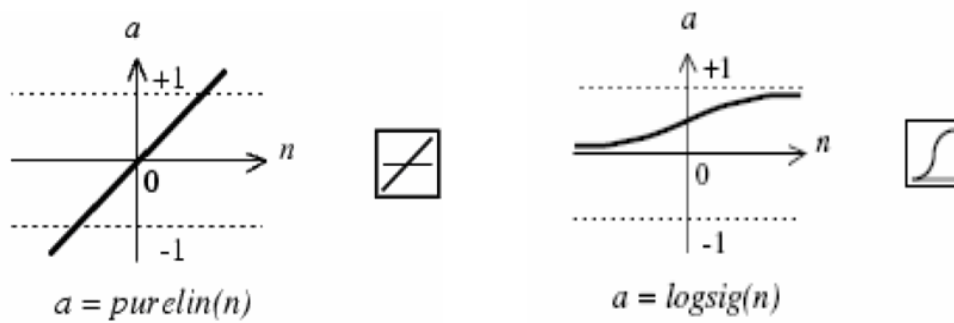


شکل (۳-۱) - شبکه های عصبی مصنوعی چندلایه^۱.

۳-۷-۱-۱- توابع محرک

تابع محرک بازه‌ی NET (رابطه‌ی (۱-۱۱)) را محدود نموده و مقدار OUT را در یک محدوده‌ی معینی قرار می‌دهد. هر تابع مشتق‌پذیر و پیوسته که بطور یکنواخت افزایش می‌یابد، می‌تواند به عنوان تابع محرک استفاده شود. معمول‌ترین این توابع، توابع زیگموئید و خطی می‌باشند که در شکل (۴-۱) نشان داده شده‌اند [۱۵]. 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 [۱]. جدول (۱-۱) معمول‌ترین توابع انتقال را نشان می‌دهد.

^۱ - Multilayer Artificial Neural Networks



تابع محرک خطی

تابع زیگموئید

شکل (۱-۴) - دو نمونه از توابع فعال ساز [۱۵]

جدول (۱-۱) معمول ترین توابع فعال ساز [۱۵]

نام تابع	تعریف
Identity	x
Logistic	$\frac{1}{1 + e^{-x}}$
Hyperbolic Tangent Sigmoid	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$
Exponential	e^{-x}
Softmax	$\frac{e^x}{\sum_i e^{x_i}}$
Unit sum	$\frac{x}{\sum_i x_i}$
Square root	$\sqrt{x}$
Sine	$\text{Sin}(x)$

۱-۱-۷-۴- آموزش^۱ شبکه‌های عصبی مصنوعی

بارزترین ویژگی شبکه‌های عصبی مصنوعی آموزش‌پذیری آن‌هاست. آموزش فرآیندی است که منجر به دستیابی به وزن‌های بهینه برای ورودی هر نرون شبکه عصبی می‌گردد. هدف آموزش آن است که خروجی شبکه را در نزدیکترین حد ممکن به خروجی مطلوب قرار دهد [۱۶]. الگوریتم یادگیری شبکه یک الگوی ریاضی است که چگونگی و سرعت یک شبکه عصبی مصنوعی در رسیدن به پارامترهای مناسب را نشان می‌دهد. در صورتیکه سری آموزش به خوبی انتخاب شود و الگوریتم آموزشی مؤثر باشد، آنگاه شبکه قادر خواهد بود به نحو صحیحی از داده‌های ورودی، خروجی قابل قبولی ارائه دهد. الگوریتم‌های مختلفی برای آموزش شبکه‌های عصبی مصنوعی بوجود آمده است، که مشهورترین آنها الگوریتم پس انتشار خطاست^۲ که در ادامه توضیح آن خواهد آمد.

۱-۱-۷-۵- الگوریتم آموزشی پس انتشار خطا

پس از طراحی اولیه، شبکه عصبی بایستی آموزش ببیند تا این که قادر به پیش‌بینی صحیحی باشد. یکی از الگوهای آموزشی که استفاده زیادی در آموزش شبکه‌های عصبی دارد، روش پس‌انتشار می باشد. این روش اساساً از الگوریتم کاهش گرادیان استفاده می‌کند [۱].

در سال ۱۹۸۶، روملهارت^۳، هینتون^۴ و ویلیامز^۵ [۱۷] تعریف روشنی از الگوریتم پس‌انتشار خطا ارائه دادند. اما قبل آن در سال ۱۹۸۲، پارکر^۶ [۱۸] شبیه کار آن‌ها را گزارش کرده بود. کمی بعد هم معلوم شد که وربوس^۷ [۱۹] در سال ۱۹۷۴ به همان نتایج رسیده بوده است! در واقع این افراد در حالی که از کارهای یکدیگر اطلاعی نداشتند، تحقیقات موازی انجام داده بودند.

^۱ - Training
^۲ - Back propagation
^۳ - Rumelhart
^۴ - Hinton
^۵ - Williams
^۶ - Parker
^۷ - Werbos

روش پس انتشار خطا از قانون کلی دلتا^۱ به عنوان یک قانون یادگیری قدرتمند استفاده می‌کند. یادگیری از طریق این روش دو مرحله دارد: مرحله‌ی پیشروی و مرحله‌ی بازگشت. در مرحله‌ی پیشروی ورودی‌ها به صورت لایه به لایه در شبکه پیش می‌رود و در پایان یک سری خروجی به عنوان جواب حقیقی شبکه به دست می‌آید. در این مرحله اوزان اتصال ثابت است. در مرحله‌ی بازگشت، اوزان اتصال بر اساس قانون دلتا، تغییر می‌کند. تفاضل پاسخ حقیقی شبکه و پاسخ مورد انتظار که خطا نامیده می‌شود، در جهت مخالف اتصالات در شبکه منتشر می‌شود و اوزان به گونه‌ای تغییر می‌یابند که پاسخ حقیقی شبکه به پاسخ مطلوب نزدیکتر شود (شکل (۱-۵)). مراحل الگوریتم پس انتشار را می‌توان به صورت زیر بیان کرد:

۱- شبکه یک مثال آموزشی دریافت می‌کند و با استفاده از اوزان موجود در شبکه خروجی‌ها را محاسبه می‌کند.

۲- خطا یعنی اختلاف بین نتیجه محاسبه شده و مقدار مورد انتظار محاسبه می‌شود.

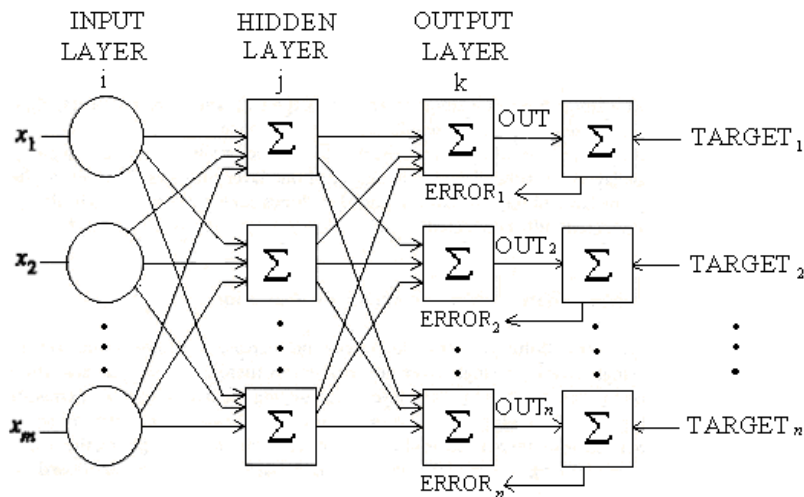
۳- خطا درون شبکه منتشر می‌شود و اوزان برای حداقل نمودن خطا از نو تنظیم می‌شوند.

مهمترین بخش این الگوریتم، تنظیم اوزان نرون‌ها است. این روش، اوزان را به گونه‌ای تغییر می‌دهد که خطا کاهش یابد. این تغییر بسیار کوچک است در واقع وزن را به آرامی به سمت مقدار بهینه‌اش سوق می‌دهد. پس از نشان دادن تعداد کافی مثال آموزشی، اوزان شبکه تغییر معنی‌داری نشان نمی‌دهند، در این صورت خطا نیز کاهش بیشتری نشان نخواهد داد. در چنین حالتی شبکه ورودی را فرا گرفته و آموزش پایان یافته است [۱۸]. همانطور که بیان شد در خلال آموزش، وزن‌ها و بایاس‌ها مکرراً برای کمینه‌کردن تابع عملکرد تعدیل می‌شوند. تابع عملکرد پیش فرض برای شبکه‌های پیش‌خور، میانگین مربعات خطا^۲ (MSE) بین خروجی‌های شبکه و خروجی‌های هدف، است. الگوریتم‌های پس‌انتشار مختلفی برای آموزش مدل‌های شبکه (تطبیق بایاس و وزن‌ها) وجود

^۱ - Delta rule

^۲ - Mean square error

دارد، که به آسانی برای تخمین آماری و تئوری بهینه‌سازی به کار می‌روند. متداول‌ترین الگوریتم که در الگوی آموزش پس‌انتشار به کار می‌رود، الگوریتم کاهش گرادیان^۱ می‌باشد [۳]. در ادامه‌ی این فصل، در مورد الگوریتم‌های مختلف آموزش برای شبکه‌های پیش‌خور، بحث می‌کنیم.



شکل (۵-۱) - شمای یک شبکه عصبی مصنوعی سه‌لایه و الگوریتم پس انتشار خطا در این الگوریتم، خروجی شبکه با جواب مسأله (Target) مقایسه می‌شود و خطای ایجاد شده در طول شبکه به عقب گزارش می‌شود. اوزان نرون‌ها لایه به لایه بر اساس این خطا تصحیح می‌گردند.

کاهش گرادیان، الگوریتمی است که در آن وزن‌ها و بایاس‌ها در جهت خلاف گرادیان خطا حرکت داده می‌شوند. هر مرحله کاهش گرادیان به خطای کمتری منتهی می‌شود تا اینکه مینیمم خطا بدست آید. معادلات برگشتی این الگوریتم بصورت زیر است:

$$w_{ji,n+1} = w_{ji,n} + \Delta w + \alpha \Delta w_{ji,n} \quad \text{رابطه‌ی (۱۳-۱)}$$

که $w_{ji,n+1}$ بردار جدید وزن‌ها و بایاس‌ها است. سطح خطا با استفاده از دستور زیر کاهش می‌یابد:

$$\Delta w_{ji,n} = -\mu \frac{\partial E}{\partial w_{ji,n}} \quad \text{رابطه‌ی (۱۴-۱)}$$

که μ و α به ترتیب، سرعت آموزش^۱ و ممنتوم^۲ هستند. عملکرد الگوریتم کاهش گرادیان تا حدودی به انتخاب مقادیر اولیه‌ی وزن بستگی دارد. هر چند در الگوریتم کاهش گرادیان با یک روند خطی

^۱ - gradient descent

(درجه‌ی اول) به کمترین مقدار می‌رسد، سرعت همگرایی آن کم می‌باشد و احتمال به دام افتادن در می‌نیم‌های محلی وجود دارد در نتیجه یادگیری شبکه سخت و کند است. افزودن عبارت ممنتوم در پس انتشار تاحدی به رفع مشکل همگرایی کمک می‌کند.

بنابراین روش‌های آموزش درجه‌ی دوم گوناگونی مانند الگوریتم‌های پس‌انتشار گوس - نیوتن^۳ (GN) و لونبرگ - مارکورت^۴ (LM) پیشنهاد شد. هر چند همگرایی الگوریتم گوس - نیوتن به میزان زیادی به انتخاب مقادیر اولیه‌ی وزن بستگی دارد، ولی خواص همگرایی درجه‌ی دوم باعث سرعت بالای آن می‌شود. روش بروزرسانی وزن‌های شبکه عصبی طبق روش گوس - نیوتن به صورت زیر می‌باشد:

$$\Delta w_{ji,n} = -H^{-1}g \quad \text{رابطه‌ی (۱-۱۵)}$$

که g گرادیان تابع، H ماتریس هسین^۵ که مشتقات مرتبه دوم تابع عملکرد نسبت به مقادیر فعلی پارامترهای شبکه را نشان می‌دهد [۳]. ماتریس هسین به شکل زیر نوشته می‌شود:

$$H = \nabla^2 F(x) = \begin{bmatrix} \frac{\partial}{\partial x_1^2} F(x) & \frac{\partial^2}{\partial x_1 \partial x_2} F(x) & \dots & \dots & \frac{\partial^2}{\partial x_1 \partial x_n} F(x) \\ \frac{\partial^2}{\partial x_2 \partial x_1} F(x) & \frac{\partial^2}{\partial x_2^2} F(x) & \dots & \dots & \frac{\partial^2}{\partial x_2 \partial x_n} F(x) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} F(x) & \frac{\partial^2}{\partial x_n \partial x_2} F(x) & \dots & \dots & \frac{\partial^2}{\partial x_n^2} F(x) \end{bmatrix}_{n \times n}$$

$$F(x) = F(x_1, x_2, \dots, x_n)$$

الگوریتم لونبرگ - مارکورت شکل تغییر یافته‌ی روش نیوتن می‌باشد که علاوه بر مزیت سرعت روش نیوتن دارای مزیت همگرایی سریع است. الگوریتم لونبرگ - مارکورت طوری طراحی

¹ - Learning rate

² - momentum

³ - Gauss-Newton

⁴ - Levenberg-Marquardt

⁵ - Hessian matrix

شده که به عنوان یک الگوریتم بهینه‌سازی حد واسط روش گوس - نیوتن و الگوریتم کاهش گرادیان عمل کند [۱].

الگوریتم لونبرگ - مارکورت از تقریب زدن ماتریس هسین بصورت زیر استفاده می‌کند:

$$H(x) = \nabla^2 F(x) \approx J^T J$$

$$g(x) = \nabla F(x) \approx J^T e$$

$$\Delta w_{ji,n} = -[J^T J + \lambda I]^{-1} J^T e \quad \text{رابطه‌ی (۱-۱۶)}$$

که e بردار خطای شبکه و J ماتریس ژاکوبین^۱ که حاوی مشتقات اول خطاهای شبکه با توجه به وزن‌ها و بایاس‌ها است.

چنانچه اسکالر λ کوچک باشد، این روش به روش نیوتن تبدیل می‌شود که از تقریب ماتریس هسین استفاده می‌کند. اگر مقدار λ بزرگ باشد، به گرادیان نزولی با اندازه گام کوچک تبدیل می‌شود. λ بعد از هر گام موفق (کاهش تابع عملکرد) کاهش می‌یابد. و تنها در صورتی افزایش می‌یابد که تابع عملکرد افزایش یابد. بنابراین تابع عملکرد در هر تکرار از الگوریتم کاهش می‌یابد [۳].

یکی از مشکلاتی که در خلال آموزش شبکه رخ می‌دهد، انطباق بیش از حد داده‌ها^۲ است. که در این حالت خطای مجموعه‌ی آموزش مقدار بسیار کوچکی می‌شود در حالی که هنگام کار با داده‌های جدید، خطای شبکه بسیار بزرگ می‌گردد. که اصطلاحاً گفته می‌شود شبکه مثال‌های آموزش را به خاطر سپرده است، اما تعمیم برای حالت‌های جدید را یاد نگرفته است [۲۰].

یک روش برای ارتقای عمومیت شبکه، بکارگیری شبکه‌هایی بزرگ برای ایجاد یک انطباق مناسب است. اگر ما از شبکه کوچکی استفاده کنیم، این شبکه توانایی لازم را برای انطباق داده‌ها، نخواهد داشت. متأسفانه دانستن این مورد از ابتدا، که اندازه‌ی شبکه برای یک کاربرد خاص چه میزان می‌بایست بزرگ باشد، مشکل است.

^۱ - Jacobian matrix

^۲ - Overfitting

روش دیگر برای ارتقای عمومیت (تعمیم)^۱ است. این روش تابع عملکرد را که عموماً

مجموع میانگین مربعات خطای شبکه برای مجموعه آموزش می‌باشد را تعدیل می‌کند:

$$F = E_D = \frac{1}{N} \sum_{i=1}^N (e_i)^2 \quad \text{رابطه‌ی (۱۷-۱)}$$

که E_D مجموع میانگین مربعات خطای شبکه می‌باشد. چنانچه تابع عملکرد را با اضافه نمودن قسمتی

که شامل میانگین مجموع مربعات وزن‌ها و بایاس‌های شبکه است، تعدیل نماییم، ارتقای عمومیت

امکان‌پذیر می‌گردد:

$$F = \beta E_D + \alpha E_W \quad \text{رابطه‌ی (۱۸-۱)}$$

$$E_W = \frac{1}{N} \sum_{i,j=1}^N w_{i,j}^2 \quad \text{رابطه‌ی (۱۹-۱)}$$

که α و β پارامترهای تابع عملکرد هستند و E_W مجموع مربعات وزن‌ها و بایاس‌های شبکه است.

استفاده از این تابع عملکرد سبب می‌شود که شبکه وزن‌ها و بایاس‌های کوچکتری داشته

باشد و این شبکه را هموارتر می‌کند و باعث می‌شود که پاسخ شبکه کمترین انطباق بیش از حد را

داشته باشد. مشکلی که در مورد تنظیم وجود دارد این است که تعیین مقدار بهینه پارامترهای تابع

عملکرد مشکل می‌باشد. یک رویکرد برای این پروسه چارچوب بایزین^۲ می‌باشد، که اولین بار توسط

دیوید مک‌کی^۳ پیشنهاد شد. در این چارچوب، فرض می‌شود که وزن‌ها و بایاس‌های شبکه متغیرهای

تصادفی با توزیع‌های مشخص و معلومی هستند. پارامترهای تعدیل به واریانس نامعلوم این توزیع‌ها

نسبت داده می‌شود. سپس این پارامترها با استفاده از تکنیک‌های آماری تخمین زده می‌شوند.

الگوریتم تنظیم بایزین^۴، در قالب تابع trainbr در جعبه ابزار نرم‌افزار MATLAB بکار برده

می‌شود [۲۱].

^۱ - Regularization

^۲ - Bayesian

^۳ - David McKay

^۴ - Bayesian regularization

۸-۱-۱- ارزیابی مدل

منظور از ارزیابی مدل، ارائه‌ی شاخص‌های کمی است که بتوان بوسیله‌ی آن دقت و صحت نتایج ارائه شده توسط مدل را ارزیابی کرد. از این شاخص‌ها ابتدا برای بهینه‌سازی پارامترهای مدل و سپس برای گزارش میزان اعتمادپذیری آن استفاده می‌شود. معمولاً برای ارزیابی شبکه، داده‌ها به سه دسته‌ی آموزشی، ارزیابی (آزمایشی) و تست تقسیم می‌شوند. ابتدا شبکه را توسط نمونه‌های آموزشی تمرین داده و وزن‌های آن را انتخاب می‌کنند. آنگاه کیفیت شبکه را در برخورد با نمونه‌های ارزیابی می‌سنجند. در صورت عدم پاسخ مناسب، مرحله‌ی آموزش را مجدداً تکرار کرده و مرحله‌ی آزمایش را اجرا می‌کنند. این کار تا زمان پاسخ مناسب به نمونه‌های ارزیابی ادامه می‌یابد. سپس برای سنجش قابلیت انطباق پذیری شبکه، آن را با نمونه‌های تست می‌سنجند [۲۱]. در این تحقیق متناسب با طبیعت پیوسته‌ی جواب‌ها از شاخص‌های ارزیابی مختلفی استفاده شد که نحوه‌ی محاسبه‌ی هر یک از پارامترها به صورت زیر می‌باشد.

۱- شاخص میانگین مربعات خطا^۱ که نحوه‌ی محاسبه‌ی آن به این شکل است:

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

رابطه‌ی (۲۰-۱)

۲- خطای استاندارد^۲

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

رابطه‌ی (۲۱-۱)

۳- پارامتر F : آزمون F یا آزمون فیشر در واقع آزمون معنی‌دار بودن آماری در تحلیل رگرسیون ساده و چندمتغیره، مشابه آزمون‌های معنی‌دار بودن آماری در تحلیل واریانس است [۲۲]. نسبت F از رابطه‌ی (۲۲-۱) به دست می‌آید:

^۱ - Mean Square Error

^۲ - Standard error

$$F = \frac{SS_{reg} / DF_{reg}}{SS_{res} / DF_{res}} \quad \text{رابطه‌ی (۲۲-۱)}$$

SS_{reg} : مجموع مجزورات رگرسیون DF_{res} : درجه آزادی رگرسیون

SS_{res} : مجموع مجزورات باقی مانده‌ها DF_{res} : درجه آزادی باقی مانده‌ها

۴- ضریب همبستگی^۱: مقدار این آماره بین ۱-۰ متغیر است. مقدار کوچک آن نشان‌دهنده‌ی این مسأله است که ارتباط خطی اندکی بین متغیر وابسته و متغیرهای مستقل وجود دارد یا چنین رابطه‌ای وجود ندارد. مجذور ضریب همبستگی چند متغیره یا R^2 ، مبین واریانس مشترک بین متغیرهای وابسته و مستقل است که به آن ضریب تعیین نیز می‌گویند، چون نسبتی از واریانس متغیر وابسته را که توسط متغیرهای مستقل به حساب آمده معین می‌کند و به صورت زیر محاسبه می‌شود [۲۲]:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad \text{رابطه‌ی (۲۳-۱)}$$

۵- خطای مطلق میانگین^۲

$$MAE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \times 100 \quad \text{رابطه‌ی (۲۴-۱)}$$

۶- خطای نسبی پیش‌بینی^۳

$$REP(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{رابطه‌ی (۲۵-۱)}$$

در فرمول‌های بالا $\hat{y}_i$ مقدار پیش‌گویی شده‌ی نمونه i ام و y_i مقدار حقیقی یا تجربی آن است، $\bar{y}$ مربوط به میانگین تجربی کل نمونه‌ها، n تعداد کل نمونه‌ها در مدل است.

^۱ - Correlation coefficient

^۲ - Mean absolute error

^۳ - Relative error of prediction

در مدل سازی فعالیت ترکیبات دارویی، بعضی از مشتقات که به طور تجربی سنتز شده اند، فاقد فعالیت دارویی (غیرفعال) می باشند. لذا لازم است مدلی ارائه شود که قادر باشد ترکیبات را به دو گروه فعال و غیرفعال تقسیم بندی نماید. در ارزیابی شبکه های عصبی به کار گرفته شده در کلاسه بندی ترکیبات به دو دسته فعال و غیرفعال، اگر N_{TP} ، تعداد ترکیبات فعال که مدل آن ها را درست دسته بندی کرده است، N_{TN} تعداد ترکیباتی غیرفعالی که مدل آن ها را درست دسته بندی کرده است، N_{FP} تعداد ترکیباتی که مدل به غلط آن ها را فعال شناخته است و N_{FN} تعداد ترکیباتی که مدل به غلط آن ها را غیرفعال شناخته است، باشند. با استفاده از این کمیت ها، می توان شاخص های کمی ارزیابی را مطابق زیر محاسبه کرد [۲۳]:

۱- نسبت صحیح^۱ (FC) درصدی از کل نمونه ها که به درستی دسته بندی شده اند:

$$FC = \frac{N_{TN} + N_{TP}}{N_{total}} \times 100 \quad \text{رابطه ی (۱-۲۶)}$$

۲- نسبت هشدار^۲ (FAR) درصدی از ترکیبات غیرفعال است که نادرست دسته بندی شده اند:

$$FAR = \frac{N_{FP}}{N_{TN} + N_{TP}} \times 100 \quad \text{رابطه ی (۱-۲۷)}$$

۳- احتمال آشکارسازی (POD) درصدی از ترکیبات است که به درستی دسته بندی شده اند:

$$POD = \frac{N_{TP}}{N_{FN} + N_{TP}} \times 100 \quad \text{رابطه ی (۱-۲۸)}$$

بر این اساس بهترین مدل آن است که بیشترین نسبت صحیح و احتمال آشکارسازی و کمترین نسبت هشدار را دارا باشد.

^۱ - Fraction Corren

^۲ - False Alarm Rate

۱-۲- کاربرد QSAR در طراحی دارو

QSAR به راحتی می‌تواند برای کمک به بهینه‌سازی سنتز داروها و مواد شیمیایی مورد استفاده قرار گیرد. تحقیقات برای پیشنهاد داروهای جدید در روش‌های داروسازی سنتی تا حد زیادی بر پایه‌ی آزمون و خطا و استخراج موادی با خاصیت دارویی از طبیعت بود. تا سال‌های میانی قرن بیستم مکانیسم عمل اکثر این داروها شناخته شده نبود. پس از شناسایی اصول مولکولی حیات و ایجاد روش‌هایی برای بررسی میان‌کنش مولکول‌های حیاتی از جمله پروتئین‌ها و اسیدهای نوکلئیک، دانشمندان به اطلاعاتی ارزشمند در مورد نحوه‌ی اثر مولکول‌های دارویی بر بدن دست یافتند. در این تلاش‌ها مشخص شد که اثر دارویی هر ترکیب به ویژگی‌های ساختاری آن ترکیب وابسته است. هر یک از این ویژگی‌های ساختاری، از جمله وجود گروه‌های فعال در ساختار ترکیبات، باعث بروز یکی از آثار دارویی مربوط به آن‌ها می‌شود. از این ویژگی‌ها می‌توان به خصوصیات وابسته به هر اتم مانند بارها و اوزان اتمی، خصوصیات مربوط به گروه‌های عاملی مانند تعداد گروه‌های عاملی و محل قرارگیری آن‌ها در مولکول و همچنین ویژگی‌های مربوط به کل مولکول همانند حجم و شکل فضایی اشاره کرد.

در علم داروسازی مدرن، با تکیه بر روش‌های QSAR تولید دارو به یک روند تحقیقاتی کاملاً جهت‌دار تبدیل شده است. در این روند، یک تیم داروسازی - که امروزه شامل متخصصان بسیاری از رشته‌هاست - به بررسی ویژگی‌های شیمیایی داروها در سطح مولکولی می‌پردازند. آن‌ها با تغییر برخی از این ویژگی‌ها، سعی در بهینه‌سازی ساختار دارو برای اهدافی مشخص دارند. در این روند پیچیده و طولانی، معمولاً مولکول‌هایی ساخته و به عنوان دارو در بازار عرضه می‌شوند که وجودشان در طبیعت بی سابقه است. شاید بتوان ادعا کرد بیشتر داروهای مدرن امروزی از روش طراحی بر اساس الهام از نمونه‌های طبیعی به دست می‌آیند.

یک ویژگی مهم داروهای تهیه شده با روش‌های QSAR گران بودن آن‌ها به علت پشتوانه‌ی سنگین تحقیقاتی برای تهیه‌ی آن‌هاست. هم چنین به طور کلی مشکل روش‌های QSAR وابستگی

آن‌ها به صحت داده‌های آزمایشگاهی است؛ به طوری که اگر خطا در آزمایش نمونه‌های مورد استفاده از مجموعه‌ی آموزشی مدل‌ها زیاد باشد، اعتمادپذیری مدل‌ها به شدت کاهش می‌یابد. برای انجام یک مطالعه QSAR ابتدا باید مقادیر فعالیت دارویی ترکیبات از طریق تجربی به دست آیند و همچنین توصیفگرهای مناسب از طریق تجربی یا محاسبه توسط کامپیوتر تولید شوند، مرحله‌ی بعد دستیابی به ارتباط ریاضی بین ساختار و فعالیت است. معمول‌ترین تکنیک چندمتغیری، رگرسیون خطی چندگانه (MLR) می‌باشد. با این حال این روش تنها زمانی به کار برده می‌شود که تعداد مولکول‌های مورد بررسی بسیار بیشتر از تعداد توصیف‌گرها باشند و همچنین متغیرها با هم همبستگی زیادی نداشته باشند. استفاده از شبکه‌های عصبی مصنوعی به جای روش رگرسیون خطی چندگانه در مطالعات QSAR، خصوصاً هنگامی که ارتباط بین توصیف‌گرها و فعالیت مورد نظر خطی نبوده و یا اینکه بین آن‌ها برهمکنش‌هایی وجود داشته باشد، باعث بهبود مدل حاصله خواهد شد. یکی از مهم‌ترین ویژگی‌های شبکه‌های عصبی مصنوعی در این است که سیستمی برنامه‌ریزی شده با قواعد از پیش تعیین شده نمی‌باشند و می‌توانند طی فرایندی که آموزش نامیده می‌شود، ساختار درونی خود را به نحوی تنظیم کنند که بهترین پاسخ ممکن را برای داده‌های ورودی ایجاد نمایند. با توجه به جنبه‌های کاربردی شبکه‌های عصبی مصنوعی و کارایی آن‌ها، در این پروژه برای مدل‌سازی و پیش‌بینی فعالیت دارویی مولکول‌های ضد سرطان و همچنین برای مدل‌سازی و پیش‌بینی فعالیت مشتقات بازدارنده‌ی یک نوع مرگ سلولی استفاده شده است [۲۴ و ۲۵].

۱-۳- سرطان و اهمیت مطالعه‌ی آن

در زمان‌های گذشته سرطان توسط محققان شناخته شد و به آن نام کانسر^۱ دادند، زیرا مانند پاهای یک خرچنگ به تمام نقاط مختلف بدن چنگ می‌اندازد. این اصطلاح تا حدی کلی است و مترادف با تومور بدخیم بکار می‌رود.

سرطان تنها ویژه‌ی انسان نیست و همه‌ی جانوران و گیاهان پرسلولی نیز ممکن است به سرطان دچار شوند. در مقایسه با سرطان بیماری‌های معدودی وجود دارند که موجب نگرانی و اضطراب می‌شوند. تأثیر فیزیولوژیکی و روانشناختی سرطان بر روی بیماران و خانواده‌هایشان منجر به تغییرات عمده‌ای در شیوه‌ی زندگی آنان می‌گردد. سرطان ممکن است برای برخی به معنای مرگ و برای برخی دیگر به مفهوم فلج شدن باشد تاکنون پیشرفت‌های زیادی در زمینه‌ی پیشگیری، کشف به موقع و درمان سرطان ایجاد شده است و تحقیق در این زمینه همچنان ادامه دارد. آمارها نشان می‌دهد که در بعضی موارد بیش از یک سوم جمعیت عمومی با بیماری سرطان برخورد دارند و بیش از ۱۰٪ تمامی مرگ‌ها در جهان را موجب شده است. به طوری که در تمام کشورهای توسعه یافته سرطان مسؤول اختصاص بیش از ده درصد تمام هزینه‌های مراقبت پزشکی است. بنابراین، تحقیق و کاوش پیرامون علل بوجود آمدن سرطان و در نهایت راه‌های پیشگیری و درمان آن بسیار پر اهمیت و حیاتی به نظر می‌رسد. حتی مردم عادی هم می‌دانند که در یک تشخیص پزشکی وجود سرطان چه معنایی دارد و بیشتر پزشکان علی‌رغم تخصصی که دارند در طول زندگی حرفه‌ای خود با بیماران سرطانی برخورد کنند. از همین رو اهمیت مطالعه‌ی سرطان نیاز به تأکید ندارد، تنها ذکر این نکته کافی است که سرطان‌های بدخیم دومین علت مرگ و میر در ایالات متحده آمریکا به شمار می‌روند و در این میان تنها بیماری‌های قلب و عروقند که قربانیانی بیشتر از مردم را می‌گیرد.

^۱ - Cancer

سرطان، رشد و تکثیر و گاهی انتشار غیرطبیعی سلول‌های بدن می‌باشد. تمام سرطان‌ها دارای الگوی رشد مهارگسیخته و تمایل به جدا شدن از منشاء اصلی و متاستاز^۱ هستند. به عبارت دیگر سرطان بیماری است که در آن سلول‌های بدن به طور کنترل نشده‌ای تکثیر می‌یابند. در یک جاندار سالم همیشه بین میزان تقسیم سلول، مرگ سلولی و تمایز، یک تعادل وجود دارد.

سرطان شامل همه تومورهای بدخیم می‌شود که در پزشکی بیشتر آن را نئوپلاسم^۲ می‌شناسند. به عبارتی سرطان یا نئوپلاسم به توده‌ای از بافت غیرطبیعی که رشدی فزاینده و ناهماهنگ با رشد نسوج طبیعی دارد و با قطع تحریکاتی که باعث این دگرگونی شده‌اند، باز هم به رشد نابهنجار خود ادامه می‌دهند، اطلاق می‌شود. چیزی که زیربنا و پایه و اساس تمام سرطان‌ها است، پاسخ ندادن و بی‌جواب گذاشتن تدابیر کنترل طبیعی رشد است. به این علت می‌گویند سلول‌های نئوپلاستیک ماهیتی دگرگون شده دارند و به طور پی در پی و مستمر تقسیم می‌شوند، بی آنکه به تحریکات تنظیم‌کننده‌ای که رشد طبیعی را تحت کنترل خود دارند هیچ‌گونه توجهی داشته باشند. علاوه بر این، نئوپلاست‌ها دو خصوصیت دیگر هم دارند، یکی این که به نظر می‌آید رفتار انگلی داشته باشند و بر سر نیازهای سوخت و سازشان با سلول‌ها و بافت‌های طبیعی به رقابت می‌پردازند، بنابراین در بیمارانی که از قبل ضعیف و رنجور هستند بیشتر خودنمایی می‌کنند. دیگر اینکه نئوپلاست‌ها تا حدی از خود مختاری برخوردارند و بدون توجه به محیط و شرایط تغذیه‌ای میزبان کم و بیش به طور ثابت و مستمر به افزایش اندازه خود ادامه می‌دهند، بنابراین همان‌طور که ملاحظه شد تا اندازه‌ای رویش‌های غیر قابل کنترل محسوب می‌شوند [۲۶ و ۲۷].

امروزه اکثر سرطان‌ها درمان قطعی ندارند و برای جلوگیری از رشد و پیشرفت آن‌ها از روش‌های جراحی، پرتو درمانی، شیمی درمانی، هورمون درمانی، پیوند مغز استخوان و یا استفاده از روش‌های ترکیبی درمانی مانند استفاده از ترکیب جراحی، رادیوتراپی و شیمی درمانی در جهت

^۱ - Metastasis

^۲ - Neoplasm

درمان یا افزایش عمر فرد بیمار استفاده می‌نمایند و مطالعات مستمر در خصوص داروهای شیمی درمانی به درمان‌های ترکیبی جدیدی برای بسیاری از انواع سرطان منتج شده است [۲۶ و ۲۷].

۱-۳-۱- تیروزین کیناز^۱ - نقش و اهمیت آن در سرطان

آنزیم‌های تیروزین کیناز میانجی‌های مهم آبشار پیام دهی^۲ هستند که نقش کلیدی تعیین‌کننده‌ای در پاسخ به محرک‌های درونی و بیرونی در فرایندهای بیولوژیکی گوناگون مثل رشد، تکثیر، تفکیک، متابولیسم و آپوپتوسیس^۳ دارند. فعالیت تیروزین کیناز در سلول‌های نرمال بطور پیوسته تنظیم می‌شود. اما ممکن است که طی فرایندهایی مثل جهش^۴، بیان فراوان^۵، تحریکات خودترشحی^۶، فعالیت آن‌ها تغییر کرده و به بیماری منتهی شود.

حدود ۵۸ تیروزین کیناز گیرنده^۷ (RTK) شناخته شده است. مکانیسم عمل تیروزین کیناز در شکل (۱-۶) نشان داده شده است. تیروزین کیناز گیرنده بوسیله اتصال لیگاند به ناحیه‌ی خارج سلول فعال می‌شود، به دنبال آن دیمریزاسیون گیرنده انجام می‌شود. مطالعات ساختاری هسته‌ی کاتالیتیکی چندین تیروزین کیناز گیرنده همراه با مطالعات سینتیکی و بیوشیمیایی فسفوریلاسیون اثبات کرد که الیگومر شدن گیرنده غلظت موضعی تیروزین کیناز گیرنده را افزایش می‌دهد و باعث فعال شدن ناحیه‌ی درون سلولی که دارای فعالیت کینازی است، می‌شود. آمینواسیدهای تیروزینی در C انتهایی گیرنده در ناحیه درون سلولی فسفریله می‌شود. بمحض شروع فسفوریلاسیون تیروزین، گروه فسفات از ATP، به تیروزین منتقل می‌شود و ATP به ADP تبدیل می‌شود. این مناطق فسفریله جایگاه اتصال بسیاری از پروتئین‌هایی است که در فعال‌سازی و کنترل چندین مسیر انتقال پیام نقش دارند.

^۱ - Tyrosine kinase

^۲ - Signaling cascade

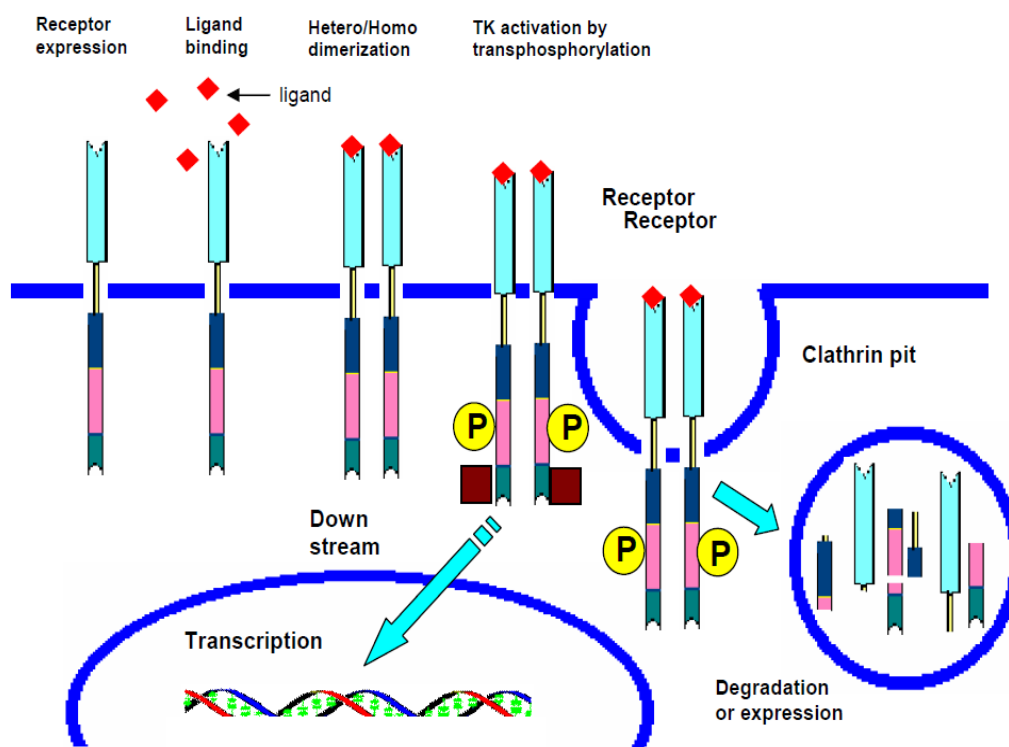
^۳ - Apoptosis

^۴ - Mutation

^۵ - Overexpression

^۶ - Autocrine simulation

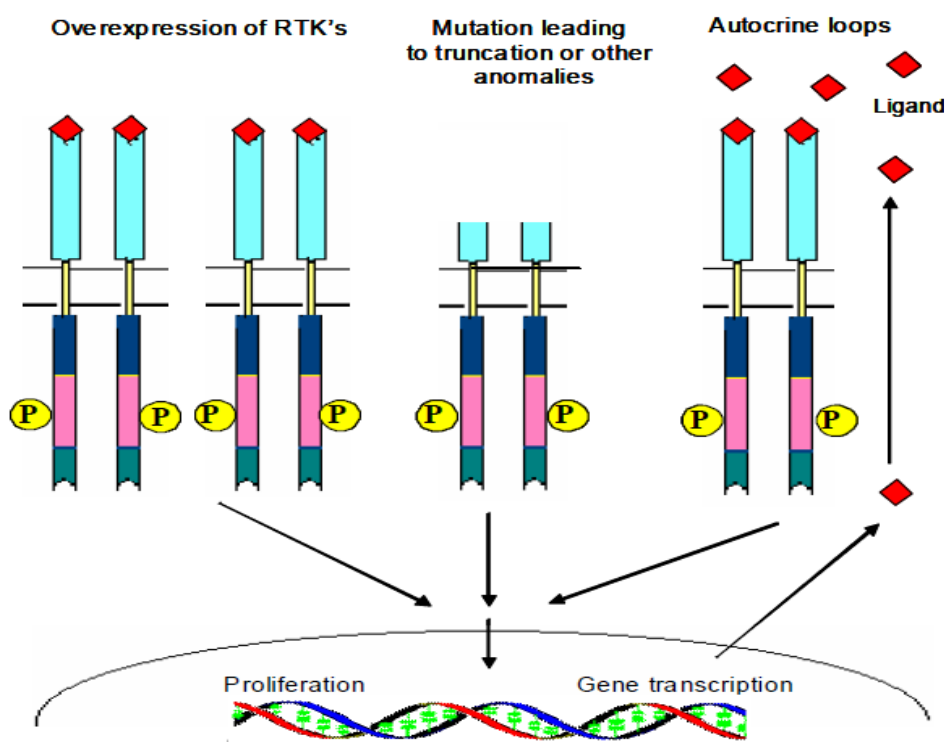
^۷ - Receptor tyrosine kinase



شکل (۶-۱) - مکانیسم عمل تیروزین کیناز.

به طور نرمال سطح فسفوریلاسیون تیروزین کیناز در سلول کنترل می‌شود. چندین مکانیزم وجود دارد که باعث تغییر عمل تیروزین کیناز می‌شوند (مانند جهش، بیان فراوان) اما نتیجه‌ی نهایی افزایش فعالیت تیروزین کیناز می‌باشد که منتهی به فعال شدن دیگر پروتئین‌های سیگنال‌دهنده و پیام‌رسان‌های ثانویه می‌شوند که مانع عملکرد تنظیمی در پاسخ‌های سلولی مانند تقسیم سلولی، رشد و مرگ سلولی می‌شوند (شکل (۷-۱)). فعال شدن تیروزین کیناز در فسفوریلاسیون بافت سرطانی در اثر فرایندهایی در طی انتقال پیام^۱ - فرایندهایی که بوسیله‌ی آن اطلاعات بیولوژیکی منتقل می‌شود - که یک مرحله‌ی مهم در ایجاد تومورهای سرطانی می‌باشد، اتفاق می‌افتد. بنابراین اثر کلیدی تیروزین کیناز در ایجاد تومور سرطانی بسیار قابل تأمل است و اخیراً تیروزین کینازها بعنوان هدف برای داروهای ضد سرطان بکار می‌روند [۲۸].

^۱ - Signal transduction



شکل (۷-۱) - طرح کلی مکانیسم‌های مختلف که منجر به فعالیت تومورزایی تیروزین کیناز می‌شوند.

۱-۳-۲- گیرنده‌ی فاکتور رشد اپیدرمی^۱ و رابطه‌ی آن با سرطان

گیرنده‌ی فاکتور رشد اپیدرمی (EGFR) یکی از اعضای مهم خانواده تیروزین کیناز گیرنده (RTK) محسوب می‌شود. گیرنده‌ی فاکتور رشد اپیدرمی دارای گیرنده‌های بزرگی در سطح سلول است که پس از اتصال لیگاند EGF فعال می‌شوند. گیرنده‌ی فاکتور رشد اپیدرمی از نظر ساختمانی یک گلیکوپروتئین ۱۷۰ کیلو دالتونی است و ژن آن روی کروموزوم ۷q۲۷ قرار دارد. وقتی گیرنده‌ی فاکتور رشد اپیدرمی فعال شود، بسیاری از فرایندهای سلولی از قبیل تکثیر سلولی، تمایز، مهاجرت، چسبندگی، محافظت در برابر آپوپتوسیس، تحریک رگ‌زایی، آغاز می‌گردد. افزایش بیان گیرنده‌ی فاکتور رشد اپیدرمی و لیگاند آن باعث ایجاد بسیاری از سرطان‌ها از جمله سرطان ریه با یاخته‌های

1- Epidermal Growth Factor Receptor

غیرکوچک، سرطان مثانه، سرطان سینه و سرطان مولتی فرم گلیوبلاستوما^۱ می شود. بیان فراوان گیرنده‌ی فاکتور رشد اپیدرمی در سرطان ریه با یاخسته‌های غیرکوچک حدود ۸۰-۴۰ درصد و در سرطان ریه با یاخسته‌های کوچک حدود ۵۰ درصد افزایش می‌یابد. فعالیت تومورزایی تیروزین کیناز در سلول‌های سرطانی می‌تواند بوسیله انتخاب مهارکننده‌های تیروزین کیناز متوقف شود. روش‌های مختلفی برای مهار تیروزین کیناز وجود دارد، که یکی از آن روش‌ها بازدارنده‌ی به وسیله بازدارنده‌های فسفریلاسیون تیروزین با وزن مولکولی پایین انجام می‌شود. در سال ۱۹۸۰ بازدارنده‌های با وزن مولکولی پایین پیشنهاد شد که می‌توانستند مانع تکثیر سلولی وابسته به لیگاند EGF شوند. بیشتر بازدارنده‌های تیروزین کیناز تقلیدی از ATP هستند و حداقل دو حلقه آروماتیک دارند. اختلاف کوچکی در ATP منجر به تغییراتی در پیوند هیدروژنی و برهمکنش‌های هیدروفوبیکی شده و سبب تغییراتی در تمایل آن می‌شود. بازدارنده در جایگاه پیوند ATP واقع در شکاف تیروزین کیناز اتصالات پایدار با باقیمانده‌ی تیروزین کیناز برقرار کرده و از فسفریلاسیون در لوپ فعال جلوگیری می‌کند [۲۸]. در این تحقیق رابطه‌ی کمی ساختار - فعالیت بازدارنده تیروزین کیناز گیرنده‌ی فاکتور رشد اپیدرمی مورد بررسی قرار گرفت.

^۱ - Multiform Glioblastoma

۱-۴- مرگ سلول

برای هر سلول یک دوره‌ی حیات و یک لحظه‌ی مرگ وجود دارد. مرگ سلولی از لحاظ مورفولوژیک به دو نوع تقسیم می‌شود: مرگ سلولی برنامه‌ریزی شده و نکروزیس^۱. نکروزیس در نتیجه‌ی یک آسیب خارجی مثل مواد شیمیایی سمی و یا آسیب‌های مکانیکی ایجاد می‌شود، اما در موارد بسیاری نیز سلول دست به انتحار می‌زند. مرگ سلولی برنامه‌ریزی نشده، نوعی مرگ سلولی تصادفی و غیرطبیعی است، که سبب از هم گسیختگی غشاء سلولی و تورم بافتی می‌شود. این امر موجب شروع پاسخ التهابی در بافت‌های مجاور می‌گردد [۲۹]. بر خلاف آن آپوپتوزیس^۲ یا مرگ برنامه‌ریزی شده‌ی سلول، اصلی‌ترین مکانیسم در تکامل و هموستاز بافت‌های بالغ در جهت حذف سلول‌های غیرضروری، آلوده، موتاسیون‌یافته و یا آسیب‌دیده بواسطه‌ی مسیرهای خودکشی داخلی است و دارای روند مشخص و متوالی می‌باشد. واژه‌ی آپوپتوزیس برای اولین بار در سال ۱۹۷۲ میلادی توسط محقق بنام کر^۳ معرفی شد. هر گونه اختلال در روند آپوپتوزیس، منجر به بیماری می‌شود که می‌تواند ناشی از کاهش مرگ سلولی باشد که منجر به ایجاد و رشد سلول‌های سرطانی و یا اختلالات خودایمنی می‌گردد. بالعکس افزایش غیرطبیعی مرگ سلولی در بیماری‌هایی نظیر اختلالات نورودژنراتیو^۴ و ایدز دیده می‌شود. داروهای شیمی درمانی سبب القاء آپوپتوزیس در سلول‌های سرطانی می‌شوند [۳۰].

همانطور که بیان شد دو مکانیسم اصلی مرگ سلولی عبارتند از آپوپتوزیس و نکروزیس که این دو با هم تفاوت‌هایی دارند که می‌توان به شرح زیر خلاصه کرد:

^۱ - Necrosis

^۲ - Apoptosis

^۳ - Kerr

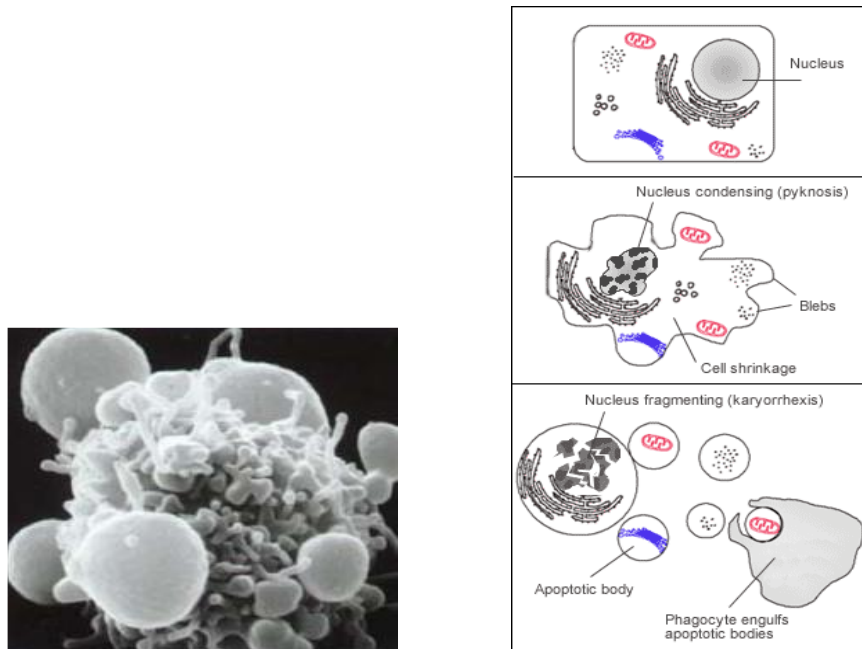
^۴ - Neurodegenerative

- آپوپتوزیس مرگ فیزیولوژیک سلولی است و در طی تحریکات خاصی اتفاق می‌افتد. در حالی که نکروزیس مرگ پاتولوژیک سلول بوده و در طی آسیب‌های شدید به سلول از قبیل هیپوکسی، هیپرترمی و سموم خارجی، این نوع مرگ سلولی اتفاق می‌افتد.
- نکروزیس فرایند غیرفعال است و در غیاب ATP نیز اتفاق می‌افتد در حالی که فرایند آپوپتوزیس فعال بوده و به انرژی وابسته است.
- در فرایند آپوپتوزیس سلول چروکیده و کوچک می‌شود در حالی که در نکروزیس سلول متورم و بزرگ می‌شود.
- در آپوپتوزیس غشاء سیتوپلاسمی به صورت اجسام آپپتوتیک در می‌آیند، در حالی که در نکروزیس غشاء تخریب می‌شود و سبب آزاد شدن محتویات داخلی سلولی می‌شود.
- ارگانل‌های سیتوپلاسمی در فرایند آپوپتوزیس دست نخورده باقی می‌مانند، در حالی که در نکروزیس تخریب می‌شوند.
- تراکم کروماتین و قطعه قطعه شدن آن در روند آپوپتوزیس مشاهده می‌گردد.
- در بافت وقوع نکروزیس همراه با واکنش‌های التهابی است، در حالی که آپوپتوزیس بدون التهاب رخ می‌دهد.

اگرچه در روند آپوپتوزیس در سلول‌های مختلف تفاوت‌هایی وجود دارد، اما به طور کلی مرگ سلولی برنامه ریزی شده به نوع سلول و شرایطی که موجب مرگ سلول می‌شود بستگی دارد. در این پدیده فرایندهای بیوشیمیایی و مورفولوژیک بسیاری وجود دارد که موجب کنترل جمعیت سلولی می‌شود [۳۱]. شکل (۸-۱) ویژگی‌های مورفولوژی آپوپتوسیس را نشان می‌دهد.

روند آپوپتوزیس طی چند مسیر مختلف انجام می‌شود که یکی از این مسیرها، آپوپتوزیس القاء شده توسط گیرنده‌های مرگ است. در غشاء پلاسمایی اغلب سلول‌ها گیرنده‌های مرگ وجود دارد. زمانی که این گیرنده‌ها توسط لیگاندهای مربوطه تحریک شوند، سبب فعال‌شدن کاسپازها و القاء آپوپتوزیس می‌گردند. ویژگی این گیرنده‌ها وجود توالی غنی از سیستئین در بخش خارج سلولی

است. این گیرنده‌ها در بخش سیتوپلاسمی خود دارای توالی به نام ناحیه‌ی مرگ^۱ بوده و از این رو در انتقال پیام آپوپتوزی به درون سلول شرکت می‌نمایند.



شکل (۸-۱) - ویژگی‌های مورفولوژی آپوپتوسیس

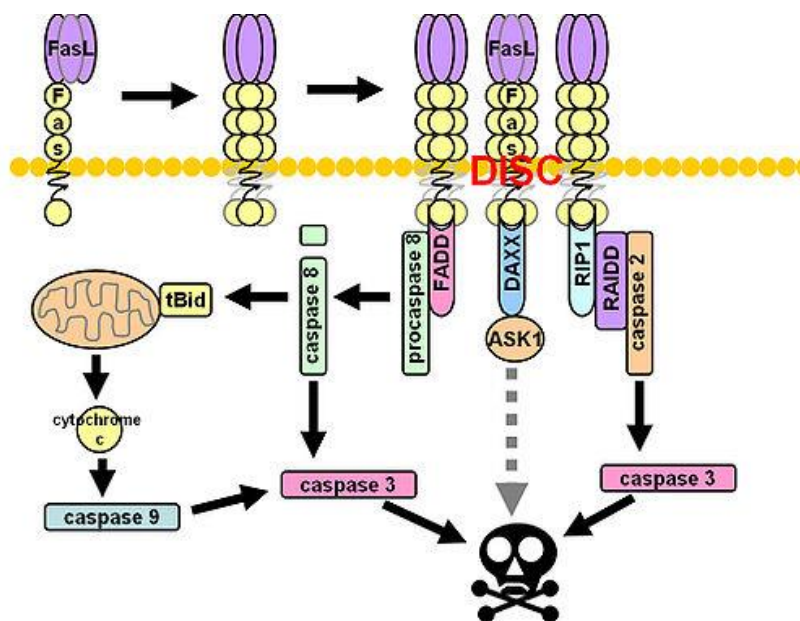
شناخته‌ترین گیرنده‌های مرگ عبارتند از $TNF\ \alpha$ ، $TNF\ \beta$ ، Fas. تحریک گیرنده‌های مرگ توسط لیگاندی‌های مربوط، منجر به تریمریزاسیون گیرنده و به کارگیری پروتئین‌های آداپتور می‌گردد. برای مثال لیگاند Fas (FasL) با گیرنده مربوط به خود (Fas) میان‌کنش نموده و باعث القاء تریمریزاسیون آن می‌شود. این عمل باعث تشکیل خوشه‌ی مرگ در ناحیه سیتوزولی گیرنده گردیده و باعث می‌شود که پروتئین‌های آداپتور از قبیل FADD^۲ به آن متصل‌شود. FADD دارای ناحیه‌ی مرگ (DD) در C - ترمینال است که این پروتئین را قادر می‌سازد تا به گیرنده تریمریزه شده از طریق میان‌کنش ناحیه مرگ - ناحیه مرگ متصل‌گردد. همچنین این پروتئین در ناحیه N-ترمینال دارای ناحیه‌ی موثر مرگ^۳ (DED) می‌باشد که با ناحیه‌ی DED مشابه در پرودومین کاسپاز ۸-

^۱ - Death domain

^۲ - Fas-associate death domain

^۳ - Death effector domain

برهم‌کنش می‌کند. به این کمپلکس پروتئین‌ها (مجموعه لیگاند - گیرنده مرگ، مولکول آداپتور و پروکاسپاز)، کمپلکس علامت‌دهنده‌ی القاء مرگ^۱ (DISC) گفته می‌شود. به این طریق پروکاسپاز-۸ به کاسپاز-۸ فعال تبدیل می‌شود. کاسپاز-۸ فعال سبب فعال شدن کاسپازهای اجرایی شده و در نتیجه فرایند آپوپتوزیس صورت می‌گیرد. در شکل (۹-۱) به طور شماتیک مسیر خارجی آپوپتوزیس نشان داده شده است [۳۲].



شکل (۹-۱)- واکنش‌های آبخاری فعال شدن کاسپازها توسط گیرنده‌های مرگ.

نقش آپوپتوزیس در سیستم ایمنی بسیار مهم است. به عنوان مثال اختتام پاسخ‌های ایمنی، حذف سلول‌های خود واکنش‌گر^۲ و در مجموع ایجاد هموستاز و تحمل در سیستم ایمنی، مدیون پدیده‌ی آپوپتوزیس می‌باشد [۳۰].

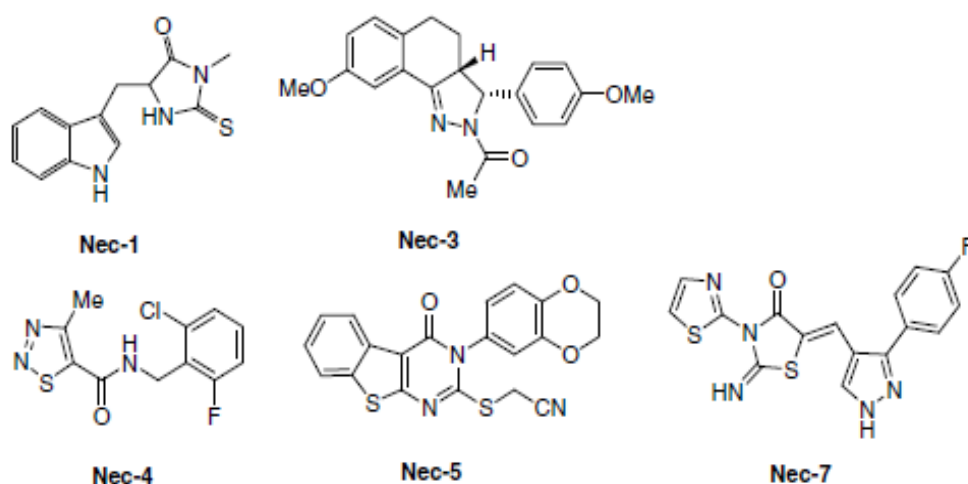
در چند سال اخیر جذابیت آپوپتوزیس افزایش یافته و بیشتر از نکروزیس مورد اهداف درمانی قرار گرفته است، زیرا نکروزیس غیرویه‌زه، تنظیم نشده و تصادفی می‌باشد. اما اخیراً الکسی دگترو^۳ و همکارانش یک نوع از نکروزیس را کشف کردند که مانند آپوپتوزیس کنترل شده

^۱ - Death-inducing signaling complex

^۲ - Autoreactive

^۳ - Alexei Degterev

می‌باشد و مسیر بیوشیمیایی ویژه‌ای دارد ولی ویژگی‌های مورفولوژی نکروزیس را داراست و آن را نکروپتوزیس نام نهادند. این نوع از مرگ سلولی در گستره‌ی وسیع پاتولوژی انسان شامل ضربه، کم‌خونی موضعی، نورودژنراتیو، حمله‌ی قلبی و مغزی مشاهده می‌شود. با تحریکات گیرنده‌های ناحیه‌ی مرگ خانواده‌ی Fas و TNF نکروپتوزیس القاء می‌شود که از ویژگیهای ساختاری هم آپوپتوزیس و هم نکروزیس سهم می‌برد. کشف نکروپتوزیس استراتژی‌های درمانی جدیدی ایجاد می‌کند. لذا شناخت و بهینه‌سازی مولکول‌های با وزن مولکولی پایین (به عنوان مثال نکروستاتین‌ها) با قابلیت بازدارندگی نکروپتوزیس در حذف نقش این فرایند در پاتو - فیزیولوژی بیماری برای توسعه‌ی درمانی قابل استفاده می‌باشد. با این هدف ساختار چندین نکروستاتین شناخته شده‌است. مکانیزم چگونگی عمل بازدارنده کاملاً واضح نیست. احتمالاً از طریق مهار برخی آنزیم‌های آبشار پیام‌دهی (کاسپازها) صورت می‌گیرد. این اجزاء کوچک با وزن مولکولی پایین با اتصال به گیرنده‌های خاص موجود در پروتئین باعث غیر فعال شدن آن در پیام‌دهی می‌شود. ساختار چندین نکروستاتین در شکل (۱-۱۰) نشان داده شده است [۳۳]. در این تحقیق ما رابطه کمی ساختار - فعالیت نکروستاتین-۵ (Nec-5) را مورد بررسی قرار دادیم.



شکل (۱-۱۰) - ساختار نکروستاتین‌ها

۱-۵- مروری بر کارهای انجام شده در QSAR ترکیبات دارویی

در سال ۱۹۹۷، ترونکت^۱ با استفاده از مطالعات QSAR برهمکنش ^۲HEPT را با ^۳HIV، با استفاده از شبکه عصبی مصنوعی پیش‌بینی نمود [۳۴].

کای^۴ و همکارانش در سال ۱۹۹۸ با استفاده از مدل شبکه عصبی سایت‌های گسسته‌ی پروتئاز را پیش‌گویی کردند [۳۵].

جلالی هروی و همکارانش در سال ۲۰۰۰ در یک مطالعه‌ی QSAR، فعالیت ترکیبات ضد ایدز را برای یک گروه بزرگ از مشتقات HEPT بررسی کردند. در این کار روش رگرسیون خطی با شبکه‌ی عصبی مقایسه شده است که در نتیجه مدل شبکه عصبی مصنوعی به عنوان پیش‌گویی‌کننده‌ی بهتری نسبت به مدل رگرسیون خطی معرفی گردیده است [۳۶].

در سال ۲۰۰۰ بوردن^۵ مقاله‌ای تحت عنوان کاربرد روش تعیین اتوماتیک ارتباط^۶ (ARD) در مطالعات QSAR با استفاده از شبکه عصبی بایزین ارائه داد و نشان داد که کوپل شبکه‌های عصبی مصنوعی بایزین و ARD، برای حل مشکلاتی که در مدل‌های QSAR ایجاد می‌شود، (مانند انتخاب مدل، قدرت مدل، انتخاب سری متغیرها و ویژگی شبکه بهینه) توانایی زیادی دارند.

روش ARD تعیین می‌کند که اندیس‌هایی که همبستگی زیادی با هم دارند، باید نادیده گرفته شوند یا در مدل‌سازی متغیر مهمی هستند. سری داده‌ها شامل سه سری مجزا هستند که از چند مقاله جمع‌آوری شده‌اند و قدرت پیش‌بینی روش‌های رگرسیون خطی چندگانه، شبکه عصبی پس‌انتشار عادی، شبکه عصبی با الگوریتم تنظیم بایزین (BRANN) و شبکه عصبی با الگوریتم تنظیم بایزین کوپل روش تعیین اتوماتیک ارتباط (ARD-BRANN) برای هر سری داده‌ها مقایسه شده است [۳۷].

^۱ - Tronchet

^۲ - 1-(2-hydroxyethoxymethyl)-6-(phenylthio)thymine

^۳ - Human immunodeficiency virus

^۴ - Cai

^۵ - Burden

^۶ - Automatic Relevance Determination

بودن برای تکمیل کار خود در سال ۲۰۰۱ مطالعات QSAR همان سری از ترکیبات را با روش گوسین که توسط مک‌کای ارائه شده است، شرح داد و با دیگر روش‌های نامبرده مقایسه کرد [۳۸].

در سال ۲۰۰۱ مقاله‌ای تحت عنوان تشخیص ایمنی آزمایشگاهی HIV با استفاده از شبکه‌ی عصبی توسط هات زاکیس^۱، ارائه شده است [۳۹].

در سال ۲۰۰۲ وابستگی قدرت جهش‌زایی رنگینه‌های آمینوآزوبنزن به ساختار آن‌ها توسط آشیش گارگ^۲ و همکارانش مورد بررسی قرار گرفت. در این تحقیق توصیف‌گرهای ۴۳ مشتق آمینوبنزن با استفاده از روش نیمه‌تجربی AM1 محاسبه شدند. مدل‌ها بر اساس رگرسیون خطی چندگانه و شبکه‌های عصبی مصنوعی برای بیش از ۸۰٪ ترکیبات درست گزارش شده بود. نتایج حاصل از انتخاب ۳، ۴ و ۵ توصیف‌گر بوسیله‌ی رگرسیون خطی چندگانه گزارش شده است. شبکه با کلاس‌های سه تایی از ۵۰ توصیف‌گر انتخاب شده با روش رگرسیون خطی چندگانه انجام شد و شبکه با چندین ۳- توصیف‌گر آموزش داده شد که مقدار R^2 بین ۰/۸۴ – ۰/۸۸ می‌باشد [۴۰].

در همان سال (۲۰۰۲) آرجونا^۳ و همکارانش شبکه‌های عصبی پیش‌خور پرسپترون چند لایه با الگوی آموزش پس انتشار (BP) برای مدل‌سازی QSAR غیرخطی را مورد مطالعه قرار دادند. همچنین این روش با روش‌های دیگری چون رگرسیون خطی چندگانه (MLR) حداقل مربعات جزئی (PLS) و حداقل مربعات جزئی درجه‌ی دوم (QPLS) مقایسه شد. در این تحقیق گیرنده‌های آلفا آدرنالین متعلق به سه گروه (آلفا ۱، آلفا ۲ و آلفا ۱ و ۲) مطابق با اثرات دارویی استفاده شد. هر یک از آن‌ها با ۱۵ ویژگی شیمیایی شرح داده شدند. همچنین برای هر یک از آن‌ها ۶ پاسخ دارویی برای ساختن ماتریکس پاسخ‌های بیولوژیکی در نظر گرفته شد [۴۱].

¹ - Hatzakis

² - Ashish Garg

³ - Arjona

آشیش گارگ و همکارانشان در سال ۲۰۰۳ مطالعات QSAR و QSPR رنگینه‌های آمینو آزوبنزن را با گسترش سری داده‌ها مورد بررسی قرار دادند. تکنیک رگرسیون خطی چندگانه با استفاده از ۸ توصیف‌گر برای ۷۰٪ این ترکیبات پیش‌گویی صحیحی نشان داد. این مقدار با استفاده از شبکه‌ی عصبی مصنوعی کوپل شده با منطق فازی به ۹۵٪ رسید [۴۲].

در سال ۲۰۰۳ با استفاده از مطالعات QSAR فعالیت ضد ایدز برای مشتقات HEPT بررسی شد. در این مطالعه سه روش شامل رگرسیون خطی، شبکه عصبی پس‌انتشار و بهینه‌سازی تکاملی ادغام شده با شبکه عصبی، مقایسه گردید. در روش بهینه‌سازی تکاملی ادغام شده با شبکه عصبی برای بهینه‌سازی وزن‌ها و بایاس‌های شبکه عصبی بوسیله حداقل نمودن خطای بین مقادیر واقعی و خروجی‌ها، روش تکاملی استفاده شده است نتایج تأیید می‌کند که محاسبات تکاملی می‌تواند به عنوان یک روش آموزش شبکه عصبی اجرا شود. همچنین نتایج حاکی از این است که بایاس‌ها در آموزش اولیه و داده‌های تست روی عملکرد محاسبات کامپیوتر می‌توانند اثر بگذارند. شبکه بهینه شده ساختار ۱-۱۱-۹ دارد که نتایج شبکه با محاسبات تکاملی بهتر از نتایج شبکه پس انتشار می‌باشد [۴۳].

در سال ۲۰۰۴ مدل‌سازی روابط ساختار - فعالیت جهش‌زایی ۹۵ ترکیب آمین هتروآروماتیک و آروماتیک توسط والکوا^۱ و همکارانش به وسیله‌ی شبکه‌ی عصبی مصنوعی پس‌انتشار رقابتی انجام شد. الگوریتم ممانعت کروی^۲ برای تقسیم مجموعه‌ی داده‌ها به دو سری تست و آموزش استفاده شد. ارزیابی بهترین مدل حاصله بوسیله‌ی تست تصادفی، آموزش مبادله شده و سری تست انجام شد [۴۴].

^۱ - Valkova

^۲ - Sphere exclusion

دوبل^۱ و همکارانش در سال ۲۰۰۵ مطالعات روابط کمی ساختار - فعالیت آنالوگ‌های پالائول^۲ را مورد بررسی قرار دادند. آنالوگ‌های پالائول شامل مشتقات ۱- (۲- هیدروکسی فیل) اتانول و بازدارنده‌های تراکم پلاکت‌های خون می باشند. مدل شبکه‌ی عصبی پسانتشار و رگرسیون خطی چندگانه جهت ارتباط ساختار - فعالیت ترکیبات به کار گرفته شد [۴۵].

در همان سال (۲۰۰۵) حیات شاهی و همکارانش رابطه‌ی ساختار - فعالیت را برای بازدارنده‌های رقابتی آدنوزین دامیناز^۳ بوسیله‌ی دو مدل غیرخطی رگرسیون لوجستیک^۴ و شبکه‌ی عصبی مصنوعی مورد بررسی قرار دادند. آنزیم آدنوزین دامیناز یا به اختصار ADA یکی از آنزیم‌های مهم در مکانیسم پورین‌هاست. این آنزیم در همه‌ی بافت‌های بدن انسان فعالیت دارد البته بیشترین فعالیت آن در بافت‌های ایمنی از جمله تیموس و لنفوسیت‌هاست. بیش فعالیت این آنزیم در بیماری‌های پارکینسون و ایدز دیده شده است. علاوه بر این اختلالات این آنزیم در برخی آنمی‌ها و سرطان‌ها نیز مشاهده شده است. بازدارنده‌های آنزیم آدنوزین دامیناز علاوه بر کاهش برخی بیماری‌های یاد شده، در سرکوب دستگاه ایمنی بدن در پیوند اعضا برای جلوگیری از دفع آن‌ها اهمیت دارد. مدل‌ها برای طبقه‌بندی ترکیبات به دو کلاس به کار گرفته شد. استفاده از رگرسیون لوجستیک برای ۹۱/۶۶٪ ترکیبات درست پیش بینی کرده در حالی که با استفاده از شبکه‌ی عصبی ۲-۳-۱ و ۳-۴-۱، این مقدار به ۹۵/۸۳٪ رسید [۴۶].

فانگ زنگ^۵ و همکارانش در سال ۲۰۰۶ از شبکه عصبی پسانتشار برای مدل‌سازی ۴۲ ترکیب از نمک‌های چند گانه مونو و بیس آمونیوم بعنوان داروهای عصبی ترک دخانیات استفاده کردند. مدل QSAR غیرخطی (ANN) سطح قابل قبول همبستگی بین داده‌های تجربی و محاسباتی

¹ - Doble

² - Paeonol

³ - Adenosine deaminase

⁴ - logistic

⁵ - Fang Zheng

برای $\log(1/IC_{50})$ ترکیبات ارائه کرد. که r^2 برای سری آموزش ۰/۷۶ و برای سری ارزیابی ۰/۶۴ می‌باشد [۴۷].

در سال ۲۰۰۶ کاتریتزکی^۱ و همکارانش مقاله‌ای تحت عنوان فعالیت ضد هجوم ترکیبات آلی با استفاده از توصیف‌گرهای ساختاری ارائه دادند. هجوم یک معیار اندازه‌گیری فعالیت سلول تومور می‌باشد که نهایتاً منجر به سرطان می‌شود. فعالیت ۱۳۹ ترکیب بوسیله شبکه عصبی مصنوعی پیش‌گویی شد. مدل QSAR غیر خطی (ANN) برای ۷۱٪ از ۶۰ ترکیب سری آموزش، ۷۰٪ از ۳۲ ترکیب سری تست به طور صحیح ترکیبات را به دو دسته‌ی فعال و غیرفعال کلاسه‌بندی کرده است [۴۸].

کیونکسو^۲ نیز در سال ۲۰۰۶ از رگرسیون خطی چندگانه و شبکه عصبی مصنوعی برای مطالعه رابطه ساختار فعالیت عامل مخالف گیرنده‌ی AT_1 استفاده کرده است. فعالیت گیرنده AT_1 در رگ‌ها باعث تنگی جداره‌ی آن‌ها و افزایش فشار خون می‌شود، لذا عامل مخالف عمل گیرنده‌ی AT_1 ، دارویی برای درمان فشار خون می‌باشد. یک سری از ایمادازول‌ها (بیو فیل متیل) بعنوان سری داروها عمل مخالف گیرنده AT_1 انجام می‌دهند که بوسیله گروه دوپونت ارائه گردید. طبق نتایج این مقاله روش ANN سه لایه‌ی (۱-۶-۶) عملکرد بهتری را نسبت به روش رگرسیون خطی چندگانه نشان می‌دهد [۴۹].

در سال ۲۰۰۷ عرب چم جنگلی و همکارانش ارتباط ساختار - فعالیت ترکیبات آنتی ایدز را به وسیله‌ی شبکه‌ی عصبی مصنوعی با الگوریتم آموزشی لونبرگ - مارکورت پیش‌بینی کردند. توصیف‌گرها به وسیله‌ی نرم‌افزار DRAGON محاسبه شد و توصیف‌گرهای مهم بوسیله‌ی رگرسیون مرحله‌ای انتخاب شد. ترکیبات بطور تصادفی ۲ مجموعه شامل ۳۱ ترکیب سری آموزش و ۱۱ ترکیب سری ارزیابی تقسیم شدند. ارزیابی توسط سری ارزیابی و روش رد مرحله‌ای تک تک انجام شد. خطای

¹ - Katritzky

² - Qing Su

جذر مجذور میانگین (RMSE) و خطای مطلق میانگین (MAE) برای سری ارزیابی به ترتیب ۰/۰۴۲ و ۰/۰۲۴ می باشد. نتایج بیانگر توانایی مدل برای پیش بینی داده های روابط ساختار - فعالیت داروهای آنتی HIV می باشد [۲].

در سال ۲۰۰۸، نانتاسنامات^۱ و همکارانش فعالیت بازدارنده های فورین را با استفاده از شبکه عصبی مصنوعی مدل سازی کردند. بازدارنده های فورین به عنوان داروی ضد سیاه زخم مصرف می شود. سیاه زخم یک بیماری مسری می باشد که توسط باکتری میله ای آنتراسیس ایجاد می شود. فعالیت سمی که از باکتریهای باسیل آنتراسیس خارج می شود، هنگامی رخ می دهد که فورین شروع به تقسیم پروتئین های سم می کند و چرخه تقسیم آنقدر زیاد می شود که به مرگ سلول منجر می شود. پیش بینی ها از طریق دو مرحله انجام شد: ۱- کلاسه بندی که فعال یا غیرفعال بودن ترکیب را پیش گویی می کند. ۲- توسعه مدل QSAR مولکول های فعال. بر اساس نتایج هم کلاسه بندی و هم مدل های QSAR راندمان خوبی داشتند. محاسبه ی توصیف گر ها توسط دو نرم افزار E-Dragon و RECON انجام شد. E-Dragon ورژن الکترونیکی نرم افزار Dragon می باشد که ۱۶۶۶ توصیف گر در ۲۰ بخش مجزا محاسبه می کند. نرم افزار RECON بوسیله بادر و همکارانش توسعه یافت که محاسبات بر اساس کوانتوم اتمی در مولکول ها انجام می شود و توصیف گر های محاسبه شده، هم ارز اتم انتقال پذیر نامیده می شوند که مولکول ها را در ترم های دانسیته الکترونی و انرژی توصیف می کنند [۵۰].

عرب چم جنگلی در سال ۲۰۰۹ شبکه ی عصبی مصنوعی را به عنوان یک شیوه ی مدل سازی در مطالعه ی رابطه ی ساختار - فعالیت برای پیش بینی داده های سیتوتوکسی سی تی (CC₅₀) مشتقات ۱- [۵- کلروفنیل]-H-۱-پیرول بکار گرفت و توانایی مدل حاصله را در پیش گویی داده های سیتوتوکسی سی تی مورد ارزیابی قرار داد [۳].

¹ - Nantasenamat

در سال ۲۰۰۹ باسینسکی^۱ و همکارانش فعالیت ضد قارچی یک سری از مشتقات پیریدین که عامل ضد یک نوع عفونت قارچی کاندیدا در انسان است، با استفاده از مطالعات QSAR توسط شبکه عصبی مصنوعی بررسی کردند. بکارگیری شبکه عصبی مصنوعی برای پیش بینی توانایی ضد میکروبی مشتقات پیریدین بر اساس توصیف‌گرهای ساختاری که بوسیله محاسبات بدست آوردند، امکان پذیر شد. مجموعه آموزشی شامل سه سری آموزش، ارزیابی و تست می‌باشد و R^2 برای سری آموزش، ۰/۹۲۴۶، برای سری ارزیابی، ۰/۸۵۴۲ و سری تست ۰/۹۱۲ می‌باشد [۵۱].

^۱ - Bucinski

فصل دوم

محاسبات و مدل سازی

۲-۱- محاسبات و مدلسازی بازدارندهی تیروزین کیناز گیرندهی فاکتور رشد اپیدرمی

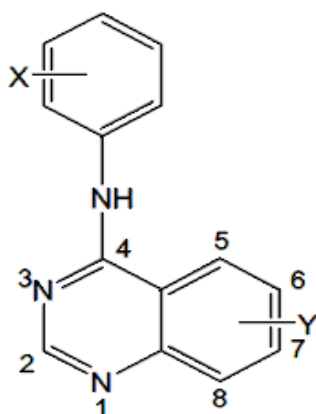
در این تحقیق، از شبکهی عصبی مصنوعی به عنوان یک شیوه مدلسازی در مطالعهی رابطهی ساختار - فعالیت برای پیش‌بینی داده‌های $\log(1/IC_{50})$ (که نشان‌دهندهی قدرت مهارکنندگی مشتقات ۴- (۳- برموانیلینو)-۶ و ۷ دی متوکسی کینازولین^۱ است، استفاده شده است و توانایی مدل حاصله در پیشگویی قدرت بازدارندگی در مقایسه با مدل رگرسیون خطی چند گانه مورد ارزیابی قرار گرفت.

۲-۱-۱- سری داده‌ها

سری داده‌های آماده برای ۶۰ ترکیب از مشتقات ۴- (۳- برموانیلینو)-۶ و ۷ دی متوکسی کینازولین از مرجع [۵۲] گردآوری شده است. ساختار پایه مشتقات در شکل (۱-۲) آمده است. در جدول (۱-۲) نیز جزئیات استخلاف‌ها و مقادیر عددی $\log(1/IC_{50})$ (بر حسب مولار متناظر هر ترکیب آورده شده است. IC_{50} یا غلظت مهارکنندگی نصف ماکزیمم^۲، غلظتی از دارو می‌باشد که برای بازدارندگی ۵۰٪ از محیط کشت مصنوعی مورد نیاز است. همانطور که داده‌های جدول (۱-۲) نشان می‌دهد، نوع استخلاف‌ها و نحوه‌ی آرایش آن‌ها در اطراف اسکلت اصلی مولکول تأثیر زیادی بر روی فعالیت دارویی ترکیب دارد. لذا در این تحقیق سعی شده است مدلی ارائه گردد که بتواند قدرت بازدارندگی هر ترکیب با هر نوع استخلاف در هر موقعیتی را پیش‌بینی نماید، بدون اینکه نیاز به سنتز و آزمایش ترکیب مورد نظر باشد.

^۱ - 4-(3-bromoanilino)-6,7-dimethoxyquinazoline derivatives

^۲ - Half maximal inhibitory concentration



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

۲-۱-۲- نرم افزارهای مورد استفاده

در این تحقیق از چندین نرم افزار آماده استفاده شده است که لیست کامل آن به شرح زیر

است:

۱- Hyperchem 7

۲- Dragon

۳- Spss-13

۴- MATLAB 7.5

از Hyperchem 7 [۵۳] برای ترسیم و بهینه سازی اولیه ی ساختار ترکیب استفاده شده است.

این نرم افزار ساختار ترکیبات را با چهار روش کلی آغازین، نیمه تجربی، کوانتوم مکانیک و تابع

دانشیته بهینه می کند. روش اول با عملکرد بسیار کند برای ترکیبات کوچک مناسب است؛ در حالی

که سه روش دیگر با عملکرد سریعتر برای ترکیبات متوسط و بزرگ مناسب ترند. نرم افزار Dragon

برای محاسبه ی توصیف گرهای مولکولی توسط گروه QSAR و کمومتریکس میلانو [۵۴] طراحی شده

است. SPSS [۵۵] یک نرم افزار تخصصی برای محاسبات آماری است. قابلیت این نرم افزار در

تحقیقات آماری در اکثر موارد بیشتر از نرم افزار Exell است. این نرم افزار برای انتخاب توصیف گرهای

مولکولی مناسب و ایجاد مدل های آماری از جمله رگرسیون مورد استفاده قرار گرفت. برنامه ی شبکه

عصبی مصنوعی در نرم افزار MATLAB 7.5 [۵۶] نوشته شده است. MATLAB یک زبان برنامه‌نویسی مناسب برای اهداف علمی و مهندسی است. ویژگی مهم این نرم‌افزار قابلیت‌های بالای آن برای کار با بردار و ماتریس هاست. تمام محاسبات با استفاده از یک کامپیوتر پنتیوم-۴ با سیستم عامل ویندوز XP انجام شده است.

۲-۱-۳- انجام محاسبات و مدل‌سازی

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

دومین مرحله در مطالعات QSAR پس از جمع‌آوری داده‌ها، رسم و بهینه‌سازی ساختار مولکول‌های مورد بررسی می‌باشد. بدین منظور ساختار شیمیایی مشتقات با استفاده از نرم‌افزار Hyperchem 7 رسم شد، سپس ساختار نهایی به روش نیمه تجربی و با استفاده از روش AM1 بهینه گردید و بهینه‌سازی تا زمانی ادامه یافت که گرادیان به $0/001$ کیلوکالری برمول برسد.

۲-۱-۳-۲- محاسبه‌ی توصیف‌گرهای مولکولی

برای توصیف مشخصات مختلف ترکیبات آلی از توصیف‌کننده‌های کمی استفاده می‌شود. توصیف‌کننده‌های مولکولی مقادیر عددی هستند که ویژگی‌های ساختاری یا شکل مولکول را توصیف می‌کنند. گروه‌های مختلفی از این توصیف‌گرها تاکنون معرفی شده‌اند که هر کدام جنبه‌های خاصی از ترکیبات را بررسی و توصیف می‌کنند. نرم‌افزار Dragon تعداد زیادی از این توصیف‌گرها را برای ترکیبات محاسبه می‌کند. ورودی این نرم‌افزار ساختار سه بعدی ترکیبات آلی در حالت بهینه است و خروجی آن شامل ۱۴۸۱ توصیف‌کننده در ۱۸ گروه مختلف است.

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

شماره	X	Y	$\log(1/IC_{50}) (M)$
۱	H	H	۶/۴۶
۲	3-F	H	۷/۲۵
۳	3-Cl	H	۷/۶۴
۴	3-Br	H	۷/۵۷
۵	3-I	H	۷/۱۰
۶	3-CF ₃	H	۶/۲۴
۷	H	6-OMe	۷/۲۶
۸	3-Br	6-OMe	۷/۵۲
۹	H	6-NH ₂	۶/۱۱
۱۰	3-CF ₃	6-NH ₂	۶/۲۴
۱۱	3-Br	6-NH ₂	۹/۱۱
۱۲	H	6-NO ₂	۵/۳۰
۱۳	3-Br	6-NO ₂	۶/۰۵
۱۴	H	7-OMe	۶/۹۲
۱۵	3-Br	7-OMe	۸/۰۰
۱۶	H	7-NH ₂	۷/۰۰
۱۷	3-F	7-NH ₂	۸/۷۰
۱۸	3-Cl	7-NH ₂	۹/۶۰
۱۹	3-Br	7-NH ₂	۱۰/۰۰
۲۰	3-I	7-NH ₂	۹/۴۶
۲۱	3-CF ₃	7-NH ₂	۸/۴۸
۲۲	H	6-NO ₂	۴/۹۲
۲۳	3-F	6-NO ₂	۵/۲۲
۲۴	3-Cl	6-NO ₂	۶/۰۹
۲۵	3-Br	6-NO ₂	۶/۰۰
۲۶	3-I	6-NO ₂	۶/۲۷
۲۷	H	6,7-Di-OMe	۷/۵۴
۲۸	3-F	6,7-Di-OMe	۸/۴۲
۲۹	3-Cl	6,7-Di-OMe	۹/۵۱
۳۰	3-Br	6,7-Di-OMe	۱۰/۶۰

ادامه‌ی جدول (۱-۲)

شماره	X	Y	$\log(1/IC_{50})$ (M)
۳۱	3-I	6,7-Di-OMe	۹/۰۵
۳۲	3-CF ₃	6,7-Di-OMe	۹/۶۲
۳۳	3-Br	6-NHMe	۸/۴۰
۳۴	3-Br	6-NMe ₂	۷/۰۸
۳۵	3-Br	6-NHCOOMe	۷/۹۲
۳۶	3-Br	7-OH	۸/۳۳
۳۷	3-Br	7-NHCOMe	۷/۴۰
۳۸	3-Br	7-NHMe	۸/۱۶
۳۹	3-Br	H ₂ 7-NHC ₅	۷/۹۲
۴۰	3-Br	7-NMe ₂	۷/۹۶
۴۱	3-Br	6,7-Di-NH ₂	۹/۹۲
۴۲	3-Br	6-NH ₂ , 7-NHMe	۹/۱۶
۴۳	3-Br	, 7-NMe ₂ 6-NH ₂	۶/۸۰
۴۴	3-Br	, 7-OMe ₂ 6-NH	۸/۴۲
۴۵	3-Br	, 7-Cl ₂ 6-NH	۸/۱۹
۴۶	3-Br	, 7-NH ₂ 6-NO ₂	۷/۲۸
۴۷	3-Br	, 7-NHMe ₂ 6-NO	۷/۱۷
۴۸	3-Br	, 7-NMe ₂ 6-NO ₂	۵/۷۰
۴۹	3-Br	, 7-NHCOMe ₂ 6-NO	۷/۵۵
۵۰	3-Br	, 7-OMe ₂ 6-NO	۷/۸۲
۵۱	3-Br	, 7-Cl ₂ 6-NO	۷/۶۰
۵۲	3-Br	6,7-Di-OH	۹/۷۷
۵۳	3-Br	H ₂ 6,7-Di-OC ₅	۱۱/۲۲
۵۴	3-Br	H ₃ 6,7-Di-OC ₇	۹/۷۷
۵۵	3-Br	H ₄ 6,7-Di-OC ₉	۶/۹۸
۵۶	3-Br	5,6,7-Tri-OMe	۹/۱۷
۵۷	2-Br	6,7-Di-OMe	۶/۸۹
۵۸	4-Br	6,7-Di-OMe	۹/۰۲
۵۹	3,4-di-Br	6,7-Di-OMe	۱۰/۱۴
۶۰	3,5-di-Br	6,7-Di-OMe	۶/۹۵

۲-۱-۳- انتخاب توصیف‌گرهای مهم

از آنجایی که تعداد توصیف‌کننده‌های محاسبه شده با استفاده از نرم افزار Dragon بسیار زیاد می‌باشد و تمام این توصیف‌گرها با داده‌های $\log(1/IC_{50})$ (ارتباط ندارند. از میان این توصیف‌گرها، لازم است که توصیف‌گرهای مهم را استخراج شود. این کار را در سه مرحله پیاپی انجام شد:

الف- در مرحله‌ی اول تمام توصیف‌گرهایی که مقادیر یکسانی برای حداقل ۹۵ درصد از ترکیبات مورد بررسی داشتند از لیست توصیف‌گرها حذف شدند.

ب- در مرحله‌ی دوم همبستگی بین متغیرها محاسبه گردید و با توجه به اینکه نبایستی متغیرهای ورودی در روش رگرسیون خطی چندگانه، با یکدیگر همبستگی داشته باشند. لذا از میان جفت توصیف‌کننده‌هایی که ضریب همبستگی بزرگتر از 0.90 ($r^2 > 0.90$) داشتند، یکی بطور تصادفی حذف شد. این کار با استفاده از نرم‌افزار SPSS برای هر گروه از توصیف‌کننده‌ها انجام شد.

ج) برای انتخاب توصیف‌کننده‌های مهم که با داده‌های $\log(1/IC_{50})$ ارتباط دارند، روش رگرسیون مرحله‌به‌مرحله در نرم افزار SPSS برای تمام مولکول‌ها و تمام توصیف‌کننده‌های باقیمانده انجام گردید. در روش رگرسیون مرحله‌به‌مرحله متغیرها یکی پس از دیگری وارد مدل می‌شوند. در ابتدا متغیری وارد مدل می‌شود که بالاترین میزان همبستگی را با متغیر وابسته خواهد داشت. ولی با ورود هر متغیر جدید، کلیه متغیرهای موجود در معادله بررسی شده و اگر هر کدام از آنها سطح معنی‌داری خود را از دست داده باشد، قبل از ورود متغیر جدید از مدل خارج می‌شود. پس برای این منظور داده‌های $\log(1/IC_{50})$ (به عنوان متغیر وابسته و توصیف‌گرها به عنوان متغیر مستقل در نظر گرفته شد و از میان تعداد زیادی توصیف‌کننده فقط ۱۴ توصیف‌گر ارتباط معنی‌داری با داده‌های $\log(1/IC_{50})$ (مولکول‌های مورد بررسی داشتند. این توصیف‌گرها به همراه کلاس مربوطه در جدول (۲-۲) آمده است.

(۲-۲) - تعداد کل توصیف‌گرهای انتخاب شده بوسیله‌ی روش رگرسیون خطی چندگانه.

The letters in the end of descriptor symbols meaning: e: weighted by Sanderson electronegativities; p: wheighted by atomic polarizibilities; v: weighted by atomic Vander Waals volumes, and m: weighted by atomic masses; s: weighted by atomic electropological states.

تعداد	نشانه	کلاس	معنی
۱	MATS1e	2D autocorrelation	Moran autocorrelation-lag 1
۲	X5sol	Topological	Solvation connectivity index chi-5
۳	H8m	GETAWAY	H autocorrelation of lag 8
۴	GATS6e	2D autocorrelation	Geary autocorrelation-lag 6
۵	R8e	GETAWAY	H autocorrelation of lag 8
۶	BEHv3	BCUT	Highest eigenvalue n. 3 of Burden matrix
۷	H8p	GETAWAY	H autocorrelation of lag 8
۸	RDF105m	RDF	Radial Distribution Function-10.5
۹	RDFO95m	RDF	Radial Distribution Function-9.5
۱۰	E3m	WHIM	3rd component accessibility directional WHIM index
۱۱	H6m	GETAWAY	H autocorrelation of lag 8
۱۲	Mor29e	3D-MoRSE	3D-MoRSE-signal 8
۱۳	G1s	WHIM	1st component accessibility directional WHIM index
۱۴	F084	Atom-centred framents	F attached to C1(sp2)

۲-۱-۳-۴- روند ایجاد مدل‌های QSAR

برای ایجاد مدل‌های توضیح‌دهنده‌ی رابطه‌ی ساختار - فعالیت ابتدا ۱۰ مورد از ترکیبات

بطور تصادفی جدا شد. البته این جداسازی به نحوی انجام گرفت که این ۱۰ ترکیب کل طیف

تغییرات متغیر هدف $\log(1/IC_{50})$ را پوشش دهند. مجموعه‌ی این ترکیبات را با توصیف‌گرهای مربوط

به آن سری ارزیابی^۱ نامیدیم که ارزیابی خطای مدل و بهینه‌سازی آن بوسیله‌ی این سری تحقق یافت. این ترکیبات بر حسب شماره‌های مربوط در جدول (۱-۲) عبارتند از ۴، ۷، ۱۵، ۲۲، ۲۹، ۳۳، ۳۷، ۴۶، ۵۰ و ۵۸ سپس ۵ مورد از ترکیبات نیز بطور تصادفی به عنوان سری تست^۲ برگزیده شد که از ابتدا کاملاً کنار گذاشته شدند و در مدل‌سازی از آن استفاده نشد که عبارتند از ترکیبات ۱، ۱۲، ۱۷، ۴۰ و ۵۴. ۴۵ ترکیب باقیمانده نیز سری آموزش نامیده شد و از آن برای ایجاد مدل استفاده گردید.

۲-۱-۳-۵- مدل رگرسیون خطی چندگانه (MLR)

با استفاده از رگرسیون خطی چند گانه مرحله به مرحله از میان تعداد زیادی توصیف‌گر، ۱۴ توصیفگر (مدل) را معنی‌دار تشخیص داد، که برای انتخاب بهترین مدل از بین مدل‌های ایجاد شده، مدلی که تعداد توصیف‌گر کمتر، آماره‌ی F بزرگتر، خطای استاندارد کوچکتر، ضریب همبستگی بالاتر و قدرت پیش‌بینی بیشتری داشته باشد، انتخاب می‌شود. مقادیر فاکتور F ، ضریب همبستگی و خطای استاندارد برای مدل‌های ۲ تا ۷ در جدول (۳-۲) آمده است.

جدول (۳-۲) - مقادیر ضریب همبستگی (R^2)، آماره‌ی F و خطای استاندارد برای مدل‌های ۲ تا ۷

خطای استاندارد	F	R^2	مدل
۰/۹۴۷	۳۸/۳۲۷	۰/۵۷۴	۲
۰/۹۰۹	۲۹/۶۹۴	۰/۶۱۴	۳
۰/۸۳۱	۲۹/۶۵۳	۰/۶۸۳	۴
۰/۸۲۵	۲۴/۴۵۲	۰/۶۹۴	۵
۰/۶۶۷	۳۶/۰۴۰	۰/۸۰۳	۶
۰/۵۸۴	۴۲/۷۴۵	۰/۸۵۲	۷

^۱ - Validation Set

^۲ - Test Set

برای اندازه‌گیری میزان قدرت پیش‌بینی هر دسته توصیف‌گر بالا، ابتدا با ۲ توصیف‌گر اول انتخاب شده، بوسیله‌ی سری آموزش به روش اجباری^۱ مدل ساخته شد و بوسیله‌ی معادله‌ی آن، سری ارزیابی آزموده شد. در مرحله بعد با مدل ایجاد شده توسط ۳ توصیف‌گر، سری ارزیابی را آزموده و همین‌طور با معادلات حاصل از ۴، ۵، ۶ و ۷ توصیف‌گر بطور مجزا سری ارزیابی مورد بررسی قرار گرفت. معادلات ایجاد شده بصورت زیر می‌باشند:

$$2 \text{ توصیف‌گر } \quad \log(1/IC_{50}) = 2/0.20 - 8/0.60 \text{ MATS1e} + 0/986 \text{ X5sol}$$

$$3 \text{ توصیف‌گر } \quad \log(1/IC_{50}) = 1/0.65 - 8/0.45 \text{ MATS1e} + 1/199 \text{ X5sol} - 3/782 \text{ H8m}$$

$$4 \text{ توصیف‌گر } \quad \log(1/IC_{50}) = 1/284 - 8/962 \text{ MATS1e} + 1/443 \text{ X5sol} - 4/474 \text{ H8m} - 0/954 \text{ GATS6e}$$

$$5 \text{ توصیف‌گر } \quad \log(1/IC_{50}) = 1/637 - 8/488 \text{ MATS1e} + 1/272 \text{ X5sol} - 4/161 \text{ H8m} - 1/133 \text{ GATS6e} \\ + 1/658 \text{ R8e}$$

$$6 \text{ توصیف‌گر } \quad \log(1/IC_{50}) = 65/927 - 6/271 \text{ MATS1e} + 2/0.61 \text{ X5sol} - 6/124 \text{ H8m} - 1/230 \text{ GATS6e} \\ + 4/353 \text{ R8e} - 21/464 \text{ BEHv3}$$

$$7 \text{ توصیف‌گر } \quad \log(1/IC_{50}) = 67/876 - 6/782 \text{ MATS1e} + 1/937 \text{ X5sol} - 8/760 \text{ H8m} - 1/148 \text{ GATS6e} \\ + 3/845 \text{ R8e} - 21/464 \text{ BEHv3} - 8/919 \text{ H8p}$$

میانگین مربعات خطا (MSE) و ضریب همبستگی (R^2) محاسبه شده برای سری ارزیابی، در جدول (۴-۲) آمده است. مطابق جدول (۳-۲) و (۴-۲)، مدل ۷ که کمترین خطای استاندارد و بیشترین مقدار آماره‌ی F و همچنین کمترین خطا را در سری ارزیابی ایجاد نموده، انتخاب گردید. نتایج استفاده از این معادله برای هر ترکیب سری ارزیابی و سری تست بترتیب در جدول (۵-۲) و جدول (۶-۲)، همچنین نتایج حاصل از ارزیابی مدل با استفاده از روش رد تک تک در جدول (۷-۲) نشان داده شده است. در روش رد مرحله‌ای تک تک به تعداد موارد موجود در مجموعه‌ی داده‌ها مدل

¹ - Enter

ایجاد می‌شود. برای این کار یکی از موارد مجموعه‌ی داده‌ها به عنوان مورد آزمایشی^۱ کنار گذاشته می‌شود. آن گاه مدل با استفاده از بقیه‌ی مجموعه‌ی ایجاد شده، بوسیله‌ی مورد آزمایشی آزموده می‌گردد. سپس همین کار با یک یک موارد مجموعه تکرار می‌شود تا جایی که به تعداد همه‌ی اجزای مجموعه یک مدل وجود داشته باشد و همه با مدل مربوط به خود آزموده شوند. شکل (۲-۲)، (۲-۳) و (۲-۴) نیز نمودارهای پراکندگی^۲ این سه مجموعه را نشان می‌دهند. مقادیر R^2 و MSE برای سری تست به ترتیب، ۰/۳۳۸۰ و ۰/۹۶۴۹ و برای رد تک تک به ترتیب، ۰/۴۰۱۹ و ۰/۸۰۰۹ می‌باشند.

جدول (۲-۴) - مقادیر میانگین مربعات خطا و ضریب همبستگی (R^2) محاسبه شده برای سری ارزیابی برای مدل‌های (توصیف‌گرهای) ۲ تا ۷

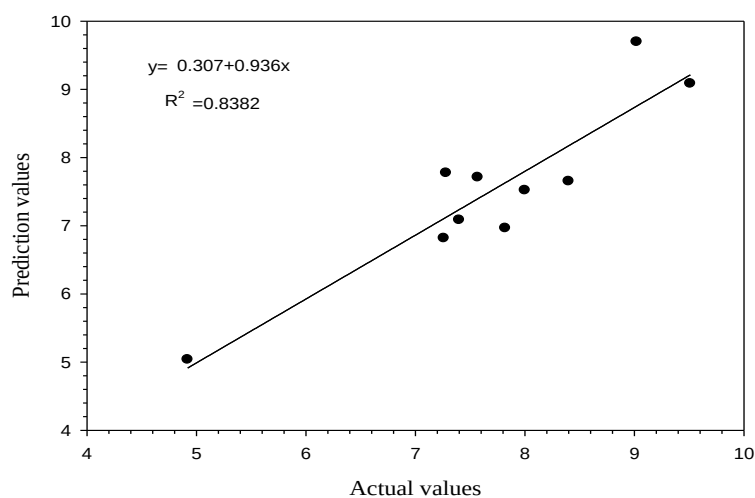
توصیف‌گر	MSE	R^2
۲	۰/۳۷۰۷	۰/۷۸۳۲
۳	۰/۳۸۰۵	۰/۷۷۸۷
۴	۰/۴۹۱۹	۰/۶۸۲۴
۵	۰/۵۱۵۷	۰/۶۷۳۷
۶	۰/۴۱۴۰	۰/۷۶۱۸
۷	۰/۲۷۳۱	۰/۸۳۸۲

جدول (۲-۵) - مقادیر مشاهده شده و پیش‌بینی شده‌ی PIC₅₀ برای ترکیبات مجموعه‌ی سری ارزیابی با معادله‌ی رگرسیون خطی چند گانه (۷ توصیف‌گر)

شماره	مقدار واقعی	مقدار پیش‌بینی	درصد خطا
۴	۷/۵۷	۷/۷۱	-۱/۸۵
۷	۷/۲۶	۶/۸۲	۶/۰۶
۱۵	۸	۷/۵۲	۶/۰۰
۲۲	۴/۹۲	۵/۰۴	-۲/۴۴
۲۹	۹/۵۱	۹/۰۹	۴/۴۲
۳۳	۸/۴	۷/۶۵	۸/۹۳
۳۷	۷/۴	۷/۰۹	۴/۱۹
۴۶	۷/۲۸	۷/۷۷	-۶/۷۳
۵۰	۷/۸۲	۶/۹۶	۱۰/۹۹
۵۸	۹/۰۲	۹/۶۹	-۷/۴۳

^۱ - Training Case

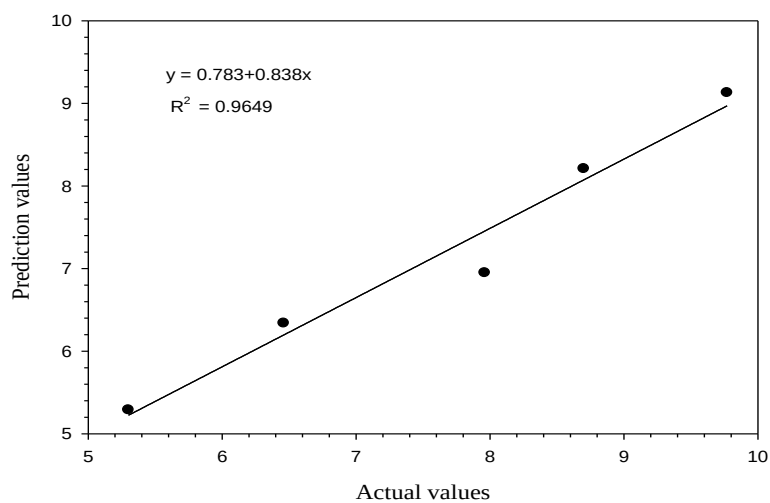
^۲ - Scatter Diagram



شکل (۲-۲) - نمودار پراکندگی مقادیر پیش‌بینی شده‌ی PIC₅₀ با مدل رگرسیون خطی چند گانه بر حسب مقادیر واقعی در سری ارزیابی

جدول (۲-۶) - نتایج حاصل از ارزیابی مدل رگرسیون خطی چند گانه با استفاده از سری تست

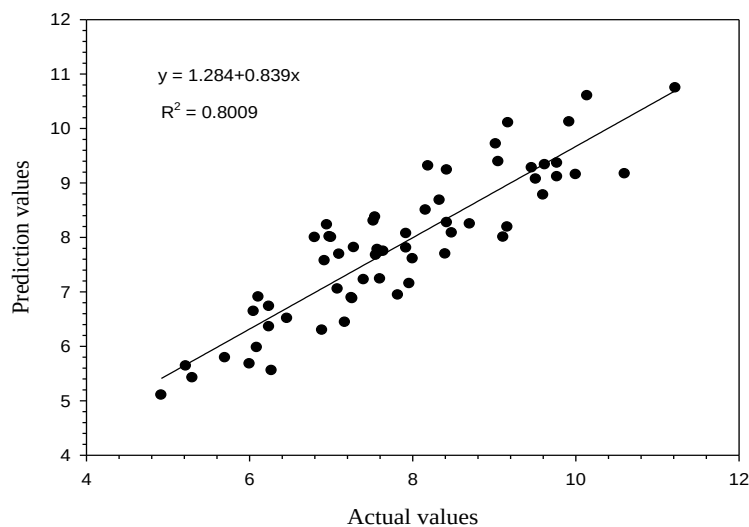
شماره	مقدار واقعی	مقدار پیش‌بینی	درصد خطا
۱	۶/۴۶	۶/۳۴	۱/۸۶
۱۲	۵/۳۰	۵/۲۹	۰/۱۹
۱۷	۸/۷۰	۸/۲۱	۵/۶۳
۴۰	۷/۹۶	۶/۹۵	۱۲/۶۹
۵۴	۹/۷۷	۹/۱۳	۶/۵۵



شکل (۲-۳) - نمودار پراکندگی مقادیر پیش‌بینی شده‌ی PIC₅₀ با مدل رگرسیون خطی چند گانه بر حسب مقادیر واقعی در سری تست

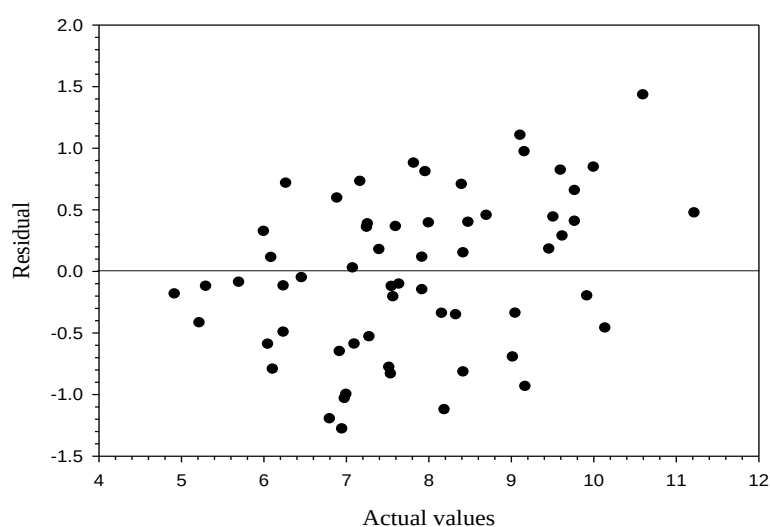
جدول (۷-۲) - نتایج حاصل از ارزیابی مدل رگرسیون خطی چندگانه با استفاده از روش رد تک تک

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



شکل (۴-۲) - نمودار پراکندگی مقادیر پیش بینی شده PIC₅₀ با مدل رگرسیون خطی چندگانه بر حسب مقادیر واقعی برای کل داده‌ها

به منظور بررسی وجود خطای معین در این مدل، باقیمانده مقادیر پیش‌بینی شده توسط مدل رگرسیون خطی چندگانه از مقادیر واقعی بر حسب مقادیر تجربی برای کل مولکول‌ها در شکل (۵-۲) رسم شده است. توزیع نسبتاً یکنواخت و اتفاقی باقی مانده‌ها در دو طرف خط صفر نشان می‌دهد که خطای معین در مدل وجود ندارد.



شکل (۵-۲) - نمودار باقیمانده‌ها برای کل داده‌ها (باقیمانده‌ی مقادیر پیش‌بینی شده‌ی PIC₅₀ توسط مدل رگرسیون خطی چندگانه از مقادیر واقعی بر حسب مقادیر تجربی برای کل مولکول‌ها)

۲-۱-۳-۵-۱- ارزیابی مدل رگرسیون خطی چندگانه‌ی ارائه شده با استفاده از پارامترهای

آماري

برای بررسی میزان کارایی مدل رگرسیون خطی چند گانه ارائه شده چهار پارامتر آماری زیر

محاسبه شد و نتایج حاصل در جدول (۸-۲) آمده است:

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

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

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

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

که y_i مقدار واقعی بازدارندگی ترکیب i ، $\hat{y}_i$ مقدار پیش بینی شده‌ی بازدارندگی ترکیب i ، $\bar{y}$ میانگین

کل مقادیر واقعی بازدارندگی و n تعداد نمونه‌های پیش بینی شده می‌باشند.

جدول (۸-۲)- پارامترهای آماری محاسبه شده برای مدل رگرسیون خطی چندگانه

پارامتر	سری ارزیابی (n = 10)	سری تست (n = 5)	کل داده‌ها (n = 60)
MAE	۵/۹۲	۵/۳۸	۶/۸۲
REP	۶/۷۷	۷/۶۱	۸/۰۹
MSE	۰/۲۷۳۱	۰/۳۳۸۰	۰/۴۰۱۹
R ²	۰/۸۳۸۲	۰/۹۶۴۹	۰/۸۰۰۹

۲-۱-۳-۶- مدل شبکه عصبی مصنوعی و بهینه‌سازی آن

در این تحقیق هر چند QSAR با استفاده از مدل خطی چندگانه نتایج خوبی ارائه کرد و مدل خطی قادر است که داده‌های مهارکنندگی را برای بازدارنده‌های تیروزین کیناز به راحتی پیش‌بینی نماید اما برای بهتر شدن نتایج از یک مدل غیر خطی نیز استفاده شده است. برای ایجاد مدل غیرخطی که ارتباط مناسب بین ساختار و فعالیت دارویی ترکیبات مورد نظر فراهم کند، شبکه‌ی عصبی مصنوعی انتخاب گردید. شبکه‌ی انتخاب شده یک شبکه عصبی پیش‌خور با الگوی آموزش پس‌انتشار می‌باشد. برای آموزش موفق شبکه و در نتیجه قدرت پیش‌بینی شبکه باید عوامل موثر در آموزش شبکه عصبی مصنوعی بررسی و بهینه شود. پارامترهای موثر و بهینه شده عبارتند از:

۱- تعداد لایه‌های پنهان

۲- نوع الگوریتم آموزشی

۳- تعداد متغیرهای ورودی شبکه

۴- تعداد نرون‌ها در لایه پنهان

۵- نوع تابع تبدیل

۶- سرعت آموزش

۷- تعداد دورهای آموزش

همانطور که گفته شد مشتقات ۴- (۳-برموانیلینو)-۶ و ۷ دی متوکسی کینازولین بطور تصادفی به ۳ گروه، سری آموزش، سری ارزیابی و سری تست تقسیم شدند. معیار در نظر گرفته شده در بهینه‌سازی به حداقل رساندن میانگین مربعات خطا (MSE) حاصله در سری ارزیابی می‌باشد. هنگامی که آموزش تکمیل شد، شبکه برای قابلیت‌های تعمیم یافتن و یادگیری و همچنین برای جلوگیری از خطای برازش اضافی تست می‌شود. تست برای قابلیت یادگیری شبکه توسط آزمایش برای تولید خروجی برای مجموعه‌ای از ورودی‌ها که در مرحله‌ی آموزش استفاده نشده‌اند، انجام می‌گیرد. زیرا در مرحله‌ی آموزش این امکان وجود دارد که شبکه الگوها را حفظ کند و برای مرحله

تست جواب درستی ندهد. در اینجا ۵ ترکیب برای سری تست استفاده شد و مدل انتخاب شده برای ارزیابی با استفاده از این سری آزمایش گردید.

۲-۱-۳-۶-۱- انتخاب تعداد لایه‌های شبکه عصبی

متأسفانه هیچ قاعده‌ی تئوری یا تجربی که بتوان بوسیله‌ی آن تعداد لایه‌های پنهان در یک شبکه را تعیین نمود، وجود ندارد. نتایج حاصل از مطالعات مختلف نشان می‌دهد که در بکارگیری شبکه عصبی در شیمی، انتخاب یک لایه پنهان نتایج رضایت‌بخشی را ایجاد می‌کند [۱]. لذا در این تحقیق شبکه عصبی سه لایه متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه خروجی بکار گرفته شد.

۲-۱-۳-۶-۲- انتخاب الگوریتم آموزشی، تعداد متغیرهای ورودی شبکه، انتخاب نوع تابع

تبدیل و تعداد گره‌های لایه پنهان

شبکه‌ی عصبی مصنوعی با استفاده از نرم افزار MATLAB طراحی و بهینه شد. پس از طراحی اولیه شبکه، برای اینکه شبکه قادر به پیش‌بینی صحیحی باشد، باید آموزش ببیند. الگوریتم‌های آموزشی مختلفی در جعبه ابزار^۱ نرم‌افزار مطلب وجود دارد. در این تحقیق دوالگوریتم آموزشی لونبرگ - مارکورت و تنظیم بایزین مورد بررسی قرار گرفت. با توجه به اینکه خروجی‌های شبکه عصبی داده‌های کمی $\log(1/IC_{50})$ (می‌باشند، لذا تابع انتقال خطی برای لایه‌ی خروجی انتخاب گردید.

1- toolbox

مطالعات اولیه نشان داد که تغییر در سرعت آموزش در ناحیه‌ی $0/1 - 0/01$ ، با استفاده از تعداد گره‌های متفاوت در لایه‌ی پنهان تاثیر چندانی روی میانگین مربعات خطای سری داده‌های ارزیابی ندارد. بنابراین سرعت آموزش $0/01$ برای آموزش شبکه عصبی مورد استفاده قرار گرفت.

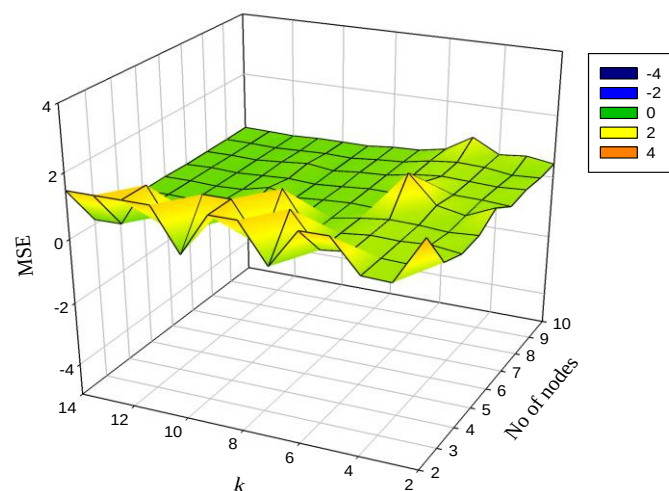
برای بهینه سازی و شناسایی بهترین پارامترهای شبکه، شبکه‌هایی با ورودی‌های از ۲ تا ۱۴ توصیف‌گر ایجاد شدند. هر شبکه با الگوریتم آموزشی لونیبرگ - مارکورت و تنظیم بایزین با تعداد گره‌های متفاوت در لایه پنهان آموزش داده شد. در حالی که برای بدست آوردن بهترین تابع انتقال در لایه‌ی پنهان مدل‌های شبکه عصبی، از توابع لگاریتم زیگموئید و تانژانت هیپربولیک به عنوان تابع فعالیت استفاده گردید. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطا به عنوان معیار انتخاب گردید. در تمام شبکه‌ها تعداد دوره‌های آموزشی یکسان و برابر ۱۰ در نظر گرفته شد. شبکه‌های آموزش یافته برای پیش‌بینی داده‌های $\log(1/IC_{50})$ (سری ارزیابی مورد استفاده قرار گرفتند. به این ترتیب در مجموع ۴۶۸ شبکه‌ی سه لایه آموزش داده شد و سپس برای سری پیش بینی مورد استفاده قرار گرفتند که نتایج حاصله در جداول (۹-۲)، (۱۰-۲)، (۱۱-۲) و (۱۲-۲) آمده است. نمودار سه بعدی این جداول نیز در شکل‌های (۶-۲)، (۷-۲)، (۸-۲) و (۹-۲) رسم شده است. با توجه به نتایج برای هر دو تابع تبدیل لگاریتم زیگموئید و تانژانت هیپربولیک، الگوریتم آموزشی تنظیم بایزین نسبت به الگوریتم آموزشی لونیبرگ - مارکورت دارای کمترین مقدار میانگین مربعات خطا می‌باشد. این الگوریتم علاوه بر سرعت آموزش بسیار زیاد دارای قدرت تعمیم پذیری بالا می‌باشد، به این معنی که قادر است میان ترکیبات که در آموزش استفاده شده و آن‌هایی که در آموزش استفاده نشده، تطابق خوبی برقرار نماید.

همانگونه که در جداول (۹-۲) و (۱۰-۲) مشاهده می‌شود؛ شبکه‌ها با الگوریتم آموزشی تنظیم بایزین و با استفاده از هر دو تابع تبدیل لگاریتم زیگموئید و تانژانت هیپربولیک در تعداد متغیرهای ۹ و بیشتر و تعداد گره‌های ۴ و بیشتر مقادیر میانگین مربعات خطای کمی در سری ارزیابی نشان می‌دهند. با توجه به اینکه هر چه تعداد گره‌ها و ورودی‌های شبکه کمتر باشد، شبکه ساده‌تر

شده و سرعت آموزش، قدرت یادگیری و تعمیم آن افزایش می‌یابد، لذا تعداد گره ۴ با تعداد ورودی‌های ۹ و تابع تبدیل لگاریتم زیگموئید به عنوان شبکه‌ی بهینه انتخاب گردید. این توصیف‌کننده‌ها به همراه کلاس مربوطه در جدول (۲-۱۳) آمده است. ماتریس همبستگی بین توصیف‌گرها نیز در جدول (۲-۱۴) نشان داده شده است، که مشاهده می‌شود که بین توصیف‌گرها همبستگی وجود ندارد.

جدول (۲-۹) - مقادیر میانگین مربعات خطا برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین (trainbr) و تابع تبدیل لگاریتم زیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

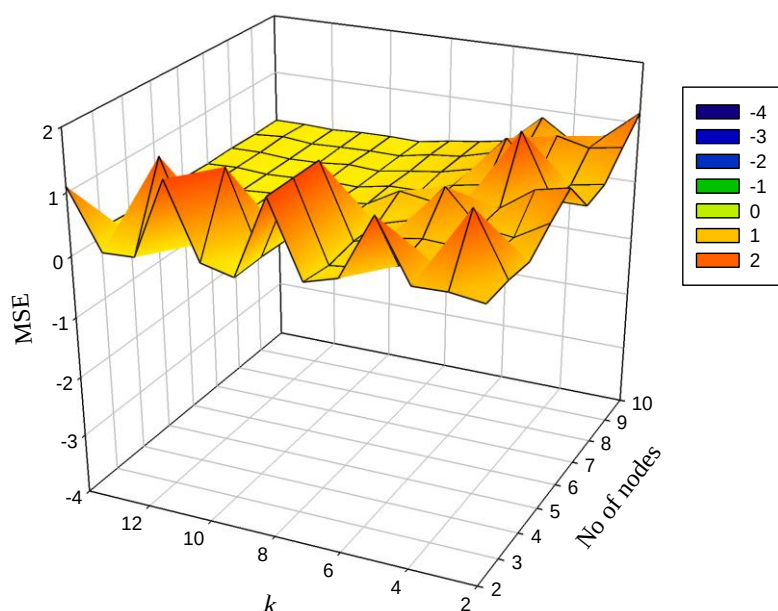
		تعداد متغیرهای ورودی شبکه (تعداد توصیف‌گرها) (k)												
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
تعداد گره‌ها	۲	۱/۸۸۴	۰/۴۶۲	۰/۴۹۷	۱/۴۷۱	۱/۴۶۴	۰/۲۸۰	۱/۴۷۱	۱/۴۷۱	۰/۱۵۴	۱/۴۷۱	۱/۴۷۱	۱/۴۷۱	۱/۴۷۱
	۳	۰/۶۸۶	۰/۴۵۰	۰/۴۸۴	۰/۵۲۵	۰/۴۱۹	۱/۴۶۴	۰/۲۲۷	۱/۴۷۱	۱/۴۷۱	۰/۱۵۷	۱/۴۷۱	۰/۱۴۶	۰/۱۳۱
	۴	۰/۴۱۹	۰/۵۳۵	۰/۴۸۶	۰/۵۱۵	۰/۴۲۳	۰/۲۶۰	۱/۴۷۱	۰/۱۳۶	۰/۲۳۶	۰/۱۵۴	۰/۱۲۲	۰/۱۴۹	۰/۱۲۴
	۵	۰/۴۷۱	۰/۴۱۶	۰/۴۸۳	۰/۵۰۷	۰/۴۰۳	۰/۲۵۹	۰/۲۲۱	۰/۱۴۶	۰/۱۵۰	۰/۱۹۰	۰/۱۱۲	۰/۱۴۷	۰/۱۳۰
	۶	۰/۸۳۱	۰/۴۴۶	۰/۶۱۶	۱/۴۷۱	۰/۲۲۵	۰/۲۶۳	۰/۲۱۹	۰/۱۵۰	۰/۳۳۶	۰/۱۳۶	۰/۱۰۴	۰/۱۴۲	۰/۱۴۱
	۷	۰/۴۰۶	۰/۴۶۸	۰/۴۸۷	۰/۷۷۱	۰/۲۴۸	۰/۲۶۴	۰/۱۹۴	۰/۱۳۷	۰/۱۸۷	۰/۱۴۴	۰/۱۰۳	۰/۱۴۲	۰/۱۱۸
	۸	۰/۴۷۷	۰/۴۲۸	۰/۴۶۸	۰/۵۰۳	۰/۴۲۰	۰/۲۶۷	۰/۲۲۵	۰/۱۳۹	۰/۲۴۹	۰/۱۶۹	۰/۱۰۴	۰/۱۳۲	۰/۱۲۹
	۹	۰/۴۱۱	۰/۳۶۶	۰/۴۸۹	۰/۵۲۴	۰/۴۳۶	۰/۲۷۲	۰/۲۱۷	۰/۱۳۸	۰/۱۶۱	۰/۱۳۱	۰/۱۰۰	۰/۱۲۷	۰/۱۳۶
	۱۰	۰/۴۳۷	۰/۵۶۶	۰/۴۸۳	۰/۹۳۳	۰/۴۱۴	۰/۲۵۰	۰/۲۲۲	۰/۱۳۷	۰/۱۴۴	۰/۱۴۲	۰/۱۰۷	۰/۱۳۵	۰/۱۵۶



شکل (۲-۶) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم زیگموئید

جدول (۲-۱۰) - مقادیر میانگین مربعات خطا برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین (trainbr) و تابع تبدیل تانژانت هیپربولیک (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

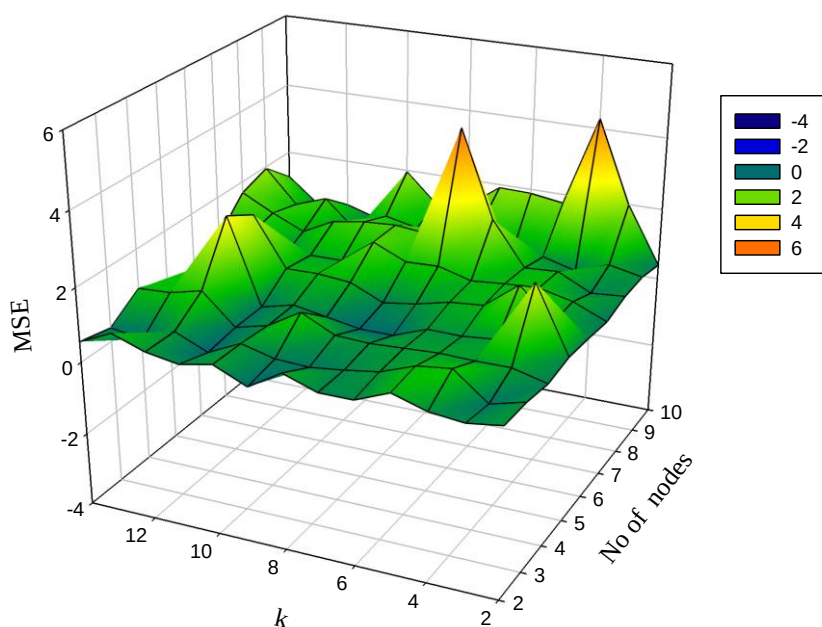
تعداد متغیرهای ورودی شبکه (تعداد توصیف گر ها)														
	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	
تعداد گر ها	۲	۰/۴۶۲	۰/۵۳۳	۰/۵۱۸	۱/۴۶۴	۰/۴۳۶	۰/۲۸۱	۱/۴۶۴	۰/۱۵۲	۰/۲۸۱	۰/۱۳۰	۱/۴۶۴	۰/۱۷۶	۰/۱۵۲
	۳	۱/۱۰۲	۰/۴۸۶	۱/۴۶۴	۰/۵۳۳	۰/۴۱۰	۰/۲۸۱	۰/۲۲۰	۱/۴۶۴	۰/۱۶۰	۱/۴۶۴	۱/۴۶۴	۰/۱۳۷	۰/۱۴۷
	۴	۰/۴۴۹	۰/۴۶۲	۰/۴۹۰	۰/۵۱۸	۰/۴۳۶	۰/۲۹۲	۱/۴۶۴	۰/۱۳۷	۰/۱۷۶	۰/۱۳۷	۰/۱۰۹	۰/۱۱۶	۰/۱۲۲
	۵	۰/۷۴۰	۰/۴۲۲	۰/۴۹۰	۱/۰۴۰	۰/۴۲۲	۰/۲۷۰	۰/۲۴۰	۰/۱۶۰	۰/۱۶۸	۰/۱۳۰	۰/۱۰۹	۰/۱۴۴	۰/۱۳۰
	۶	۱/۰۰۰	۰/۸۸۴	۰/۵۰۴	۰/۵۳۳	۰/۴۳۶	۰/۵۱۸	۰/۲۲۱	۰/۱۷۶	۰/۱۶۰	۰/۱۳۷	۰/۱۰۲	۰/۱۳۰	۰/۱۳۰
	۷	۰/۴۳۶	۰/۴۳۶	۱/۴۶۴	۰/۵۱۸	۰/۴۲۲	۰/۲۴۵	۰/۲۴۰	۰/۱۳۷	۰/۱۵۲	۰/۱۲۲	۰/۱۰۱	۰/۱۳۷	۰/۱۱۶
	۸	۰/۵۰۴	۰/۴۱۰	۰/۴۹۰	۱/۰۲۰	۰/۴۲۲	۰/۲۹۲	۰/۲۲۱	۰/۱۳۷	۰/۱۶۸	۰/۱۱۶	۰/۱۰۲	۰/۱۳۷	۰/۱۳۰
	۹	۰/۸۴۶	۰/۷۵۷	۰/۸۸۳	۰/۵۴۸	۰/۴۳۶	۰/۲۹۲	۰/۲۴۰	۰/۱۳۷	۰/۱۴۴	۰/۱۳۰	۰/۱۰۹	۰/۱۶۰	۰/۱۱۹
	۱۰	۱/۱۶۶	۰/۷۰۶	۰/۵۱۸	۰/۸۶۵	۰/۳۸۴	۰/۲۹۲	۰/۲۲۱	۰/۱۳۰	۰/۱۴۴	۰/۱۳۷	۰/۱۰۹	۰/۱۳۰	۰/۱۳۰



شکل (۲-۷) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت هیپربولیک

جدول (۲-۱۱) - مقادیر میانگین مربعات خطا برای شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکورت (trainlm) و تابع تبدیل لگاریتم زیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

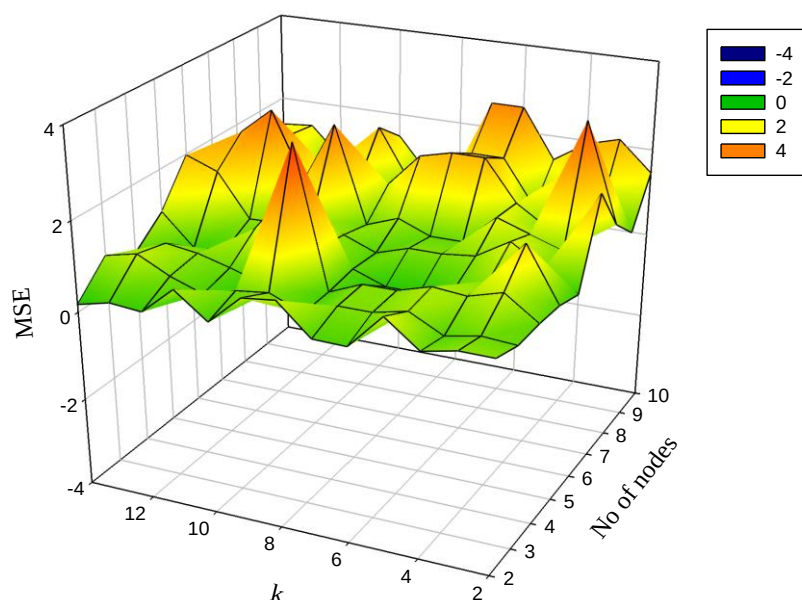
		تعداد متغیرهای ورودی شبکه (تعداد توصیف گرها)												
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
تعداد گره‌ها	۲	۰/۶۲۵	۰/۴۹۰	۰/۵۸۷	۰/۸۷۴	۰/۴۹۶	۰/۸۱۸	۰/۲۶۹	۰/۶۷۴	۰/۴۹۴	۰/۶۴۳	۰/۹۶۲	۰/۵۶۱	۱/۲۰۰
	۳	۰/۵۹۳	۰/۳۴۸	۱/۰۷۹	۰/۹۶۲	۰/۶۲۲	۰/۴۸۴	۰/۳۷۴	۰/۴۱۴	۰/۶۵۰	۰/۲۲۵	۰/۶۲۰	۰/۲۳۶	۰/۳۷۰
	۴	۰/۴۵۵	۰/۶۸۹	۰/۶۸۵	۰/۷۱۰	۰/۶۶۲	۰/۶۱۲	۰/۶۵۶	۱/۱۰۲	۰/۵۷۱	۰/۳۴۲	۱/۱۸۳	۰/۹۶۵	۰/۹۱۸
	۵	۰/۵۶۳	۲/۳۷۲	۰/۶۸۱	۰/۵۹۰	۰/۶۰۴	۰/۲۵۹	۰/۲۶۴	۰/۴۲۱	۰/۵۸۷	۰/۳۵۹	۲/۷۲۶	۱/۱۵۳	۰/۶۱۱
	۶	۰/۴۹۰	۰/۷۷۴	۱/۵۹۳	۰/۸۷۸	۰/۷۹۵	۰/۵۷۶	۰/۴۴۳	۰/۸۴۹	۰/۹۳۰	۰/۹۱۶	۲/۲۷۹	۰/۷۵۴	۰/۲۶۹
	۷	۰/۳۸۴	۰/۷۷۴	۰/۹۸۵	۰/۹۳۹	۰/۸۵۷	۱/۱۰۹	۱/۰۷۳	۱/۴۷۶	۰/۷۹۵	۰/۷۸۶	۰/۴۵۸	۱/۰۶۰	۰/۸۷۹
	۸	۰/۵۴۸	۰/۴۷۴	۰/۹۲۲	۰/۸۶۹	۲/۰۷۶	۴/۴۶۹	۱/۳۷۸	۱/۴۶۹	۱/۴۱۴	۱/۲۷۳	۰/۸۲۹	۱/۲۲۵	۱/۶۸۲
	۹	۰/۵۶۸	۰/۷۵۷	۰/۴۲۴	۰/۶۷۳	۰/۴۹۲	۱/۲۳۲	۱/۰۴۶	۰/۹۰۷	۱/۰۹۹	۰/۵۲۵	۱/۳۲۹	۱/۰۱۴	۱/۹۴۸
	۱۰	۰/۳۹۷	۱/۹۸۸	۴/۲۷۸	۱/۷۵۲	۱/۸۸۵	۱/۹۰۷	۱/۲۴۵	۱/۱۰۸	۱/۹۵۵	۰/۴۱۰	۰/۶۸۷	۰/۵۲۸	۱/۲۴۷



شکل (۲-۸) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکورت و تابع تبدیل لگاریتم زیگموئید

جدول (۲-۱۲) - مقادیر میانگین مربعات خطا برای شبکه‌هایی با الگوریتم آموزشی لونیگ - مارکورت (trainlm) و تابع تبدیل تانژانت هیپربولیک (tansig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

		تعداد متغیرهای ورودی شبکه (تعداد توصیف‌گر)												
		۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
تعداد گره‌ها	۲	۰/۷۳۹	۰/۷۴۵	۰/۵۸۸	۱/۱۶۶	۰/۳۹۲	۰/۴۰۹	۱/۰۸۹	۱/۰۰۲	۰/۳۴۴	۰/۹۱۰	۰/۲۸۴	۰/۳۴۷	۰/۱۷۲
	۳	۰/۵۲۹	۰/۵۸۳	۰/۴۳۹	۰/۸۲۴	۰/۶۶۰	۰/۶۶۰	۰/۳۶۴	۰/۶۷۴	۰/۵۸۴	۰/۳۷۷	۰/۶۳۶	۱/۰۰۲	۰/۸۴۰
	۴	۰/۵۲۸	۱/۰۵۴	۰/۸۴۱	۰/۹۱۱	۰/۶۹۴	۰/۶۹۴	۰/۳۸۸	۳/۴۹۵	۰/۳۰۳	۰/۵۳۴	۰/۷۳۲	۰/۸۳۳	۰/۴۸۸
	۵	۰/۴۶۸	۱/۶۸۷	۰/۷۴۸	۰/۴۰۷	۰/۲۰۴	۰/۲۰۴	۰/۴۷۹	۰/۸۴۶	۱/۹۲۱	۱/۰۱۹	۰/۳۸۵	۰/۳۲۳	۱/۰۳۵
	۶	۰/۲۷۳	۰/۷۲۸	۱/۱۹۰	۰/۷۴۳	۰/۶۶۱	۰/۶۶۱	۰/۴۳۱	۰/۷۱۹	۰/۹۳۳	۰/۷۴۵	۰/۷۲۲	۱/۷۰۲	۱/۹۷۷
	۷	۲/۱۰۵	۰/۷۰۴	۰/۹۵۶	۱/۲۱۳	۰/۵۳۲	۰/۵۳۲	۰/۸۲۹	۰/۵۶۱	۲/۷۵۳	۱/۳۵۹	۲/۸۷۶	۲/۲۸۰	۰/۹۱۵
	۸	۱/۰۲۱	۳/۲۳۴	۰/۶۸۵	۱/۱۵۷	۲/۰۸۸	۲/۰۸۸	۱/۹۰۷	۱/۱۳۶	۱/۰۰۶	۱/۲۷۲	۰/۳۲۳	۰/۳۲۲	۱/۳۰۶
	۹	۰/۴۴۳	۰/۷۸۱	۱/۶۶۱	۰/۶۸۷	۱/۷۲۳	۱/۷۲۳	۰/۵۰۲	۰/۸۳۹	۲/۰۱۸	۰/۵۱۲	۰/۹۷۷	۱/۵۲۵	۱/۰۱۴
	۱۰	۱/۴۶۰	۲/۱۳۴	۱/۹۲۵	۱/۴۴۴	۲/۵۷۱	۲/۵۷۱	۰/۷۸۹	۰/۶۱۹	۱/۴۷۴	۱/۲۸۱	۰/۷۱۴	۱/۴۲۱	۱/۳۵۸



شکل (۲-۹) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونیگ - مارکورت و تابع تبدیل تانژانت هیپربولیک

جدول (۲-۱۳) - توصیف گرهای بکار رفته در مدل سازی

معنی	کلاس	نشانه	تعداد
Moran autocorrelation-lag 1	2D autocorrelation	MATS1e	۱
Solvation connectivity index chi-5	Topological	X5sol	۲
H autocorrelation of lag 8	GETAWAY	H8m	۳
Geary autocorrelation-lag 6	2D autocorrelation	GATS6e	۴
R autocorrelation of lag 8	GETAWAY	R8e	۵
Highest eigenvalue n. 4 of Burden matrix	BCUT	BEHv3	۶
H autocorrelation of lag 8	GETAWAY	H8p	۷
Radial Distribution Function-10.5	RDF	RDF105m	۸
Radial Distribution Function-9.5	RDF	RDFO95m	۹

جدول (۲-۱۴) - ماتریس همبستگی برای توصیف گرهای انتخاب شده

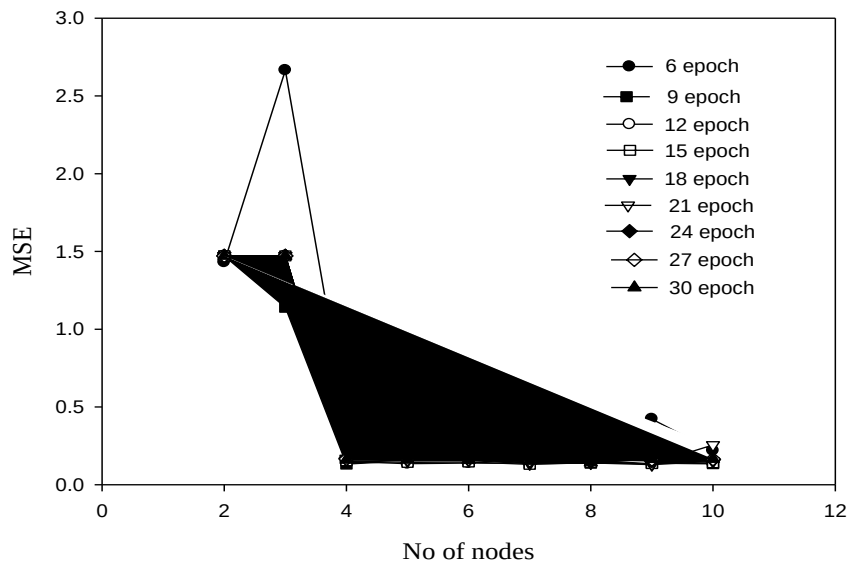
	MATS1e	X5sol	H8m	GATS6e	R8e	BEHv3	H8p	RDF105m	RDF105m
MATS1e	۱								
X5sol	۰/۱۴۹	۱							
H8m	۰/۰۵۹	۰/۳۶۵	۱						
GATS6e	-۰/۲۰۳	۰/۲۲۷	-۰/۰۲۳	۱					
R8e	-۰/۲۵۰	۰/۵۳۹	۰/۰۲۰	۰/۵۲۹	۱				
BEHv3	۰/۲۵۷	۰/۷۰۶	۰/۰۵۵	۰/۲۱۲	۰/۵۶۰	۱			
H8p	۰/۱۶۴	۰/۵۳۴	۰/۶۶۳	۰/۰۵۳	۰/۲۱۵	۰/۳۴۳	۱		
RDF105m	۰/۲۷۳	۰/۵۹۷	۰/۳۳۳	۰/۱۵۱	۰/۳۷۲	۰/۳۱۸	۰/۳۷۱	۱	
RDF105m	۰/۱۵۶	۰/۵۱۴	۰/۳۰۰	۰/۰۸۵	۰/۰۵۸	۰/۱۸۰	۰/۳۲۲	۰/۳۲۰	۱

۲-۱-۳-۶-۳- انتخاب تعداد دوره‌های آموزش

جهت بهینه کردن تعداد دوره‌های آموزشی نیز پارامتر میانگین مربعات خطا مورد بررسی قرار گرفت. به این منظور شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم زیگموئید و تعداد گره‌های لایه پنهان در محدوده ۱۰-۲ با فواصل یک و با تعداد دوره‌های آموزش مختلف با فواصل یکسان طراحی و آموزش داده شدند و شبکه‌های آموزش داده شده در پیش‌بینی داده‌های $\log(1/IC_{50})$ (سری ارزیابی بکار گرفته شدند. نتایج پیش‌بینی در جدول (۲-۱۵) و شکل (۲-۱۰) آمده است. بنابراین تعداد دوره‌های آموزشی برابر با ۹ که حداقل میانگین مربعات خطا را ایجاد می‌کند، انتخاب گردید.

جدول (۲-۱۵)- مقادیر میانگین مربعات خطا برای تعداد گره‌های مختلف در لایه پنهان

		تعداد دوره‌های آموزشی								
		۶	۹	۱۲	۱۵	۱۸	۲۱	۲۴	۲۷	۳۰
تعداد گره	۲	۱/۴۳۰۶	۱/۴۷۰۸	۱/۴۷۰۸	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹
	۳	۲/۶۶۵۴	۱/۱۴۰۷	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹	۱/۴۷۰۹
	۴	۰/۴۱۹۶	۰/۱۳۲۹	۰/۱۴۹۷	۰/۱۵۴۰	۰/۱۶۲۳	۰/۱۶۸۵	۰/۱۶۹۴	۰/۱۶۵۷	۰/۱۵۹۴
	۵	۰/۸۰۲۹	۰/۱۵۳۹	۰/۱۳۴۸	۰/۱۴۳۵	۰/۱۵۲۵	۰/۱۶۴۴	۰/۱۶۶۳	۰/۱۶۵۱	۰/۱۵۹۹
	۶	۰/۴۵۳۴	۰/۱۴۸۳	۰/۱۳۹۱	۰/۱۴۵۸	۰/۱۵۵۶	۰/۱۶۳۶	۰/۱۶۵۸	۰/۱۶۵۱	۰/۱۵۷۵
	۷	۰/۲۰۸۵	۰/۱۳۷۴	۰/۱۳۶۴	۰/۱۳۱۱	۰/۱۳۵۱	۰/۱۴۸۱	۰/۱۵۶۳	۰/۱۵۶۴	۰/۱۵۳۱
	۸	۰/۲۶۵۷	۰/۱۳۸۷	۰/۱۳۷۴	۰/۱۴۲۲	۰/۱۴۰۲	۰/۱۵۸۳	۰/۱۶۰۵	۰/۱۵۴۴	۰/۱۴۲۹
	۹	۰/۴۲۲۱	۰/۱۳۸۶	۰/۱۳۷۳	۰/۱۳۵۷	۰/۱۲۹۴	۰/۱۳۴۵	۰/۱۵۳۰	۰/۱۶۱۱	۰/۱۶۴۷
	۱۰	۰/۲۱۹۱	۰/۱۳۴۹	۰/۱۳۵۸	۰/۱۴۱۴	۰/۱۵۰۷	۰/۲۵۴۹	۰/۱۶۱۵	۰/۱۶۳۷	۰/۱۶۵۲



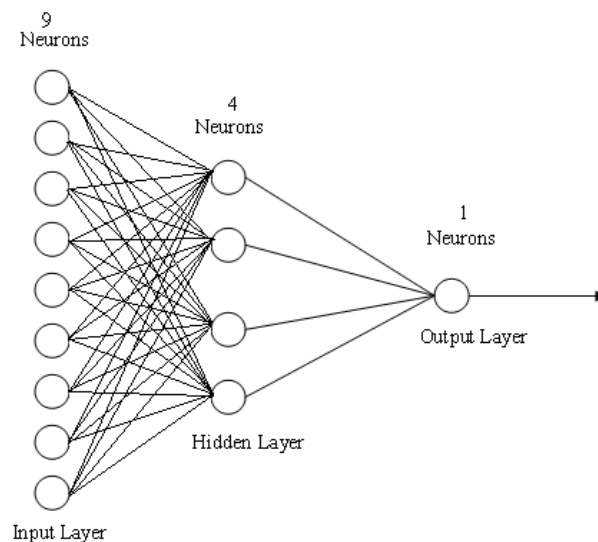
شکل (۲-۱۰) - تغییرات مقادیر میانگین مربعات خطا بر حسب تعداد گره‌های مختلف در لایه پنهان

۲-۱-۳-۶-۴ - هندسه‌ی شبکه عصبی مصنوعی بهینه شده

با توجه به نتایج حاصل در بهینه‌سازی شبکه، شبکه‌ی عصبی پیش‌خور سه لایه و آموزش

پس انتشار با الگوریتم تنظیم بایزین و تابع انتقال لگاریتم زیگموئید انتخاب گردید. ساختار هندسی

شبکه بهینه شده در شکل (۲-۱۱) و پارامترهای آن در جدول (۲-۱۶) آمده است.



شکل (۲-۱۱) - ساختار هندسی شبکه عصبی استفاده شده

جدول (۲-۱۶) - پارامترهای شبکه‌ی بهینه

۹	تعداد دوره‌های آموزش
e^{-10}	مینیمم شیب عملکرد
۰/۰۰۵	پارامتر تطبیق مارکورت (μ)
۰/۱	فاکتور کاهش برای پارامتر تطبیق مارکورت
۱۰	فاکتور افزایش برای پارامتر تطبیق مارکورت
$1e^{-10} \times$	مقدار ماکزیمم برای پارامتر تطبیق مارکورت

۲-۱-۳-۷- ارزیابی مدل شبکه عصبی مصنوعی ارائه شده

اساسی‌ترین مرحله در مطالعات رابطه کمی ساختار - فعالیت ارزیابی قدرت پیش‌بینی شبکه می‌باشد. قدرت پیش‌بینی شبکه 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 تحقیق برای پیش‌بینی داده‌های $\log(1/IC_{50})$ (مولکول‌های بازدارنده‌ی تیروزین کیناز، چندین روش بکار گرفته شد.

۲-۱-۳-۷-۱- ارزیابی شبکه با استفاده از سری ارزیابی

به این منظور، شبکه عصبی با ۴۵ مولکول سری ارزیابی آموزش داده شد. معیار برای بهینه‌سازی شبکه، کمترین میزان میانگین مربعات خطا در سری ارزیابی که شامل ۱۰ مولکول که بطور تصادفی انتخاب شده‌اند، می‌باشد. نتایج حاصله در جدول (۲-۱۷) نشان داده شده است. ترسیمی از مقادیر پیش‌بینی شده بر حسب مقدار واقعی نیز در شکل (۲-۱۲) نشان داده شده است.

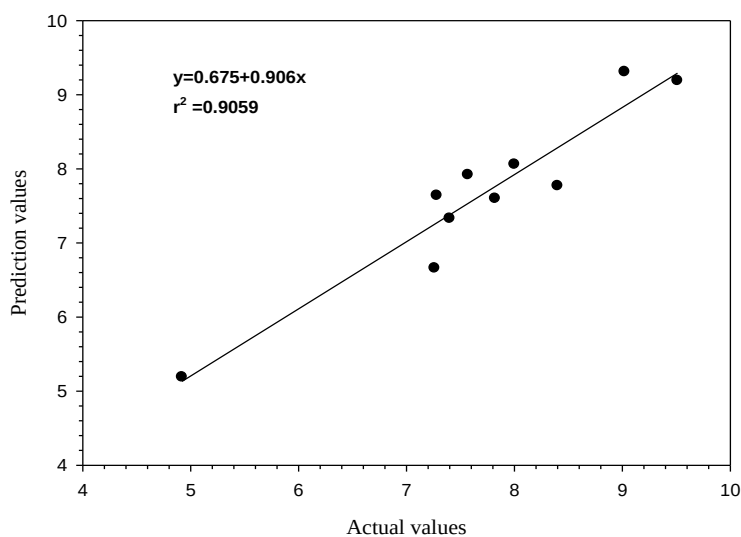
۲-۱-۳-۷-۲- ارزیابی شبکه با استفاده از سری تست

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

۵ داده که در آموزش استفاده نشده‌اند، مورد بررسی قرار گرفت. نتایج حاصله در جدول (۲-۱۸) و همچنین شکل (۲-۱۳) نشان داده شده است.

جدول (۲-۱۷) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری ارزیابی

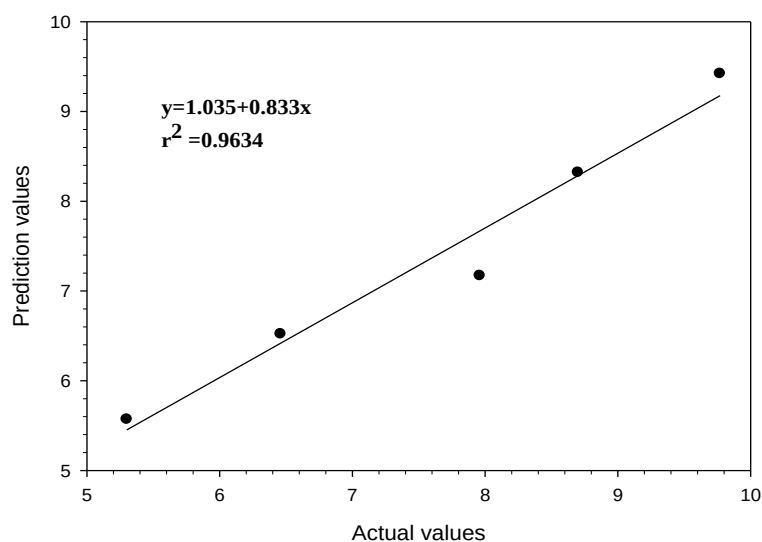
شماره	مقدار واقعی	مقدار پیش‌بینی	درصد خطا
۴	۷/۵۷	۷/۹۲	-۴/۶۲
۷	۷/۲۶	۶/۶۶	۸/۲۶
۱۵	۸	۸/۰۶	-۰/۷۵
۲۲	۴/۹۲	۵/۱۹	-۵/۴۹
۲۹	۹/۵۱	۹/۱۹	۳/۳۶
۳۳	۸/۴	۷/۷۷	۷/۵۰
۳۷	۷/۴	۷/۳۳	۰/۹۵
۴۶	۷/۲۸	۷/۶۴	-۴/۹۴
۵۰	۷/۸۲	۷/۶۰	۲/۸۱
۵۸	۹/۰۲	۹/۳۱	-۳/۲۱



شکل (۲-۱۲) - ترسیمی از مقادیر پیش‌بینی شده‌ی PIC₅₀ با شبکه عصبی مصنوعی بر حسب مقدار واقعی در سری ارزیابی

جدول (۲-۱۸) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری تست

شماره	مقدار واقعی	مقدار پیش بینی	درصد خطا
۱	۶/۴۶	۶/۵۲	-۰/۹۳
۱۲	۵/۳۰	۵/۵۷	-۵/۰۹
۱۷	۸/۷۰	۸/۳۲	۴/۳۷
۴۰	۷/۹۶	۷/۱۷	۹/۹۲
۵۴	۹/۷۷	۹/۴۲	-۳/۵۵



شکل (۲-۱۳) - ترسیمی از مقادیر پیش‌بینی شده PIC₅₀ با شبکه عصبی مصنوعی بر حسب مقدار واقعی در سری تست

۲-۱-۳-۷-۳- ارزیابی شبکه به روش رد مرحله‌ای تک تک

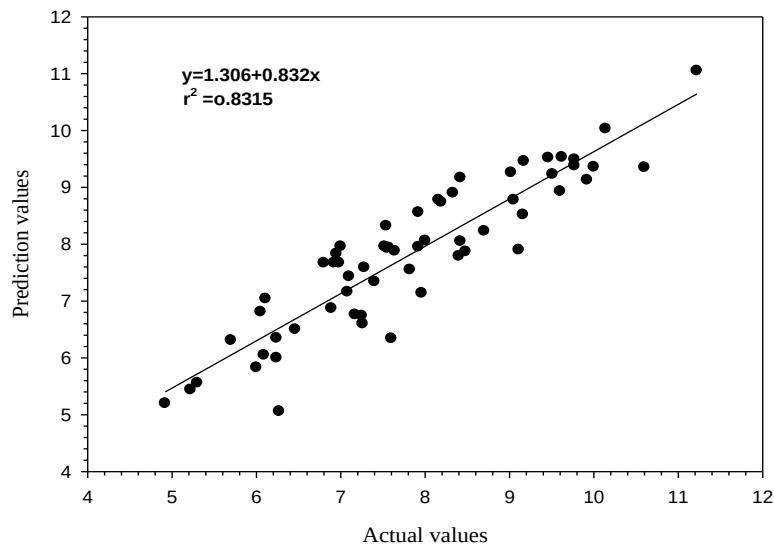
در مرحله‌ی بعد برای ارزیابی بیشتر شبکه عصبی بهینه، روش رد تک تک مورد استفاده قرار گرفت. در این روش به این صورت عمل گردید که ابتدا یک مولکول از داده‌ها کنار گذاشته شد و شبکه عصبی در شرایط بهینه با ۵۹ ترکیب باقیمانده آموزش داده شد، سپس شبکه آموزش داده شده برای پیش‌بینی $\log(1/IC_{50})$ (ترکیب کنار گذاشته شده، بکار گرفته شد. نتایج حاصله در جدول (۲-۱۹) نشان داده شده است. شکل (۲-۱۴) هم مقادیر پیش‌بینی شده بر حسب مقادیر واقعی را

نشان می‌دهد. نتایج بیانگر این است که مدل ارائه شده در این تحقیق خطای پیش‌بینی کم و صحت بالایی دارد و همچنین قدرت تعمیم شبکه را بیان می‌کند لذا با این مدل می‌توان داده‌های

$\log(1/IC_{50})$ (را برای هر ترکیب جدید با خطای قابل قبول پیش‌بینی کرد.

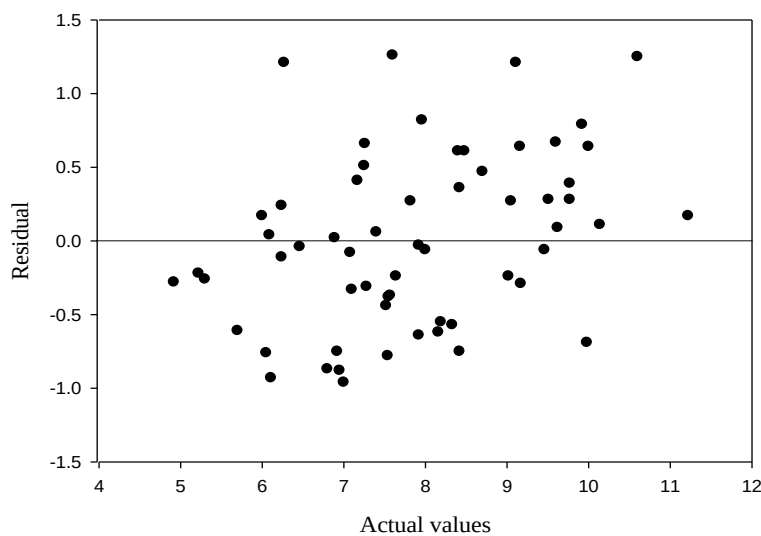
جدول (۲-۱۹)- نتایج حاصل از ارزیابی مدل شبکه عصبی با استفاده از روش رد تک تک

شماره	مقدار واقعی	مقدار پیش‌بینی شده	درصد خطا	شماره	مقدار واقعی	مقدار پیش‌بینی شده	درصد خطا
۳۱	۹/۰۵	۸/۷۸	۲/۹۸	۱	۶/۴۶	۶/۵۰	-۰/۶۲
۳۲	۹/۶۲	۹/۵۳	۰/۹۴	۲	۷/۲۵	۶/۷۴	۷/۰۳
۳۳	۸/۴۰	۷/۷۹	۷/۲۶	۳	۷/۶۴	۷/۸۸	-۳/۱۴
۳۴	۷/۰۸	۷/۱۶	-۱/۱۳	۴	۷/۵۷	۷/۹۴	-۴/۸۹
۳۵	۷/۹۲	۸/۵۶	-۸/۰۸	۵	۷/۱۰	۷/۴۳	-۴/۶۵
۳۶	۸/۳۳	۸/۹۰	-۶/۸۴	۶	۶/۲۴	۶/۳۵	-۱/۷۶
۳۷	۷/۴۰	۷/۳۴	۰/۸۱	۷	۷/۲۶	۶/۶۰	۹/۰۹
۳۸	۸/۱۶	۸/۷۸	-۷/۶۰	۸	۷/۵۲	۷/۹۶	-۵/۸۵
۳۹	۷/۹۲	۷/۹۵	-۰/۳۸	۹	۶/۱۱	۷/۰۴	-۱۵/۲۲
۴۰	۷/۹۶	۷/۱۴	۱۰/۳۰	۱۰	۶/۲۴	۶/۰۰	۳/۸۵
۴۱	۹/۹۲	۹/۱۳	۷/۹۶	۱۱	۹/۱۱	۷/۹۰	۱۳/۲۸
۴۲	۹/۱۶	۸/۵۲	۶/۹۹	۱۲	۵/۳۰	۵/۵۶	-۴/۹۰
۴۳	۶/۸۰	۷/۶۷	-۱۲/۷۹	۱۳	۶/۰۵	۶/۸۱	-۱۲/۵۶
۴۴	۸/۴۲	۹/۱۷	-۸/۹۱	۱۴	۶/۹۲	۷/۶۷	-۱۰/۸۴
۴۵	۸/۱۹	۸/۷۴	-۶/۷۲	۱۵	۸/۰۰	۸/۰۶	-۰/۷۵
۴۶	۷/۲۸	۷/۵۹	-۴/۲۶	۱۶	۷/۰۰	۷/۹۶	-۱۳/۷۱
۴۷	۷/۱۷	۶/۷۶	۵/۷۲	۱۷	۸/۷۰	۸/۲۳	۵/۴۰
۴۸	۵/۷۰	۶/۳۱	-۱۰/۷۰	۱۸	۹/۶۰	۸/۹۳	۶/۹۸
۴۹	۷/۵۵	۷/۹۳	-۵/۰۳	۱۹	۱۰/۰۰	۹/۳۶	۶/۴۰
۵۰	۷/۸۲	۷/۵۵	۳/۴۵	۲۰	۹/۴۶	۹/۵۲	-۰/۶۳
۵۱	۷/۶۰	۶/۳۴	۱۶/۵۸	۲۱	۸/۴۸	۷/۸۷	۷/۱۹
۵۲	۹/۷۷	۹/۳۸	۳/۹۹	۲۲	۴/۹۲	۵/۲۰	-۵/۶۹
۵۳	۱۱/۲۲	۱۱/۰۵	۱/۵۲	۲۳	۵/۲۲	۵/۴۴	-۴/۲۱
۵۴	۹/۷۷	۹/۴۹	۲/۸۶	۲۴	۶/۰۹	۶/۰۵	۰/۶۶
۵۵	۶/۹۸	۷/۶۷	-۹/۸۹	۲۵	۶/۰۰	۵/۸۳	۲/۸۳
۵۶	۹/۱۷	۹/۴۶	-۳/۱۶	۲۶	۶/۲۷	۵/۰۶	۱۹/۳۰
۵۷	۶/۸۹	۶/۸۷	۰/۲۹	۲۷	۷/۵۴	۸/۳۲	-۱۰/۳۴
۵۸	۹/۰۲	۹/۲۶	-۲/۶۶	۲۸	۸/۴۲	۸/۰۵	۴/۳۹
۵۹	۱۰/۱۴	۱۰/۰۳	۱/۰۸	۲۹	۹/۵۱	۹/۲۳	۲/۹۴
۶۰	۶/۹۵	۷/۸۳	-۱۲/۶۶	۳۰	۱۰/۶۰	۹/۳۵	۱۱/۷۹



شکل (۲-۱۴) - ترسیمی از مقادیر پیش‌بینی شده با شبکه عصبی مصنوعی بر حسب مقدار واقعی برای کل داده‌ها

به منظور بررسی وجود خطای معین در این مدل، باقیمانده مقادیر پیش‌بینی شده توسط مدل شبکه عصبی مصنوعی از مقادیر واقعی بر حسب مقادیر تجربی برای مولکول رسم شده است (شکل (۲-۱۵)). توزیع نسبتاً یکنواخت و اتفاقی باقی مانده‌ها در دو طرف خط صفر نشان می‌دهد که خطای معین در مدل وجود ندارد.



شکل (۲-۱۵) - نمودار باقیمانده‌ها برای کل داده‌ها (پیش‌بینی با شبکه عصبی مصنوعی)

۲-۱-۳-۷-۴- آزمون Y - تصادفی

به منظور اجتناب از همبستگی‌های تصادفی و تضمین قدرت شبکه، تست Y – تصادفی انجام شد. لذا داده‌های $\log(1/IC_{50})$ (متغیر وابسته) بطور تصادفی ایجاد شدند. مدل QSAR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از فعالیت ترکیبات توسعه یافت. نتایج حاصل از چندین مجموعه‌ی تصادفی ایجاد شده، در جدول (۲-۲۰) نشان داده شده است. با توجه به نتایج هیچ وابستگی برجسته‌ای بین خواص ساختاری و داده‌های بازداری برای ترکیبات وجود ندارد (ضریب همبستگی پایین)، که نشان‌دهنده‌ی عدم همبستگی تصادفی در مدل شبکه عصبی مصنوعی ایجاد شده می‌باشد.

جدول (۲-۲۰) - مقادیر R^2 برای سری ارزیابی و سری تست بعد از چندین تست Y - تصادفی

تکرار	R^2 سری ارزیابی (n=10)	R^2 سری تست (n=5)
۱	۰/۱۰	۰/۱۶
۲	۰/۰۸	۰/۰۳
۳	۰/۰۶	۰/۴۷
۴	۰/۰۸	۰/۱۲
۵	۰/۰۵	۰/۳۰
۶	۰/۰۴	۰/۰۹
۷	۰/۵۰	۰/۰۰
۸	۰/۴۰	۰/۱۹
۹	۰/۳۷	۰/۶۱
۱۰	۰/۰۱	۰/۰۵

۲-۱-۳-۵- ارزیابی مدل شبکه عصبی ارائه شده با استفاده از پارامترهای آماری

برای بررسی میزان کارایی مدل شبکه عصبی ارائه شده و مقایسه‌ی آن با روش رگرسیون

خطی چندگانه چهار پارامتر آماری MAE ، REP ، MSE و R^2 محاسبه شد (در بخش (۲-۱-۳-۵-۱))

توضیح داده شد). نتایج حاصل در جدول (۲-۲۱) آمده است:

جدول (۲-۲۱) - پارامترهای آماری محاسبه شده برای مدل شبکه عصبی			
پارامتر	سری ارزیابی (n = 10)	سری تست (n = 5)	کل داده‌ها (n = 60)
MAE	۴/۲	۴/۷۹	۶/۲۴
REP	۴/۱۹	۲/۳۹	۷/۴۱
MSE	۰/۱۳۲۹	۰/۱۹۳۷	۰/۳۳۶۷
R^2	۰/۹۰۵۹	۰/۹۶۳۴	۰/۸۳۱۵

نتایج حاصله نشان‌دهنده‌ی برتری مدل شبکه عصبی مصنوعی بر مدل حاصل از روش

رگرسیون خطی است. و همچنین بیانگر این است که روش QSAR ارائه شده در این پروژه قادر است

فعالیت مهارکنندگی مشتقات با هر استخلاف را با قدرت بالا پیش‌بینی کند.

۲-۲- محاسبات و مدل سازی بازدارنده‌ی نکروپتوسیس (نکروستاتین-۵)

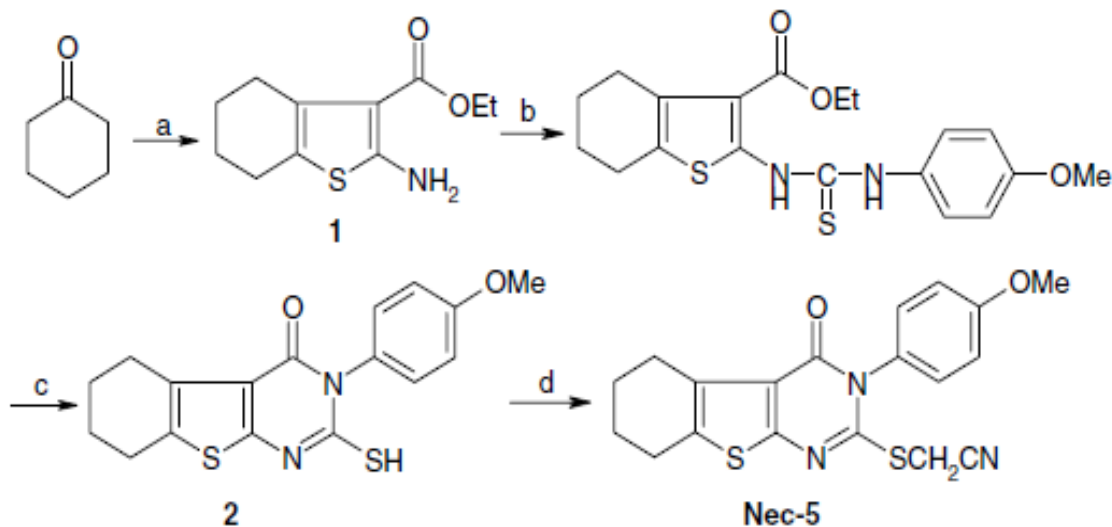
در این پروژه از شبکه‌ی عصبی مصنوعی بعنوان یک شیوه‌ی مدل سازی غیرخطی استفاده شده است. در مرحله اول ابتدا ۱۶۱ ساختار مولکولی بهینه شدند و مدلی طراحی شد که بتواند مهارکننده‌ها را به دو کلاس فعال و غیرفعال تقسیم کند. در مرحله‌ی دوم یک مدل شبکه‌ی عصبی مصنوعی طراحی و بهینه گردید که بتواند فعالیت دارویی (EC_{50}) ۵۲ ترکیب فعال از مشتقات نکروستاتین-۵ را به درستی پیش‌بینی نماید. EC_{50} بیانگر غلظتی از دارو می‌باشد که نیمی از حداکثر اثر دارویی را نشان می‌دهد.

۲-۲-۱- سری داده‌ها

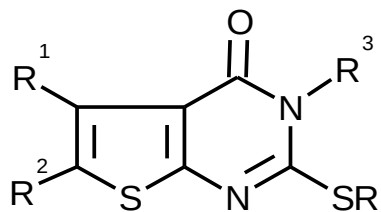
ترکیبات مورد استفاده در این پروژه، داده‌های ۱۶۱ ترکیب از مشتقات نکروستاتین-۵ ($Nec-5$) می‌باشند که از مرجع شماره‌ی [۵۷] گردآوری شده است. مقادیر منهای لگاریتم داده‌های EC_{50} بر حسب مولار ($-\log EC_{50} \times 10^{-6}$) به عنوان متغیر وابسته، در این مدل سازی مورد استفاده قرار گرفته است. به طور شیمیایی $Nec-5$ ، با عنوان $p-3$ متوکسی فنیل ۱-۵ و ۶- تترا متیلنو تینو [۲و۳- d] پیریمیدین-۴-یک-۲- مرکاپتو اتیل سیانید^۱ نامیده می‌شود و روش سنتز آن به طور تجربی مطابق با مرجع فوق در شکل (۲-۱۶) نشان داده شده است. ساختار پایه در شکل (۲-۱۷) آمده است. در جدول (۲-۲۲) نیز جزئیات استخلاف‌ها، فعال یا غیرفعال بودن هر ترکیب و مقادیر عددی pEC_{50} بر حسب مولار متناظر هر ترکیب فعال آورده شده است. با توجه به نتایج تجربی از ۱۶۱ مشتق $Nec-5$ ، ۵۲ مورد آن دارای فعالیت مهارکنندگی می‌باشد. لذا در ابتدا باید مدلی ساخته شود که قادر به پیش‌بینی نوع فعالیت ترکیبات مورد نظر باشد. به این معنی که مدل انتخاب شده توانایی تقسیم

¹ - 3-p-methoxyphenyl-5,6-tetramethylenothieno[2,3-d]pyrimidin-4-one-2-mercaptoethylcyanide

ترکیبات به دو گروه فعال و غیرفعال را داشته باشد، سپس ارتباط ساختار - فعالیت برای پیش‌بینی داده‌های EC_{50} مشتقات فعال صورت پذیرد.

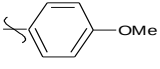
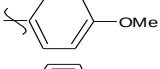
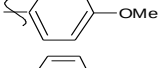
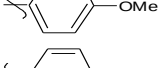
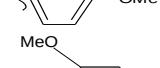
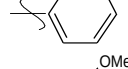
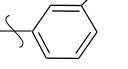
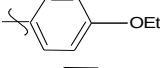
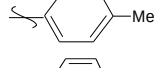
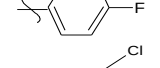
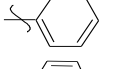
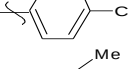
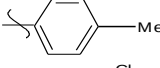
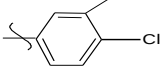
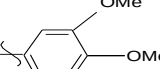
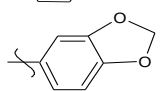
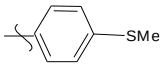
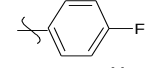
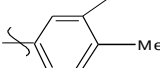
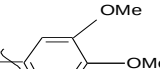
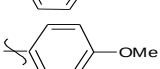


شکل (۲-۱۶) - روش سنتز Nec-5 شرایط و واکنش‌گرها: (a) سیانواستات، S_8 ، Et_2NH ، $EtOH$ ، ۱۲ ساعت رفلکس، ۷۲٪ بازده؛ (b) p-متوکسی فنیل ایزوتیوسیانات، $EtOH$ ، ۵-۶ ساعت رفلکس، ۸۵٪ بازده؛ (c) HCl اتانولی، ۱۲-۲۴ ساعت رفلکس، ۷۸٪ بازده؛ (d) KOH در $EtOH$ ، ۷۰٪ سیس $BrCH_2CN$ ، ۱-۲ ساعت، ۹۲٪ بازده.

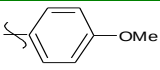
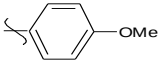
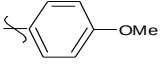
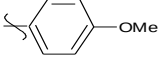
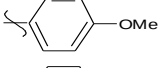
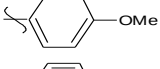
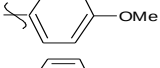
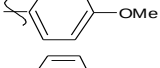
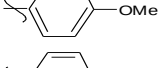
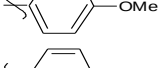
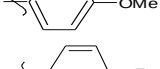
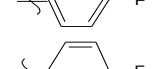
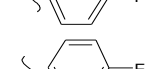
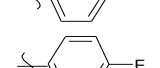
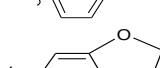
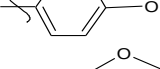
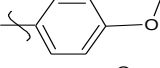
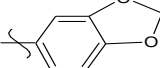
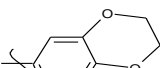
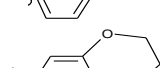
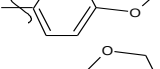
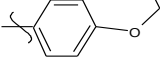


شکل (۲-۱۷) - اسکلت مولکولی ترکیبات مورد بررسی

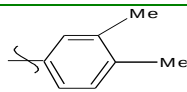
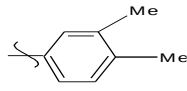
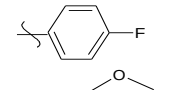
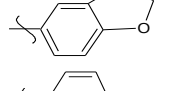
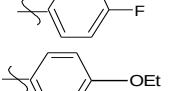
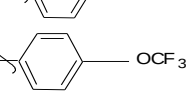
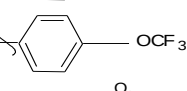
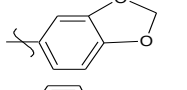
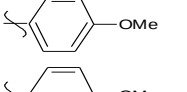
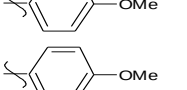
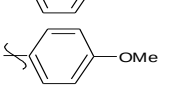
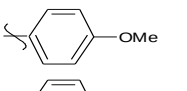
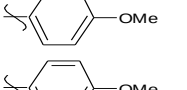
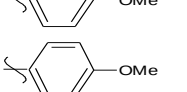
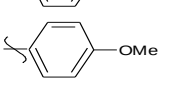
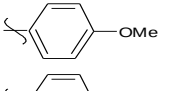
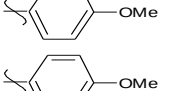
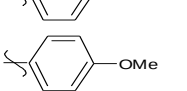
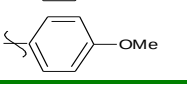
جدول (۲-۲۲) - جزئیات ساختار مولکولی مشتقات Nec-5

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۱	CH ₂ CN	—(CH ₂) ₄ —			۶/۶۲
۲	Me	—(CH ₂) ₄ —			۶/۶۲
۳	CH ₂ C≡CH	—(CH ₂) ₄ —			۵/۲۲
۴	CH ₂ CH ₂ CN	—(CH ₂) ₄ —			۵/۲۸
۵	CH ₂ Cl	—(CH ₂) ₄ —			۵/۶۵
۶	CH ₂ CN	—(CH ₂) ₄ —			۴/۷۵
۷	CH ₂ CN	—(CH ₂) ₄ —			۵/۸۳
۸	CH ₂ CN	—(CH ₂) ₄ —			۶/۲۶
۹	CH ₂ CN	—(CH ₂) ₄ —			۵/۷۴
۱۰	CH ₂ CN	—(CH ₂) ₄ —			۶/۶۲
۱۱	CH ₂ CN	—(CH ₂) ₄ —			۵/۲۹
۱۲	CH ₂ CN	—(CH ₂) ₄ —			۵/۰۷
۱۳	CH ₂ CN	—(CH ₂) ₄ —			۴/۷۸
۱۴	CH ₂ CN	—(CH ₂) ₄ —			۵/۲۴
۱۵	CH ₂ CN	—(CH ₂) ₄ —			۵/۲۴
۱۶	CH ₂ CN	—(CH ₂) ₄ —			۶/۰۵
۱۷	CH ₂ CN	—(CH ₂) ₄ —			۵/۶۳
۱۸	Me	—(CH ₂) ₄ —			۵/۸۰
۱۹	Me	—(CH ₂) ₄ —			۴/۷۸
۲۰	Me	—(CH ₂) ₄ —			۴/۷۸
۲۱	CH ₂ CN	Me	Me		۶/۳۵

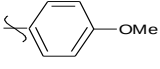
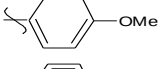
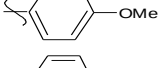
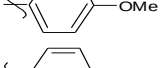
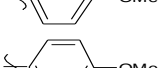
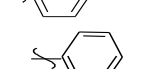
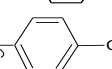
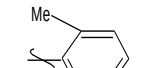
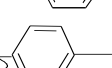
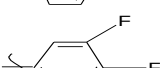
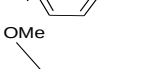
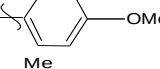
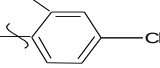
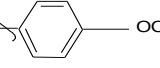
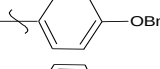
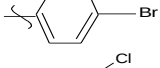
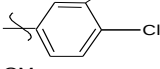
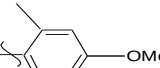
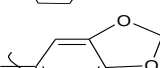
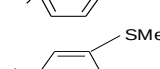
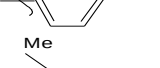
ادامہی جدول (۲-۲۲)

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۲۲	CH ₂ CN	Me	Et		۵/۲۸
۲۳	CH ₂ CN	Et	Me		۵/۹۷
۲۴	CH ₂ CN	—(CH ₂) ₃ —			۶/۳۵
۲۵	CH ₂ CN	—(CH ₂) ₅ —			۶/۰۲
۲۶	CH ₂ CN	-CH=CHCH=CH-			۶/۷۴
۲۷	Me	—(CH ₂) ₃ —			۶/۶۲
۲۸	Me	-CH=CHCH=CH-			۶/۶۲
۲۹	CH ₂ C≡CH	—(CH ₂) ₃ —			۵/۶۰
۳۰	CH ₂ CH ₂ OH	—(CH ₂) ₃ —			۵/۱۵
۳۱	CH ₂ C≡CH	Me	Me		۵/۳۱
۳۲	CH ₂ CH ₂ OH	Me	Me		۵/۱۴
۳۳	CH ₂ CN	Me	Me		۵/۳۵
۳۴	CH ₂ CN	—(CH ₂) ₃ —			۶/۰۵
۳۵	Me	Me	Me		۵/۵۸
۳۶	Me	—(CH ₂) ₃ —			۵/۵۲
۳۷	CH ₂ CN	Me	Me		۵/۷۸
۳۸	CH ₂ CN	Et	Me		۵/۷۲
۳۹	CH ₂ CN	—(CH ₂) ₃ —			۵/۹۳
۴۰	CH ₂ CN	—(CH ₂) ₃ —			۶/۶۰
۴۱	CH ₂ CN	—(CH ₂) ₄ —			۶/۶۶
۴۲	CH ₂ CN	-CH=CHCH=CH-			۶/۸۲
۴۳	CH ₂ CN	Me	Me		۶/۶۰

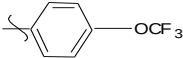
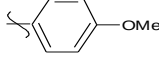
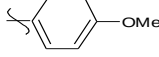
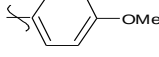
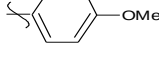
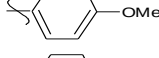
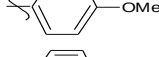
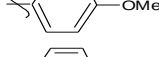
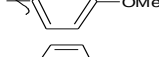
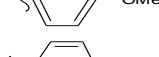
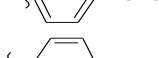
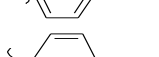
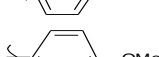
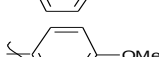
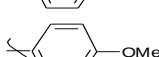
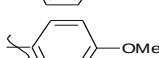
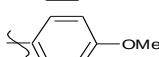
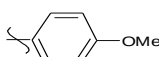
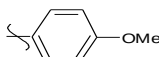
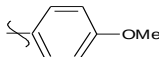
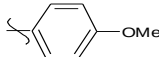
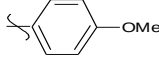
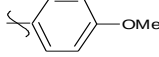
ادامہی جدول (۲-۲۲)

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۴۴	CH ₂ CN	Me	Me		۵/۵۵
۴۵	CH ₂ CN	-CH ₂ CH ₂ CH(CH ₃)CH ₂ -			۴/۷۸
۴۶	CH ₂ CN	-CH=CHCH=CH-			۵/۵۱
۴۷	CH ₂ CN	-CH=CHCH=CH-			۶/۵۷
۴۸	Me	-CH=CHCH=CH-			۵/۵۱
۴۹	CH ₂ CN	Me	Me		۵/۱۱
۵۰	CH ₂ CN	Me	Me		۵/۴۳
۵۱	CH ₂ CN	—(CH ₂) ₃ —			۵/۴۳
۵۲	Me	Me	Me		۵/۹۵
۵۳	Et	—(CH ₂) ₄ —			غیر فعال
۵۴	n-Pr	—(CH ₂) ₄ —			غیر فعال
۵۵	n-Bu	—(CH ₂) ₄ —			غیر فعال
۵۶	n-Pent	—(CH ₂) ₄ —			غیر فعال
۵۷	n-Hex	—(CH ₂) ₄ —			غیر فعال
۵۸	CH ₂ CH=CH ₂	—(CH ₂) ₄ —			غیر فعال
۵۹	CH ₂ C ₆ H ₅	—(CH ₂) ₄ —			غیر فعال
۶۰	CH ₂ (C ₆ H ₄ Me-4)	—(CH ₂) ₄ —			غیر فعال
۶۱	CH ₂ (C ₆ H ₄ OMe-4)	—(CH ₂) ₄ —			غیر فعال
۶۲	CH ₂ (C ₆ H ₄ NO ₂ -4)	—(CH ₂) ₄ —			غیر فعال
۶۳	CH ₂ COMe	—(CH ₂) ₄ —			غیر فعال
۶۴	CH ₂ COOMe	—(CH ₂) ₄ —			غیر فعال
۶۵	CH ₂ CONH ₂	—(CH ₂) ₄ —			غیر فعال
۶۶	COMe	—(CH ₂) ₄ —			غیر فعال

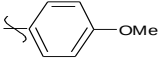
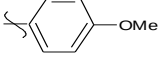
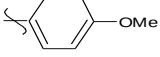
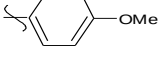
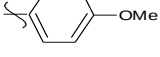
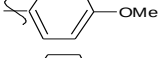
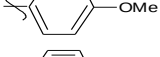
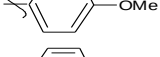
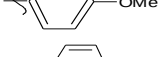
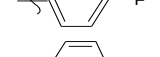
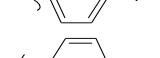
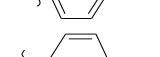
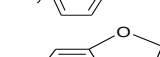
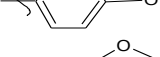
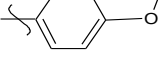
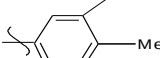
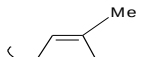
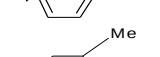
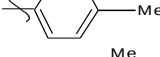
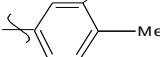
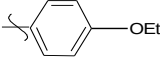
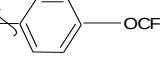
ادامہی جدول (۲-۲۲)

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۶۷	COC ₃ H _{7-n}	—(CH ₂) ₄ —			غیر فعال
۶۸	COC ₆ H ₅	—(CH ₂) ₄ —			غیر فعال
۶۹	CH ₂ NO ₂	—(CH ₂) ₄ —			غیر فعال
۷۰	CH ₂ C(O)NH(C ₆ H ₄ CF ₃ -2)	—(CH ₂) ₄ —			غیر فعال
۷۱	CH ₂ CH(OH)CH ₃	—(CH ₂) ₄ —			غیر فعال
۷۲	CH ₂ COOH	—(CH ₂) ₄ —			غیر فعال
۷۳	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۷۴	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۷۵	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۷۶	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۷۷	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۷۸	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۷۹	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۸۰	CH ₂ CN	—(CH ₂) ₄ —			غیر فعال
۸۱	Me	—(CH ₂) ₄ —			غیر فعال
۸۲	Me	—(CH ₂) ₄ —			غیر فعال
۸۳	Me	—(CH ₂) ₄ —			غیر فعال
۸۴	Me	—(CH ₂) ₄ —			غیر فعال
۸۵	Me	—(CH ₂) ₄ —			غیر فعال
۸۶	Me	—(CH ₂) ₄ —			غیر فعال
۸۷	Me	—(CH ₂) ₄ —			غیر فعال

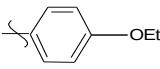
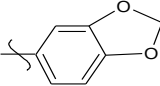
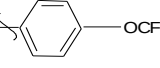
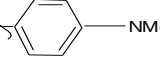
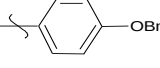
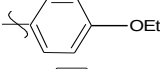
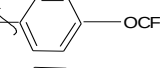
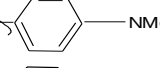
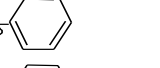
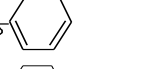
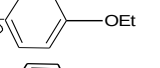
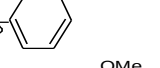
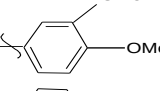
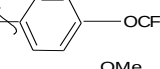
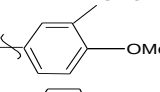
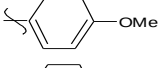
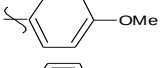
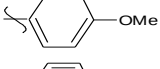
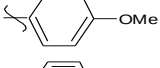
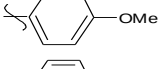
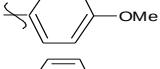
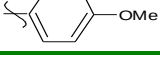
ادامه‌ی جدول (۲-۲۲)

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۸۸	Me	—(CH ₂) ₄ —			غیرفعال
۸۹	CH ₂ CN	H	H		غیرفعال
۹۰	CH ₂ CN	H	Me		غیرفعال
۹۱	CH ₂ CN	H	Et		غیرفعال
۹۲	CH ₂ CN	Me	H		غیرفعال
۹۳	CH ₂ CN	Me	<i>n</i> -Pr		غیرفعال
۹۴	CH ₂ CN	Me	<i>i</i> -Pr		غیرفعال
۹۵	CH ₂ CN	Me	C ₁₄ H ₂₉		غیرفعال
۹۶	CH ₂ CN	-CH ₂ CH ₂ CHCH ₃ CH ₂ -			غیرفعال
۹۷	CH ₂ CN	-CH ₂ CH ₂ NEtCH ₂ -			غیرفعال
۹۸	CH ₂ CN	-CH ₂ CH ₂ N(<i>i</i> -Pr)CH ₂ -			غیرفعال
۹۹	Et	—(CH ₂) ₃ —			غیرفعال
۱۰۰	<i>n</i> -Pr	—(CH ₂) ₃ —			غیرفعال
۱۰۱	<i>n</i> -Bu	—(CH ₂) ₃ —			غیرفعال
۱۰۲	<i>n</i> -Pent	—(CH ₂) ₃ —			غیرفعال
۱۰۳	CH ₂ CH=CH ₂	—(CH ₂) ₃ —			غیرفعال
۱۰۴	CH ₂ C ₆ H ₅	—(CH ₂) ₃ —			غیرفعال
۱۰۵	CH ₂ (C ₆ H ₄ NO ₂ -4)	—(CH ₂) ₃ —			غیرفعال
۱۰۶	CH ₂ COMe	—(CH ₂) ₃ —			غیرفعال
۱۰۷	CH ₂ NO ₂	—(CH ₂) ₃ —			غیرفعال
۱۰۸	CH ₂ C≡CH	H	H		غیرفعال
۱۰۹	CH ₂ C≡CH	Me	H		غیرفعال
۱۱۰	CH ₂ C≡CH	Me	Et		غیرفعال
۱۱۱	CH ₂ C≡CH	Me	<i>n</i> -Pr		غیرفعال

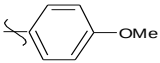
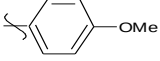
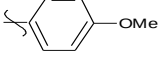
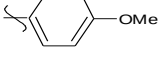
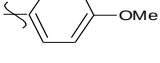
ادامہی جدول (۲-۲۲)

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۱۱۲	CH ₂ C≡CH	—(CH ₂) ₅ —			غیر فعال
۱۱۳	Et	Me	Me		غیر فعال
۱۱۴	Et	Me	Et		غیر فعال
۱۱۵	Et	—(CH ₂) ₅ —			غیر فعال
۱۱۶	CH ₂ CH ₂ CN	Me	Me		غیر فعال
۱۱۷	CH ₂ CH ₂ OH	Me	Et		غیر فعال
۱۱۸	CH ₂ CH ₂ OH	Me	<i>n</i> -Pr		غیر فعال
۱۱۹	Et	Me	<i>n</i> -Pr		غیر فعال
۱۲۰	CH ₂ C(O)OH	Me	Me		غیر فعال
۱۲۱	CH ₂ CN	Me	Et		غیر فعال
۱۲۲	CH ₂ CN	—(CH ₂) ₅ —			غیر فعال
۱۲۳	CH ₃	Me	Et		غیر فعال
۱۲۴	CH ₃	—(CH ₂) ₅ —			غیر فعال
۱۲۵	CH ₃	Et	Me		غیر فعال
۱۲۶	CH ₃	—(CH ₂) ₃ —			غیر فعال
۱۲۷	CH ₂ CN	—(CH ₂) ₃ —			غیر فعال
۱۲۸	CH ₃	Me	Me		غیر فعال
۱۲۹	CH ₃	-CH ₂ CH ₂ CH(CH ₃)CH ₂ -			غیر فعال
۱۳۰	CH ₃	—(CH ₂) ₃ —			غیر فعال
۱۳۱	CH ₂ CN	-CH=CHCH=CH-			غیر فعال
۱۳۲	CH ₂ CN	-CH=CHCH=CH-			غیر فعال
۱۳۳	CH ₂ CN	-CH=CHCH=CH-			غیر فعال

ادامه‌ی جدول (۲-۲۲)

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۱۳۴	CH ₃	-CH=CHCH=CH-			غیرفعال
۱۳۵	CH ₃	-CH=CHCH=CH-			غیرفعال
۱۳۶	CH ₃	-CH=CHCH=CH-			غیرفعال
۱۳۷	CH ₂ CN	Me	Me		غیرفعال
۱۳۸	CH ₂ CN	Me	Me		غیرفعال
۱۳۹	CH ₃	Me	Me		غیرفعال
۱۴۰	CH ₃	Me	Me		غیرفعال
۱۴۱	CH ₃	Me	Me		غیرفعال
۱۴۲	CH ₃	Me	Me		غیرفعال
۱۴۳	CH ₂ CN	H	H		غیرفعال
۱۴۴	CH ₂ CN	—(CH ₂) ₃ —			غیرفعال
۱۴۵	CH ₂ CN	—(CH ₂) ₃ —			غیرفعال
۱۴۶	CH ₂ CN	—(CH ₂) ₅ —			غیرفعال
۱۴۶	CH ₂ CN	-CH ₂ CH ₂ CH(CH ₃)CH ₂ -			غیرفعال
۱۴۸	CH ₂ CN	—(CH ₂) ₅ —			غیرفعال
۱۴۹	CH ₂ CN	—(CH ₂) ₃ —			غیرفعال
۱۵۰	CH ₃	H	H		غیرفعال
۱۵۱	CH ₃	Me	H		غیرفعال
۱۵۲	CH ₃	Me	Me		غیرفعال
۱۵۳	CH ₃	Et	Me		غیرفعال
۱۵۴	CH ₃	Me	Et		غیرفعال
۱۵۵	CH ₃	Me	<i>n</i> -Pr		غیرفعال
۱۵۶	CH ₃	Me	<i>i</i> -Pr		غیرفعال

ادامه‌ی جدول (۲-۲۲)

No	R	R ¹	R ²	R ³	P(EC ₅₀ ×10 ⁻⁶)(M)
۱۵۷	CH ₃	Me	C ₁₄ H ₂₉		غیرفعال
۱۵۸	CH ₃	—(CH ₂) ₅ —			غیرفعال
۱۵۹	CH ₃	-CH ₂ CH ₂ CHCH ₃ CH ₂ -			غیرفعال
۱۶۰	CH ₃	-CH ₂ CH ₂ NEtCH ₂ -			غیرفعال
۱۶۱	CH ₃	-CH ₂ CH ₂ N(i-Pr)CH ₂ -			غیرفعال

۲-۲-۲- نرم افزارهای مورد استفاده

همچنانکه در قسمتهای قبلی ذکر شد از Hyperchem 7 برای ترسیم و بهینه سازی اولیه‌ی ساختار ترکیبات استفاده شد که با استفاده از روش نیمه تجربی AM1 قادر به محاسبه قادر به محاسبه‌ی پارامترهای مولکول‌ها می‌باشد. برای محاسبه‌ی توصیف‌گرهای مولکولی از نرم‌افزار Dragon که توسط یک گروه تحقیقاتی ایتالیایی نوشته شده است، استفاده شد. از نرم‌افزار SPSS برای انتخاب توصیف‌گرهای مولکولی مناسب و ایجاد مدل‌های آماری از جمله رگرسیون استفاده شد. برنامه‌ی شبکه عصبی مصنوعی در نرم‌افزار MATLAB 7.5 نوشته شد. تمام محاسبات با استفاده از یک کامپیوتر پنتیوم-۴ با سیستم عامل ویندوز XP انجام گردید.

۲-۲-۳- انجام محاسبات و مدل سازی

۲-۲-۳-۱- رسم و بهینه سازی ساختار مولکول‌ها

بدین منظور ساختار دوبعدی مشتقات Nec-5 در نرم‌افزار Hyperchem7 ترسیم شدند و ساختار نهایی به روش نیمه تجربی و با استفاده از روش AM1 بهینه گردید. تمامی محاسبات در سطح

محدود شده‌ی هارتری - فاک انجام گردید. بهینه سازی تا زمانی ادامه یافت که گرادیان، $0/001$ کیلو کالری بر مول برسد.

۲-۳-۲-۲- طبقه‌بندی ترکیبات نکروستاتین-۵

همانطور که بیان شد، بعضی از ترکیبات که به طور تجربی سنتز شده‌اند، فاقد فعالیت مهارکنندگی می‌باشند. لذا در مرحله اول تحقیق، هدف ارائه مدلی است که قادر به تقسیم‌بندی ترکیبات به دو گروه فعال و غیرفعال باشد. این کار بدان جهت انجام گرفت که اگر ترکیبی از این خانواده وارد مدل شود، مدل نوع فعالیت آن را پیش‌بینی نماید، سپس رابطه‌ی کمی ساختار - فعالیت EC_{50} ترکیبات فعال، بررسی شود. با توجه به تعریف EC_{50} ، هر چه EC_{50} بر حسب درصد غلظت (مولار) افزایش یابد (مقدار $-LogEC_{50}$ بر حسب مولار کاهش یابد) فعالیت دارویی کم می‌شود. با بررسی مقادیر EC_{50} در جدول (۲-۲۲) مشاهده می‌شود که برای ترکیبات فعال مقادیر $-LogEC_{50}$ بر حسب مولار در ناحیه تقریبی $4/0-7/0$ قرار دارند. بنابراین در طبقه‌بندی ترکیبات به دو دسته‌ی فعال و غیرفعال ترکیباتی با pEC_{50} کمتر از $4/0$ غیرفعال و ترکیبات با pEC_{50} بیشتر از $4/0$ به عنوان ترکیبات فعال در نظر گرفته شدند.

درانتخاب توصیف‌گر و در مدل‌سازی شبکه عصبی، ترکیبات فعال با کد ۸ ($pEC_{50}=8$) و ترکیبات غیرفعال با کد ۳ ($pEC_{50}=3$) در نظر گرفته شدند. پس از آموزش شبکه و در ارزیابی شبکه، عمل کدبرداری انجام گرفت به این صورت که تمام مولکول‌هایی که pEC_{50} پیش‌بینی شده برای آن‌ها بیشتر از ۴ بود به عنوان ترکیب فعال و تمام مولکول‌های که pEC_{50} آن‌ها کمتر از ۴ بود به عنوان غیرفعال در نظر گرفته شدند.

۲-۲-۳-۱- محاسبه و انتخاب توصیف‌گرهای مهم

همانطور که در قسمت‌های قبلی بیان شد، نرم افزار Dragon تعداد زیادی توصیف‌گر را برای ترکیبات محاسبه می‌کند. با روش رگرسیون خطی چند گانه - در بخش (۱-۱-۶) به طور مبسوط شرح داده شد- می‌توان از میان تعداد بسیاری متغیر مستقل (توصیف‌گر مولکولی) آن‌هایی را که بیشترین تاثیر را در متغیر وابسته دارند، مشخص کرد و به هر کدام به ترتیب اهمیت ضریب داد. رگرسیون خطی چندگانه با استفاده از امکانات نرم افزار SPSS انجام شد. از میان ۱۴۸۱ توصیف‌گر برای مولکول، ۱۷ توصیف‌گر انتخاب گردید.

۲-۲-۳-۲- مدل شبکه عصبی مصنوعی برای طبقه‌بندی ترکیبات نکروستاتین-۵ و

بهینه‌سازی آن

به طور کلی در روند ایجاد مدل‌های پیش‌گو مجموعه داده‌ها به سه دسته‌ی آموزشی، ارزیابی و تست تقسیم شدند. آن گاه مدل مورد نظر با گروه آموزشی ایجاد شده، با گروه آزمایشی ارزیابی شده و با سری تست برای جلوگیری از خطای برازش اضافی سنجیده می‌شود. لذا ۱۱۷ ترکیب از ۱۶۱ ترکیب برای سری آموزشی، ۳۰ ترکیب برای سری ارزیابی و ۱۴ ترکیب برای سری تست انتخاب شد. هر چه تعداد موارد گروه آموزشی بیشتر باشد، مدل دقیق‌تری خواهیم داشت و هر چه تعداد گروه ارزیابی بیشتر شود مدل قابل اعتمادتر خواهد بود.

شبکه عصبی مصنوعی با استفاده از نرم افزار MATLAB 7.5 طراحی و بهینه‌سازی شد. شبکه‌ی انتخاب شده یک شبکه عصبی پیش‌خور با الگوی پس انتشار خطا و الگوریتم آموزشی تنظیم بایزین می‌باشد. در روند بهینه‌سازی و شناسایی بهترین پارامترهای شبکه، شبکه‌هایی با ورودی‌های از ۲ تا ۱۷ توصیف‌گر ایجاد شدند. هر شبکه با الگوریتم آموزشی تنظیم بایزین با تعداد گره‌های متفاوت در لایه پنهان آموزش داده شد و در نرون‌های هر لایه از تابع لگاریتم زیگموئید به عنوان تابع فعالیت

استفاده شد. در تمام شبکه‌ها تعداد دوره‌های آموزشی یکسان و برابر ۱۰ در نظر گرفته شد. که پس از ایجاد هر شبکه قدرت پیش‌گویی آن با سه شاخص FC، FAR و POD (در قسمت (۸-۱-۱) توضیح داده شده است) ارزیابی شد. نتایج این ارزیابی در جدول (۲-۲۳) آمده است.

با توجه به نتایج جدول (۲-۲۳) مدل با ۱۱ متغیر ورودی (توصیف‌گر) و ۲ نرون در لایه پنهان شاخص‌های خوبی را ایجاد کرد و شبکه با این ویژگی، پیش‌گویی خوبی داشت و قادر بود ۱۰۰٪ ترکیبات را به درستی گروه بندی کند. این توصیف‌کننده‌ها به همراه کلاس مربوطه در جدول (۲-۲۴) آمده است. ماتریس همبستگی بین توصیف‌گرها نیز در جدول (۲-۲۵) نشان داده شده است، که مشاهده می‌شود که بین توصیف‌گرها همبستگی وجود ندارد.

جدول (۲-۲۳) - نتایج ارزیابی شبکه‌های ایجاد شده با الگوریتم آموزشی تنظیم بایزین (trainbr) و تابع تبدیل لگاریتم زیگموئید (logsig) با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

			تعداد گره‌ها				
			۲	۳	۴	۵	۶
تعداد متغیرهای ورودی شبکه (توصیف‌گر)	۲	FC	٪۵۶/۶۷	٪۵۶/۶۷	٪۵۶/۶۷	٪۵۶/۶۷	٪۵۶/۶۷
		FAR	٪۵۸/۸۲	٪۵۸/۸۲	٪۵۸/۸۲	٪۷۰/۵۹	٪۵۸/۸۲
		POD	٪۶۶/۶۷	٪۶۶/۶۷	٪۶۶/۶۷	٪۸۸/۸۹	٪۶۶/۶۷
	۳	FC	٪۶۰/۱۰۰	٪۶۶/۶۷	٪۶۳/۳۳	٪۶۳/۳۳	٪۶۳/۳۳
		FAR	٪۵۵/۵۶	٪۴۰/۱۰۰	٪۴۷/۳۷	٪۴۷/۳۷	٪۴۷/۳۷
		POD	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸
	۴	FC	٪۷۰/۱۰۰	٪۶۰/۱۰۰	٪۶۰/۱۰۰	٪۷۰/۱۰۰	٪۶۰/۱۰۰
		FAR	٪۳۳/۳۳	٪۵۵/۵۶	٪۵۵/۵۶	٪۳۳/۳۳	٪۵۵/۵۶
		POD	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸
	۵	FC	٪۶۳/۳۳	٪۶۰/۱۰۰	٪۶۰/۱۰۰	٪۶۰/۱۰۰	٪۶۰/۱۰۰
		FAR	٪۴۷/۳۷	٪۵۵/۵۶	٪۵۵/۵۶	٪۵۵/۵۶	٪۵۵/۵۶
		POD	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸
	۶	FC	٪۷۶/۶۷	٪۷۶/۶۷	٪۷۶/۶۷	٪۷۶/۶۷	٪۷۶/۶۷
		FAR	٪۲۱/۷۴	٪۲۱/۷۴	٪۲۱/۷۴	٪۲۱/۷۴	٪۲۱/۷۴
		POD	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸
	۷	FC	٪۸۰/۱۰۰	٪۸۰/۱۰۰	٪۸۰/۱۰۰	٪۸۰/۱۰۰	٪۷۶/۶۷
		FAR	٪۱۶/۶۷	٪۱۶/۶۷	٪۱۶/۶۷	٪۱۶/۶۷	٪۲۱/۷۴
		POD	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸	٪۷۷/۷۸

تعداد متغیرهای ورودی شبکه (توصیفگر)	۸	FC	%۸۶/۶۷	%۸۶/۶۷	%۸۶/۶۷	%۸۰/۰۰	%۸۶/۶۷
		FAR	%۱۱/۵۴	%۱۱/۵۴	%۱۱/۵۴	%۲۰/۸۳	%۱۱/۵۴
		POD	%۸۸/۸۹	%۸۸/۸۹	%۸۸/۸۹	%۸۸/۸۹	%۸۸/۸۹
	۹	FC	%۸۶/۶۷	%۸۳/۳۳	%۸۰/۰۰	%۸۶/۶۷	%۸۶/۶۷
		FAR	%۱۱/۵۴	%۱۶/۰۰	%۲۵/۰۰	%۱۱/۵۴	%۱۱/۵۴
		POD	%۸۸/۸۹	%۸۸/۸۹	%۱۰۰	%۸۸/۸۹	%۸۸/۸۹
	۱۰	FC	%۹۰/۰۰	%۹۰/۰۰	%۸۶/۶۷	%۸۶/۶۷	%۸۶/۶۷
		FAR	%۷/۴۱	%۷/۴۱	%۱۵/۳۸	%۱۵/۳۸	%۱۵/۳۸
		POD	%۸۸/۸۹	%۸۸/۸۹	%۱۰۰	%۱۰۰	%۱۰۰
	۱۱	FC	%۱۰۰	%۹۳/۳۳	%۱۰۰	%۱۰۰	%۹۶/۶۷
		FAR	%۰	%۰	%۰	%۰	%۳/۴۵
		POD	%۱۰۰	%۷۷/۷۸	%۱۰۰	%۱۰۰	%۱۰۰
	۱۲	FC	%۹۶/۶۷	%۹۳/۳۳	%۹۳/۳۳	%۱۰۰	%۹۶/۶۷
		FAR	%۳/۴۵	%۷/۱۴	%۰	%۰	%۰
		POD	%۱۰۰	%۱۰۰	%۷۷/۷۸	%۱۰۰	%۸۸/۸۹
	۱۳	FC	%۱۰۰	%۱۰۰	%۹۶/۶۷	%۱۰۰	%۱۰۰
		FAR	%۰	%۰	%۰	%۰	%۰
		POD	%۱۰۰	%۱۰۰	%۸۸/۸۹	%۱۰۰	%۱۰۰
	۱۴	FC	%۱۰۰	%۹۳/۳۳	%۸۶/۶۷	%۱۰۰	%۱۰۰
		FAR	%۰	%۷/۱۴	%۱۵/۳۸	%۰	%۰
		POD	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰
	۱۵	FC	%۱۰۰	%۹۶/۶۷	%۸۰/۰۰	%۹۰/۰۰	%۹۳/۳۳
		FAR	%۰	%۳/۴۵	%۲۵/۰۰	%۱۱/۱۱	%۷/۱۴
		POD	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰
	۱۶	FC	%۹۶/۶۷	%۹۶/۶۷	%۱۰۰	%۹۳/۳۳	%۹۳/۳۳
		FAR	%۳/۴۵	%۳/۴۵	%۰	%۷/۱۴	%۷/۱۴
		POD	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰
	۱۷	FC	%۸۶/۶۷	%۹۳/۳۳	%۹۳/۳۳	%۹۳/۳۳	%۳۳/۳۳
		FAR	%۱۵/۳۸	%۷/۱۴	%۷/۱۴	%۷/۱۴	%۲۰۰
		POD	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰	%۱۰۰

جدول (۲-۲۴) - توصیف‌گرهای بکار رفته در مدل‌سازی

شماره	نشانه	کلاس	معنی
۱	TE2	Charge	Topographic electronic descriptor (bond rescricted)
۲	GATS4v	2D autocorrelation	Geary autocorrelation-lag 4
۳	R7u-A	GETAWAY	R maximal autocorrelation of lag 7
۴	Mor03e	3D-MorSE	3D-MorSE-signal 3
۵	R6u-A	GETAWAY	R maximal autocorrelation of lag 6
۶	Mor8p	3D-MorSE	3D-MorSE-signal 8
۷	GATS5e	2D autocorrelation	Geary autocorrelation-lag 5
۸	RDF125m	RDF	Radial Distribution Function-12.5
۹	PJI3	Geometrical	3D Petijean shap index
۱۰	R3m-A	GETAWAY	R maximal autocorrelation of lag 3
۱۱	ATS7P	2D autocorrelation	Broto-Moreau autocorrelation of a topological structure-lag 7

جدول (۲-۲۵) - ماتریس همبستگی برای توصیف‌گرهای انتخاب شده

	TE2	GATS 4v	R7u-A	Mor 03e	R6u-A	Mor8p	GATS 5e	RDF 125m	PJI3	R3m- A	ATS 7P
TE2	۱										
GATS4v	۰/۲۲۷	۱									
R7u-A	-۰/۴۴۵	-۰/۲۱۰	۱								
Mor03e	-۰/۶۷۱	۰/۱۰۹	۰/۴۰۶	۱							
R6u-A	-۰/۳۹۹	-۰/۱۸۵	۰/۶۰۴	۰/۳۴۱	۱						
Mor8p	۰/۱۹۶	۰/۲۸۸	-۰/۳۰۸	-۰/۰۲۴	-۰/۳۴۸	۱					
GATS5e	-۰/۱۳۷	۰/۰۰۳	۰/۰۴۳	-۰/۰۳۷	-۰/۰۲۵	-۰/۰۹۳	۱				
RDF125m	۰/۷۱۶	۰/۰۰۴	-۰/۲۹۹	-۰/۷۰۶	-۰/۱۹۴	۰/۰۵۳	-۰/۲۰۴	۱			
PJI3	-۰/۰۱۷	۰/۲۸۹	-۰/۰۲۳	۰/۲۰۳	۰/۰۲۶	۰/۰۵۰	-۰/۰۷۱	۰/۰۴۶	۱		
R3m-A	-۰/۲۷۷	-۰/۱۳۰	۰/۱۷۶	۰/۰۲۶	۰/۲۵۸	-۰/۱۹۹	۰/۱۰۱	-۰/۱۱۱	۰/۰۸۲	۱	
ATS7P	-۰/۴۳۲	-۰/۲۰۶	۰/۰۹۸	۰/۱۵۹	۰/۱۹۹	-۰/۳۶۵	۰/۲۵۵	-۰/۳۷۵	-۰/۱۹۹	۰/۳۵۰	۱

جهت انتخاب تعداد دوره‌های آموزشی شبکه‌هایی با ۲ گره در لایه پنهان و با تعداد دوره‌های آموزش از ۴ تا ۱۵ طراحی و آموزش داده شدند و شبکه‌های آموزش داده شده برای پیش‌بینی نوع فعالیت ترکیبات سری ارزیابی بکار گرفته شدند. نتایج در جدول (۲-۲۶) آمده است. با توجه به نتایج جدول از دور آموزشی ۱۱ به بعد نتایج تغییر نکرده است و کمترین دور آموزشی که بهترین شاخص‌ها را ایجاد می‌کند ۸ می‌باشد. بنابراین تعداد دور آموزشی برابر با ۸، انتخاب گردید.

جدول (۲-۲۶) - نتایج ارزیابی شبکه‌هایی با ۲ نرون در لایه پنهان و تعداد دوره‌های آموزشی مختلف

شاخص‌های ارزیابی تعداد دور آموزشی	FC	FAR	POD
۴	٪۳۰/۰۰	٪۲۳۳/۳۳	٪۱۰۰
۵	٪۷۳/۳۳	٪۲۷/۲۷	٪۷۷/۷۸
۶	٪۸۰/۰۰	٪۱۶/۶۷	٪۷۷/۷۸
۷	٪۸۳/۳۳	٪۱۲/۰۰	٪۷۷/۷۸
۸	٪۱۰۰	٪۰	٪۱۰۰
۹	٪۱۰۰	٪۰	٪۱۰۰
۱۰	٪۱۰۰	٪۰	٪۱۰۰
۱۱	٪۹۳/۳۳	٪۰	٪۷۷/۷۸
۱۲	٪۹۳/۳۳	٪۰	٪۷۷/۷۸
۱۳	٪۹۳/۳۳	٪۰	٪۷۷/۷۸
۱۴	٪۹۳/۳۳	٪۰	٪۷۷/۷۸
۱۵	٪۹۳/۳۳	٪۰	٪۷۷/۷۸

با توجه به نتایج بهینه سازی شبکه، شبکه‌ی عصبی پیش‌خور سه لایه و آموزش پس انتشار با الگوریتم تنظیم بایزین انتخاب گردید. که این شبکه دارای ۱۱ متغیر ورودی، ۲ گره در لایه پنهان و دور آموزشی برابر ۸ می‌باشد.

۲-۲-۳-۳- ارزیابی شبکه برای طبقه‌بندی ترکیبات نکروستاتین-۵ با استفاده از سری

تست

بعد از انتخاب شبکه عصبی بهینه با استفاده از سری ارزیابی قدرت پیش‌بینی شبکه، بوسیله ۱۴ داده‌ی سری تست که در آموزش استفاده نشده‌اند، مورد بررسی قرار گرفت. نتایج حاصله در جدول (۲-۲۷) نشان داده شده است. نتایج بیانگر این است که شبکه قادر است که ۱۰۰٪ ترکیبات سری ارزیابی را به درستی گروه‌بندی کند.

جدول (۲-۲۷) - نتایج حاصل از ارزیابی مدل با استفاده از سری تست

شاخص‌های ارزیابی	FC	FAR	POD
سری تست	۱۰۰٪	۰٪	۱۰۰٪

۲-۲-۳-۳- مجموعه داده‌های مولکول‌های فعال

پس از اجرای کلاسه‌بندی و حصول اطمینان از توانایی مدل اجرا شده برای کلاسه‌بندی کامل ترکیبات، رابطه ساختار - فعالیت بازدارنده‌های نکروپتوسیس (Nec-5)، مجموعه داده‌های حاوی ۵۲ ترکیب فعال که ساختار و قدرت بازدارندگی (EC_{50}) آن‌ها در جدول (۲-۲۲) نشان داده شده است، مورد بررسی قرار گرفت.

۲-۲-۳-۴- محاسبه‌ی توصیف‌گرهای مولکولی

برای محاسبه توصیف‌گرها، ساختار بهینه‌شده‌ی هر مولکول به نرم‌افزار Dragon منتقل گردید و تمام ۱۴۸۱ توصیف‌گر برای هر مولکول محاسبه گردید.

۲-۳-۵- انتخاب توصیف‌گرهای مهم

انتخاب توصیف‌گرهای مهم که داده‌های EC_{50} را به ساختار مولکولی ارتباط دهند، مهمترین مرحله در مطالعات QSAR می‌باشد. جهت انتخاب توصیف‌گرهای مناسب که ارتباط بیشتری با داده‌های EC_{50} داشته باشند و بتوانند به عنوان ورودی‌های شبکه عصبی مصنوعی مورد استفاده قرار گرفته شوند، بسته‌ی نرم‌افزاری SPSS استفاده شد. بدین منظور مراحل زیر به ترتیب در استخراج پارامترهای مؤثر بکار گرفته شد:

الف) تمام توصیف‌گرهایی که مقادیر یکسانی حداقل برای ۹۵ درصد از ترکیبات مورد بررسی داشتند، حذف شدند.

ب) بین توصیف‌گرهای محاسبه شده ضریب همبستگی را برقرار کرده و از بین هر زوج توصیف‌گر که ضریب همبستگی بیش از ۹۰٪ دارند، یکی که ارتباط کمتری با متغیر وابسته داشت، حذف شد.

ج) برای انتخاب توصیف‌گرهای مناسب و مدل‌سازی، روش رگرسیون خطی چندگانه مرحله به مرحله در نرم افزار SPSS انجام گردید. نتایج حاصله نشان داد که از این تعداد زیاد توصیف‌گر، فقط تعداد ۸ توصیف‌گر ارتباط معنی‌داری با داده‌های EC_{50} ترکیبات مورد بررسی داشته‌اند.

۲-۳-۶- روند ایجاد مدل‌های QSAR

برای ایجاد مدل‌های توضیح دهنده‌ی رابطه‌ی ساختار - فعالیت از ۵۲ ترکیب فعال، ۱۰ ترکیب به طور تصادفی کنار گذاشته شد. مجموعه این ۱۰ ترکیب با توصیف‌گرهای مربوط به آن سری ارزیابی نامیده شد که ارزیابی خطای شبکه و بهینه‌سازی شبکه بوسیله‌ی این سری تحقق یافت. این سری برحسب شماره‌های مربوط در جدول (۲-۲۲) شامل ۲، ۹، ۱۳، ۱۵، ۲۴، ۳۰، ۳۷، ۴۰، ۴۵ و ۵۰ می‌باشند. سپس ۶ مورد نیز بطور تصادفی به عنوان سری تست کنار گذاشته شد که در مدل‌سازی از

آن استفاده نشد که عبارتند از: ۵، ۱۹، ۲۶، ۳۲، ۴۲ و ۴۸. از ۳۶ ترکیب باقیمانده برای آموزش شبکه عصبی مصنوعی و ایجاد مدل استفاده شد و سری آموزش نامیده شد.

۲-۳-۷- مدل شبکه عصبی مصنوعی و بهینه‌سازی آن

شبکه عصبی مصنوعی که روش مدل‌سازی غیرخطی است، جهت مطالعه‌ی روابط ساختار - فعالیت ترکیبات نکروستاتین-۵، انتخاب گردید. برنامه شبکه عصبی مصنوعی در نرم‌افزار MATLAB 7.5 نوشته شد. شبکه با الگوی آموزش پس‌انتشار و الگوریتم آموزشی تنظیم بایزین آموزش می‌یابد. تعداد نرون‌های لایه‌ی ورودی شبکه، برابر با تعداد توصیف‌گرهای موجود در مدل می‌باشد. پارامترهای قابل بهینه عبارتند از:

۱- تعداد لایه‌های پنهان

۲- تعداد متغیرهای ورودی شبکه

۳- نوع تابع تبدیل

۴- تعداد نرون‌ها در لایه پنهان

۵- تعداد دوره‌های آموزش

۲-۳-۷-۱- انتخاب تعداد لایه‌های شبکه عصبی

همانطور که در بخش قبل بیان شد، شبکه عصبی سه لایه متشکل از یک لایه ورودی، یک لایه پنهان و یک لایه‌ی خروجی بکار گرفته شد.

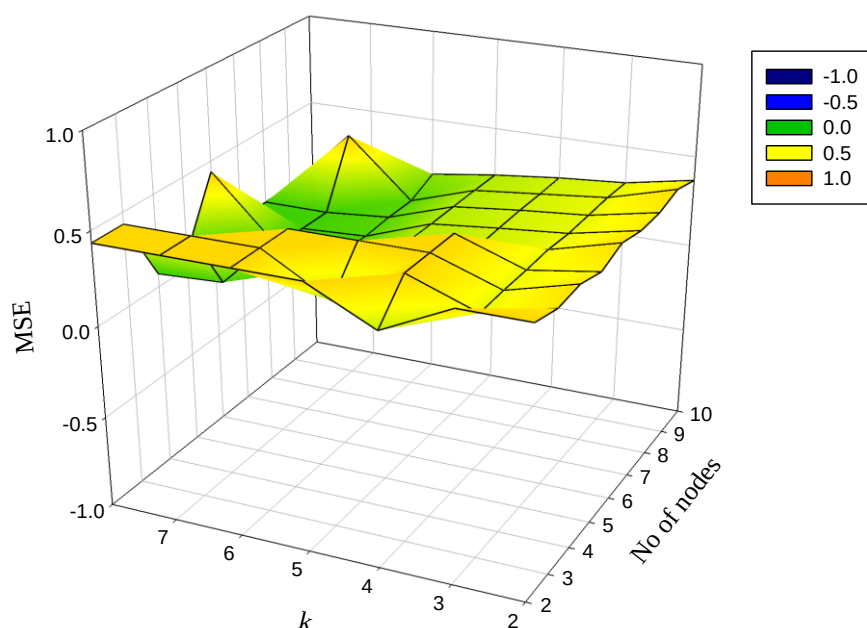
۲-۲-۳-۷-۲- انتخاب تعداد متغیرهای ورودی شبکه، انتخاب نوع تابع تبدیل و تعداد

گره‌های لایه پنهان

برای بهینه سازی و شناسایی بهترین پارامترهای شبکه، شبکه‌هایی با ورودی‌های از ۲ تا ۸ توصیف‌گر ایجاد شدند. هر شبکه با الگوریتم آموزشی تنظیم بایزین با تعداد گره‌های متفاوت در لایه پنهان آموزش داده شد. در حالی که برای بدست آوردن بهترین تابع انتقال در لایه‌ی پنهان مدل‌های شبکه عصبی، از توابع لگاریتم زیگموئید و تانژانت هیپربولیک به عنوان تابع فعالیت استفاده گردید. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطا به عنوان معیار انتخاب گردید. در تمام شبکه‌ها تعداد دورهای آموزشی یکسان و برابر ۱۰ در نظر گرفته شد. شبکه‌های آموزش یافته برای پیش‌بینی داده‌های $p(EC_{50} \times 10^{-6})$ سری ارزیابی مورد استفاده قرار گرفتند که نتایج حاصله در جداول (۲-۲۸) و (۲-۲۹)، آمده است. نمودار سه بعدی این جداول نیز در شکل‌های (۲-۱۸) و (۲-۱۹)، رسم شده است. با توجه به نتایج تعداد گره ۴ با تعداد متغیر ورودی ۸ و تابع تبدیل لگاریتم زیگموئید به عنوان شبکه‌ی بهینه انتخاب گردید. این توصیف‌کننده‌ها به همراه کلاس مربوطه در جدول (۲-۳۰) آمده است. ماتریس همبستگی بین توصیف‌گرها نیز در جدول (۲-۳۱) نشان داده شده است، که مشاهده می‌شود که بین توصیف‌گرها همبستگی وجود ندارد.

جدول (۲-۲۸) - مقادیر میانگین مربعات خطا برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم زیگموئید با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

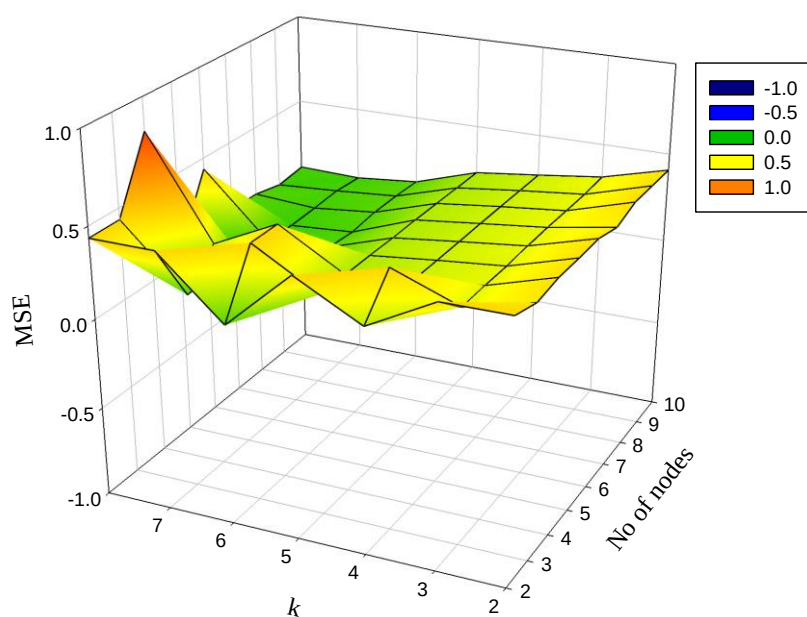
		تعداد متغیرهای ورودی شبکه (تعداد توصیف‌گر)						
		۲	۳	۴	۵	۶	۷	۸
تعداد گره‌ها	۲	۰/۴۳۷۴	۰/۴۳۷۳	۰/۲۵۹۱	۰/۴۳۷۳	۰/۴۳۷۳	۰/۴۳۷۵	۰/۴۳۷۳
	۳	۰/۳۹۵۵	۰/۳۰۳۶	۰/۴۳۷۳	۰/۲۳۰۴	۰/۴۳۷۳	۰/۴۳۷۳	۰/۴۳۷۳
	۴	۰/۴۰۸۱	۰/۳۰۹۸	۰/۴۳۷۴	۰/۴۳۷۳	۰/۴۳۷۳	۰/۰۸۹۰	۰/۰۷۱۴
	۵	۰/۳۶۴۶	۰/۳۰۹۱	۰/۴۳۷۳	۰/۲۲۹۱	۰/۱۲۱۲	۰/۱۲۶۵	۰/۰۷۹۵
	۶	۰/۴۱۲۲	۰/۳۱۴۱	۰/۲۶۳۷	۰/۲۳۳۹	۰/۱۱۲۴	۰/۰۸۴۷	۰/۴۴۰۹
	۷	۰/۳۷۷۲	۰/۳۰۱۲	۰/۲۶۸۹	۰/۲۳۲۲	۰/۱۰۲۹	۰/۰۸۵۶	۰/۰۷۸۰
	۸	۰/۳۷۶۸	۰/۳۰۳۲	۰/۲۷۲۲	۰/۲۳۳۲	۰/۱۱۴۱	۰/۰۸۳۲	۰/۰۸۴۸
	۹	۰/۴۱۰۲	۰/۳۰۷۱	۰/۲۶۹۸	۰/۲۳۳۳	۰/۱۱۵۷	۰/۴۳۹۹	۰/۰۸۷۶
	۱۰	۰/۳۶۸۸	۰/۲۹۸۷	۰/۲۷۱۰	۰/۲۳۱۹	۰/۱۸۱۴	۰/۰۸۵۸	۰/۰۷۳۱



شکل (۲-۱۸) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم زیگموئید

جدول (۲-۲۹) - مقادیر میانگین مربعات خطا برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت هیپربولیک با تعداد متغیرهای ورودی متفاوت و تعداد گره‌های مختلف در لایه پنهان

		تعداد متغیرهای ورودی شبکه (تعداد توصیف‌گر)						
		۲	۳	۴	۵	۶	۷	۸
تعداد گره‌ها	۲	۰/۴۳۷۳	۰/۴۳۷۲	۰/۲۴۳۶	۰/۴۳۷۵	۰/۱۰۹۱	۰/۴۳۴۷	۰/۴۳۷۳
	۳	۰/۳۹۱۵	۰/۲۹۹۷	۰/۴۳۷۲	۰/۲۲۹۲	۰/۴۳۷۲	۰/۰۹۳۸	۰/۴۳۷۲
	۴	۰/۴۱۱۳	۰/۳۰۲۹	۰/۲۶۷۸	۰/۲۲۸۵	۰/۴۳۷۲	۰/۰۸۷۱	۰/۰۸۱۸
	۵	۰/۴۰۷۴	۰/۳۰۰۱	۰/۲۶۷۱	۰/۲۲۹۶	۰/۱۰۵۴	۰/۰۹۱۳	۰/۰۷۱۵
	۶	۰/۴۰۸۷	۰/۳۰۶۳	۰/۲۶۷۵	۰/۲۲۶۶	۰/۱۱۱۴	۰/۰۹۰۰	۰/۴۳۷۶
	۷	۰/۳۷۰۰	۰/۳۰۸۳	۰/۲۶۹۲	۰/۲۲۹۵	۰/۰۹۹۴	۰/۰۷۶۹	۰/۱۱۰۴
	۸	۰/۴۱۲۱	۰/۳۰۵۳	۰/۲۶۹۴	۰/۲۲۸۶	۰/۱۱۸۲	۰/۰۸۴۰	۰/۱۰۹۶
	۹	۰/۴۰۹۸	۰/۳۰۴۱	۰/۲۶۸۱	۰/۲۲۶۹	۰/۱۲۰۷	۰/۰۹۴۵	۰/۰۷۴۴
	۱۰	۰/۴۰۸۰	۰/۳۱۲۶	۰/۲۶۷۹	۰/۲۲۶۴	۰/۱۱۴۹	۰/۰۸۳۰	۰/۰۹۴۸



شکل (۲-۱۹) - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت هیپربولیک

جدول (۲-۳۰) - توصیف‌گرهای بکار رفته در مدل‌سازی

No	symbol	Class	Meaning
۱	H7u	GETAWAY	H autocorrelation of lag 7
۲	Ke	WHIM	K global shap index
۳	MATS7e	2D autocorrelation	Moran autocorrelation-lag 7
۴	PJ13	Geometrical	3D Petijean shap index
۵	RDF140u	RDF	Radial Distribution Function-14.0
۶	RDF020p	RDF	Radial Distribution Function-2.0
۷	RDF090v	RDF	Radial Distribution Function-9.0
۸	nBnz	Constitutional	Number of benzene

جدول (۲-۳۱) - ماتریس همبستگی برای توصیف‌گرهای انتخاب شده

	H7u	Ke	MATS7e	PJ13	RDF140u	RDF020p	RDF090v	nBnz
H7u	۱							
Ke	۰/۳۶۶	۱						
MATS7e	-۰/۳۴۵	-۰/۳۱۶	۱					
PJ13	۰/۰۳۱	۰/۳۳۸	۰/۰۳۷	۱				
RDF140u	۰/۱۴۴	۰/۲۹۹	۰/۰۱۹	-۰/۱۶۳	۱			
RDF020p	۰/۴۳۰	۰/۳۱۹	-۰/۶۳۷	۰/۰۹۲	۰/۴۱۹	۱		
RDF090v	۰/۵۳۱	۰/۳۵۹	-۰/۵۷۶	-۰/۲۴۹	۰/۳۲۸	۰/۶۴۳	۱	
nBnz	-۰/۱۲۴	۰/۰۸۹	۰/۰۵۲	-۰/۰۸۷	-۰/۲۰۲	-۰/۲۰۲	۰/۰۱۵	۱

۲-۲-۳-۷-۳- انتخاب تعداد دوره‌های آموزش

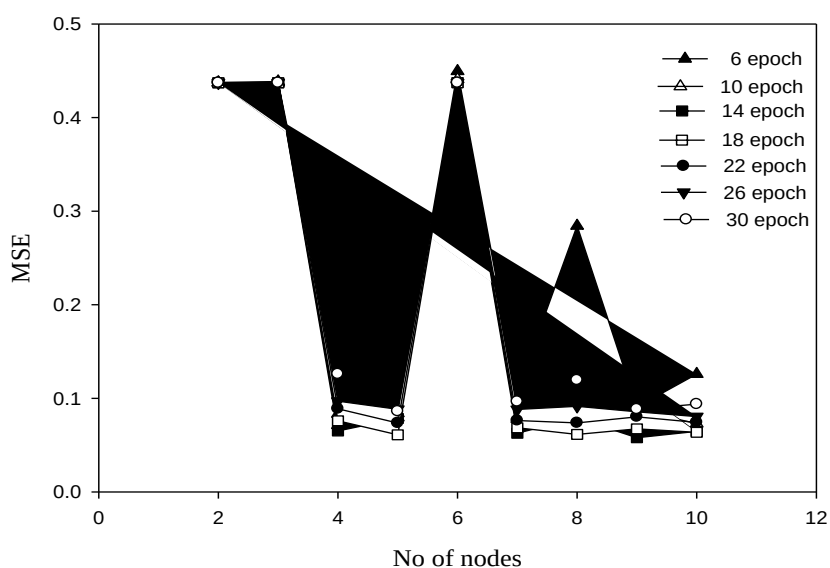
به منظور انتخاب تعداد دوره‌های آموزشی شبکه‌هایی با تعداد گره‌های لایه پنهان در

محدوده‌ی ۲ تا ۱۰ و با تعداد دوره‌های آموزش مختلف طراحی و آموزش داده شدند و شبکه‌های

آموزش داده شده در پیش‌بینی داده‌های EC_{50} بکار گرفته شدند. نتایج در جدول (۲-۳۲) و شکل (۲-۲۰) آمده است. بنابراین تعداد دور آموزشی برابر با ۱۴ که حداقل میانگین مربعات خطا را ایجاد می‌کند، انتخاب گردید.

جدول (۲-۳۲) - مقادیر میانگین مربعات خطا برای تعداد گره‌های مختلف در لایه پنهان

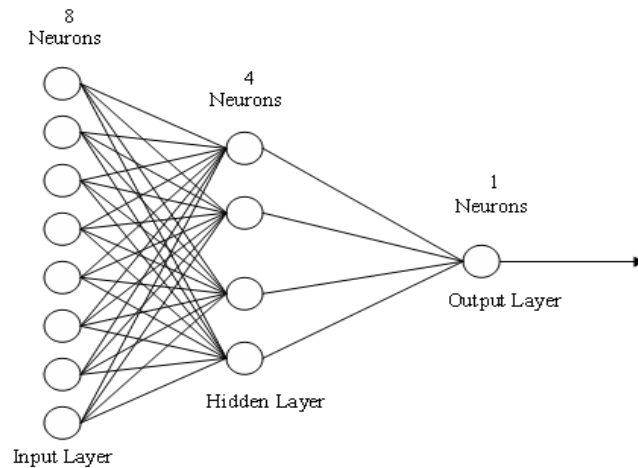
تعداد گره	تعداد دورهای آموزشی						
	۶	۱۰	۱۴	۱۸	۲۲	۲۶	۳۰
۲	۰/۴۳۷۶	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲
۳	۰/۴۳۸۶	۰/۴۳۷۳	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲
۴	۰/۰۸۳۴	۰/۰۷۱۴	۰/۰۶۵۱	۰/۰۷۶۰	۰/۰۸۸۹	۰/۰۹۷۰	۰/۱۲۶۰
۵	۰/۰۸۴۱	۰/۰۷۹۵	۰/۰۷۸۶	۰/۰۶۱۱	۰/۰۷۳۶	۰/۰۸۸۳	۰/۰۹۶۲
۶	۰/۴۴۹۶	۰/۴۴۰۹	۰/۴۳۷۴	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲	۰/۴۳۷۲
۷	۰/۱۰۸۲	۰/۰۷۸۰	۰/۰۶۲۷	۰/۰۶۸۵	۰/۰۷۶۳	۰/۰۸۸۱	۰/۰۹۶۵
۸	۰/۲۸۴۳	۰/۰۸۴۸	۰/۰۷۸۸	۰/۰۶۱۵	۰/۰۷۳۷	۰/۰۹۱۵	۰/۱۱۹۴
۹	۰/۰۹۰۷	۰/۰۸۷۶	۰/۰۵۸۰	۰/۰۶۷۴	۰/۰۸۰۲	۰/۰۸۵۶	۰/۰۸۸۴
۱۰	۰/۱۲۶۰	۰/۰۷۳۱	۰/۰۶۴۷	۰/۰۶۳۷	۰/۰۷۴۳	۰/۰۸۰۵	۰/۰۹۳۸



شکل (۲-۲۰) - تغییرات مقادیر MSE بر حسب تعداد گره‌های مختلف در لایه پنهان

۴-۷-۳-۲-۲- هندسه‌ی شبکه عصبی مصنوعی بهینه شده

با توجه به نتایج حاصل در بهینه‌سازی شبکه، شبکه‌ی عصبی پیش‌خور سه لایه و آموزش پس انتشار با الگوریتم تنظیم بایزین انتخاب گردید. ساختار هندسی شبکه بهینه شده در شکل (۲-۲) و پارامترهای بهینه‌ی آن در جدول (۲-۳۳) آمده است. سرعت آموزش بسیار بالا است.



شکل (۲-۲۱) - ساختار هندسی شبکه عصبی استفاده شده

جدول (۲-۳۳) - پارامترهای شبکه‌ی بهینه

۱۴	تعداد دوره‌های آموزش
e^{-10}	مینیمم شیب عملکرد
۰/۰۰۵	پارامتر تطبیق مارکورت (μ)
۰/۱	فاکتور کاهش برای پارامتر تطبیق مارکورت
۱۰	فاکتور افزایش برای پارامتر تطبیق مارکورت
$1e^{-10} \times$	مقدار ماکزیمم برای پارامتر تطبیق مارکورت

۲-۳-۸- ارزیابی مدل ارائه شده

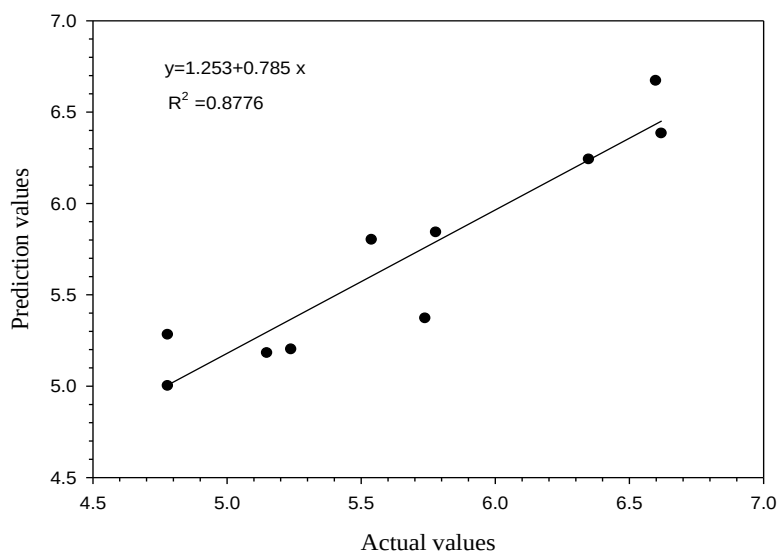
در ارزیابی مدل ارائه شده در این تحقیق برای پیش‌بینی داده‌های EC₅₀ بازدارنده‌های یک نوع مرگ سلول (نکروپتوزیس)، روش‌های زیر بکار گرفته شد:

۲-۳-۸-۱- ارزیابی شبکه با استفاده از سری ارزیابی

به این منظور، شبکه عصبی با استفاده از ۳۶ مولکول سری آموزش، آموزش داده شد. معیار برای بهینه‌سازی شبکه کمترین مقدار میانگین مربعات خطا در سری ارزیابی (شامل ۱۰ مولکول که بطور تصادفی انتخاب شده‌اند) می‌باشد. نتایج حاصله در جدول (۲-۳۴) نشان داده شده است. ترسیمی از مقادیر پیش‌بینی شده بر حسب مقدار واقعی نیز در شکل (۲-۲۲) نشان داده شده است.

جدول (۲-۳۴) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری ارزیابی

درصد خطا	مقدار پیش‌بینی	مقدار واقعی	شماره
۳/۶۳	۶/۳۸	۶/۶۲	۲
۶/۴۵	۵/۳۷	۵/۷۴	۹
-۱۰/۴۶	۵/۲۸	۴/۷۸	۱۳
۰/۷۶	۵/۲۰	۵/۲۴	۱۵
۱/۷۳	۶/۲۴	۶/۳۵	۲۴
-۰/۵۸	۵/۱۸	۵/۱۵	۳۰
-۱/۰۴	۵/۸۴	۵/۷۸	۳۷
-۱/۰۶	۶/۶۷	۶/۶۰	۴۰
-۴/۶۰	۵/۰۰	۴/۷۸	۴۵
-۶/۸۱	۵/۸۰	۵/۴۳	۵۰



شکل (۲۲-۲) - ترسیمی از مقادیر پیش‌بینی شده با شبکه عصبی مصنوعی بر حسب مقدار واقعی در سری ارزیابی

۲-۲-۳-۸-۲- ارزیابی شبکه با استفاده از سری تست

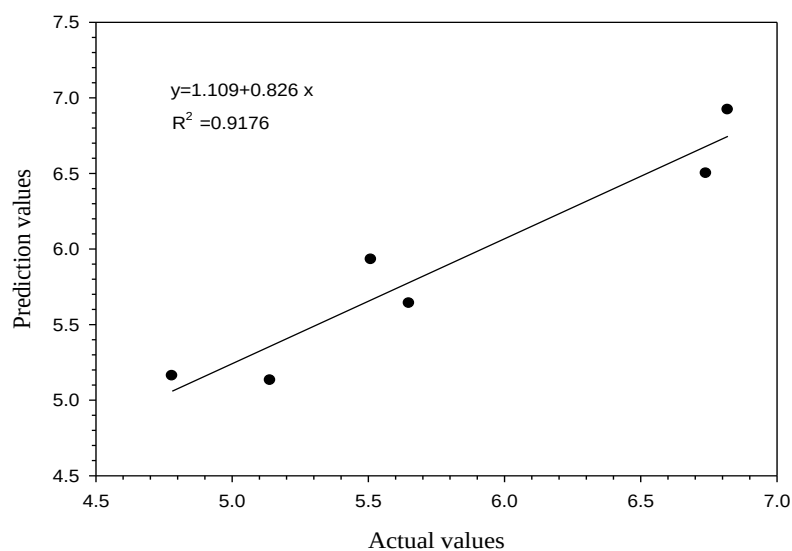
بعد از انتخاب شبکه عصبی بهینه با استفاده از سری ارزیابی قدرت پیش‌بینی شبکه، بوسیله

۶ داده سری تست که در آموزش استفاده نشده‌اند، مورد بررسی قرار گرفت. نتایج حاصله در جدول

(۳۵-۲) و همچنین شکل (۲۳-۲) نشان داده شده است.

جدول (۳۵-۲) - نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از سری تست

شماره	مقدار واقعی	مقدار پیش‌بینی	درصد خطا
۵	۵/۶۵	۵/۶۴	۰/۱۸
۱۹	۴/۷۸	۵/۱۶	-۷/۹۵
۲۶	۶/۷۴	۶/۵۰	۳/۵۶
۳۲	۵/۱۴	۵/۱۳	۰/۱۹
۴۲	۶/۸۲	۶/۹۲	-۱/۴۷
۴۸	۵/۵۱	۵/۹۳	-۷/۶۲



شکل (۲-۲۳) - ترسیمی از مقادیر پیش‌بینی شده بر حسب مقدار واقعی در سری تست

۲-۱-۳-۸-۳- ارزیابی شبکه به روش رد مرحله‌ای تک تک

نتایج حاصل از ارزیابی شبکه با این روش در جدول (۲-۳۶) نشان داده شده است. شکل

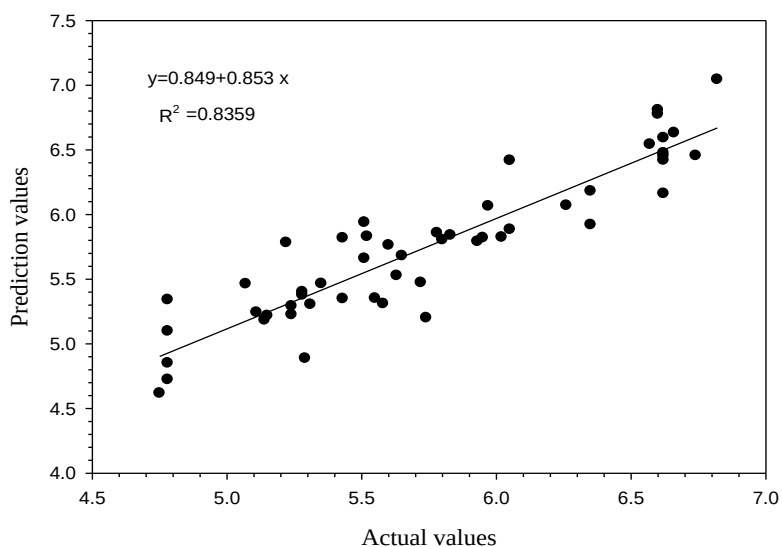
(۲-۲۴) هم مقادیر پیش‌بینی شده بر حسب مقادیر واقعی را نشان می‌دهد. نتایج بیانگر این است که

مدل ارائه شده خطای پیش‌بینی کم و قدرت تعمیم بالایی دارد. لذا با این مدل می‌توان داده‌های

EC₅₀ را برای هر ترکیب جدید با خطای کم و قابل قبول پیش‌بینی کرد.

جدول (۲-۳۶) - نتایج حاصل از ارزیابی مدل با استفاده از روش رد تک تک

درصد خطا	مقدار پیش‌بینی	مقدار واقعی	شماره	درصد خطا	مقدار پیش‌بینی	مقدار واقعی	شماره
۶/۹۵	۶/۱۶	۶/۶۲	۱	۳/۰۲	۶/۴۲	۶/۶۲	۲۷
۲/۵۷	۶/۴۵	۶/۶۲	۲	۰/۴۵	۶/۵۹	۶/۶۲	۲۸
-۱۰/۷۳	۵/۷۸	۵/۲۲	۳	-۲/۸۶	۵/۷۶	۵/۶۰	۲۹
-۲/۲۷	۵/۴۰	۵/۲۸	۴	-۱/۳۶	۵/۲۲	۵/۱۵	۳۰
-۰/۵۳	۵/۶۸	۵/۶۵	۵	۰/۱۹	۵/۳۰	۵/۳۱	۳۱
۲/۷۵	۴/۶۲	۴/۷۵	۶	-۰/۷۸	۵/۱۸	۵/۱۴	۳۲
-۰/۱۷	۵/۸۴	۵/۸۳	۷	-۲/۰۶	۵/۴۶	۵/۳۵	۳۳
۳/۰۴	۶/۰۷	۶/۲۶	۸	۲/۸۱	۵/۸۸	۶/۰۵	۳۴
۹/۴۱	۵/۲۰	۵/۷۴	۹	۴/۸۴	۵/۳۱	۵/۵۸	۳۵
۲/۲۶	۶/۴۷	۶/۶۲	۱۰	-۵/۶۲	۵/۸۳	۵/۵۲	۳۶
۷/۵۶	۴/۸۹	۵/۲۹	۱۱	-۱/۳۸	۵/۸۶	۵/۷۸	۳۷
-۷/۶۹	۵/۴۶	۵/۰۷	۱۲	۴/۳۷	۵/۴۷	۵/۷۲	۳۸
-۱۱/۷۲	۵/۳۴	۴/۷۸	۱۳	۲/۳۶	۵/۷۹	۵/۹۳	۳۹
-۰/۹۵	۵/۲۹	۵/۲۴	۱۴	-۲/۷۳	۶/۷۸	۶/۶۰	۴۰
۰/۳۸	۵/۲۲	۵/۲۴	۱۵	۰/۴۵	۶/۶۳	۶/۶۶	۴۱
-۶/۱۲	۶/۴۲	۶/۰۵	۱۶	-۳/۲۲	۷/۰۴	۶/۸۲	۴۲
۱/۷۸	۵/۵۳	۵/۶۳	۱۷	-۳/۱۸	۶/۸۱	۶/۶۰	۴۳
۰/۰۰	۵/۸۰	۵/۸۰	۱۸	۳/۶۰	۵/۳۵	۵/۵۵	۴۴
۱/۲۵	۴/۷۲	۴/۷۸	۱۹	-۶/۶۹	۵/۱۰	۴/۷۸	۴۵
-۱/۴۶	۴/۸۵	۴/۷۸	۲۰	-۲/۷۲	۵/۶۶	۵/۵۱	۴۶
۶/۷۷	۵/۹۲	۶/۳۵	۲۱	-۰/۴۲	۶/۵۴	۶/۵۷	۴۷
-۱/۸۹	۵/۳۸	۵/۲۸	۲۲	-۷/۸۰	۵/۹۴	۵/۵۱	۴۸
-۱/۵۱	۶/۰۶	۵/۹۷	۲۳	-۲/۵۴	۵/۲۴	۵/۱۱	۴۹
۲/۶۸	۶/۱۸	۶/۳۵	۲۴	-۷/۱۸	۵/۸۲	۵/۴۳	۵۰
۳/۳۲	۵/۸۲	۶/۰۲	۲۵	۱/۴۷	۵/۳۵	۵/۴۳	۵۱
۴/۳۰	۶/۴۵	۶/۷۴	۲۶	۲/۱۸	۵/۸۲	۵/۹۵	۵۲

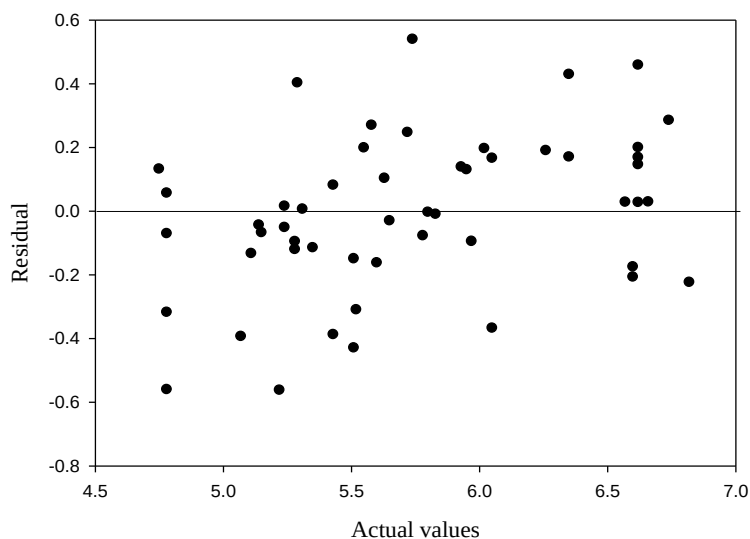


شکل (۲-۲۴) - ترسیمی از مقادیر پیش‌بینی شده با شبکه عصبی مصنوعی بر حسب مقدار واقعی برای کل داده‌ها

باقیمانده مقادیر محاسبه شده از مقادیر تجربی بر حسب مقادیر تجربی برای مولکول‌ها در

شکل (۲-۲۵) رسم شده است. توزیع نسبتاً یکنواخت و اتفاقی باقی‌مانده‌ها در دو طرف خط صفر

نشان‌دهنده‌ی این است که در مدل حاصل از شبکه‌ی عصبی مصنوعی، خطای معین وجود ندارد.



شکل (۲-۲۵) - نمودار باقیمانده‌ها برای کل داده‌ها (باقیمانده‌ی مقادیر پیش‌بینی شده‌ی PEC_{50} توسط شبکه عصبی

مصنوعی از مقادیر واقعی بر حسب مقادیر تجربی برای کل ترکیبات)

۲-۱-۳-۸-۴-آزمون Y - تصادفی

برای حصول اطمینان از تصادفی نبودن نتایج، یک مجموعه اعداد تصادفی در دامنه‌ی متغیر هدف تعریف کرده، بهترین شبکه با آن آزموده شد. لذا داده‌های $\log EC_{50} \times 10^{-6}$ (متغیر وابسته بطور تصادفی ایجاد شدند. مدل QSAR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از بازدارنده‌ها توسعه یافت. نتایج حاصل از چندین مجموعه‌ی تصادفی ایجاد شده، در جدول (۳۷-۲) نشان داده شده است. با توجه به نتایج، هیچ وابستگی برجسته‌ای بین خواص ساختاری و داده‌های بازدارندگی برای ترکیبات وجود ندارد (ضریب همبستگی پایین)، که نشان‌دهنده‌ی عدم همبستگی تصادفی در مدل شبکه عصبی مصنوعی ایجاد شده می‌باشد.

جدول (۳۷-۲) - مقادیر R^2 برای سری ارزیابی و سری تست بعد از چندین تست Y - تصادفی

تکرار	R^2 سری ارزیابی (n=10)	R^2 سری تست (n=6)
۱	۰/۲۵	۰/۱۴
۲	۰/۰۱	۰/۰۰
۳	۰/۳۵	۰/۰۳
۴	۰/۰۲	۰/۰۶
۵	۰/۰۰	۰/۰۳
۶	۰/۰۶	۰/۱۸
۷	۰/۴۱	۰/۱۵
۸	۰/۰۷	۰/۰۵
۹	۰/۰۳	۰/۱۷
۱۰	۰/۱۲	۰/۰۸

۲-۱-۳-۸-۵- ارزیابی مدل شبکه عصبی ارائه شده با استفاده از پارامترهای آماری

برای ارزیابی کارایی شبکه عصبی مصنوعی ارائه شده برای پیش‌بینی مهارکنندگی ترکیبات

نکروستاتین-۵ چهار پارامتر آماری MAE ، REP ، MSE و R^2 محاسبه شد - در بخش (۲-۱-۳-۵-۱)

توضیح داده شد - و نتایج حاصل در جدول (۲-۳۸) آمده است. این نتایج حاکی از قدرت پیش‌بینی

بالای مدل غیرخطی شبکه‌ی عصبی می‌باشد.

جدول (۲-۳۸) - پارامترهای آماری محاسبه شده

پارامتر	سری ارزیابی (n = 10)	سری تست (n = 6)	کل داده‌ها (n = 52)
MAE	۳/۴۹	۳/۴۶	۳/۳۸
REP	۴/۲۶	۴/۳۸	۴/۲۵
MSE	۰/۰۶۵۱	۰/۰۷۶۷	۰/۰۵۹۸
R^2	۰/۸۷۷۶	۰/۹۱۷۶	۰/۸۳۵۹

فصل سوم

بحث و نتیجه گیری

۳-۱- بررسی نتایج مدل سازی بازدارنده‌ی تیروزین کیناز

در مطالعه‌ی روابط کمی ساختار - فعالیت برای پیش‌بینی IC₅₀ مهارکننده‌های تیروزین کیناز با کار دوباره روی مجموعه‌ای از داده‌های آماده به نتایجی متفاوت با آنچه شی^۱ و همکارانش [۵۲] در روش بهینه‌سازی کلونی مورچه اصلاح شده^۲ (ACO) و رگرسیون خطی چندگانه به دست آورده بودند، رسیدیم. الگوریتم (ACO) اولین بار توسط دوریگو^۳ و همکارانش از رفتار کلونی‌های واقعی مورچه الهام گرفته شد [۵۲]. شی و همکارانش از آن، جهت انتخاب توصیف‌گرهای مناسب استفاده کردند و روابط کمی ساختار - فعالیت ترکیبات را با استفاده از مدل رگرسیون خطی چندگانه مورد بررسی قرار دادند. در تحقیق حاضر مدلی بر مبنای شبکه عصبی مصنوعی (غیرخطی) برای پیش‌بینی فعالیت مهارکنندگی این ترکیبات ارائه گردیده است.

۳-۱-۱- توصیف‌گرهای انتخاب شده

توصیف‌گرهای مولکولی به روش‌های مختلفی دسته‌بندی می‌شوند. ساده‌ترین روش برای دسته‌بندی توصیف‌کننده‌ها بر اساس تجربی یا تئوری بودن آن‌ها است. دسته‌بندی دیگر بر اساس ابعاد توصیف مولکول است. این نوع دسته‌بندی بصورت زیر است:

توصیف‌کننده‌های صفربعدی

توصیف‌کننده‌های صفربعدی در برگیرنده‌ی توصیف‌کننده‌های اساسی هستند که مستقل از صورتبندی مولکولی می‌باشند و از فرمول شیمیایی مولکول مشتق می‌شوند. این توصیف‌کننده‌ها به توصیف‌کننده‌های ساختاری نیز معروفند که تعداد و نوع اتم‌ها، جرم مولکولی و هر عاملی از خصوصیات اتمی را تعیین می‌کنند. همچنین شامل توصیف‌کننده‌های مربوط به نوع پیوندها، تعداد

¹ - Shi

² - Modified ant colony optimization

³ - Dorigo

پیوندها (ساده، دوگانه، سه‌گانه، آروماتیک) و تعداد کل حلقه‌ها بر اساس تعداد اتم‌های آن (حلقه‌های شش‌تایی، پنج‌تایی و ...)، تعداد گروه‌های عاملی خاص و تعداد یک عنصر شیمیایی خاص (C, N, O, H, F,...) در مولکول نیز می‌شوند.

توصیف‌کننده‌های یک‌بعدی

توصیف‌گرهای یک‌بعدی خواصی هستند که به دانش توپولوژی و ساختار سه بعدی ترکیبات نیاز ندارد و به خصوصیات کلی مولکول ارتباط دارد. توصیف‌گرهای یک بعدی شامل گروه‌های عاملی، قطعات اتم مرکزی^۱، پارامترهای تجربی و خصوصیات مولکولی محاسبه‌شده از مدل‌های رگرسیون می‌باشند.

توصیف‌کننده‌های دوبعدی

در نمایش دوبعدی از مولکول، چگونگی اتصال اتم‌ها در مولکول بر حسب حضور و طبیعت پیوندهای شیمیایی، در نظر گرفته می‌شود. توصیف‌کننده‌های دوبعدی در برگیرنده‌ی توصیف‌کننده‌های ساختمانی (اندازه، شکل، تقارن، شاخه‌دار شدن، حلقوی بودن)، خودهمبستگی دوبعدی، شمارش‌های مسیر مولکولی^۲، شاخص‌های بار، مقادیر مثبت و منفی ماتریس بارگیری می‌باشند.

توصیف‌کننده‌های سه‌بعدی

دسته‌ی دیگری از توصیف‌کننده‌های مولکولی شامل توصیف‌کننده‌های سه‌بعدی هستند که با شروع از یک نمایش هندسی یا سه‌بعدی مولکول محاسبه می‌شوند و نه تنها یک نمایش از طبیعت و اتصال اتم‌ها بلکه از صورتبندی فضایی مولکول نیز فراهم می‌کند. شاخص‌های آروماتیکی،

^۱ - Atom-centered fragments

^۲ - Molecular walk counts

توصیف‌کننده‌های هندسی، توصیف‌کننده‌های مربوط به بار، پروفایل مولکولی، توابع توزیع شعاعی، توصیف‌کننده‌های شامل شاخص‌های آماری اجزاء اصلی و توصیف‌کننده‌های محاسبه‌شده از عناصر ماتریس‌های مربوط به نفوذ مولکولی در دسته توصیف‌کننده‌های سه‌بعدی قرار می‌گیرند.

مهمتر از قدرت پیش‌گویی میزان فعالیت ترکیبات جدید، رسالت دیگر مدل‌های QSAR ایجاد ارتباط بین این فعالیت و پارامترهای ساختاری یا توصیف‌گرها است. پس از انتخاب توصیف‌گرهای مناسب در نهایت ۹ توصیف‌گر که شامل ۵ گروه مختلف هستند، از میان کل توصیف‌گرها در مدل‌سازی شبکه‌ی عصبی مورد استفاده قرار گرفت که لیست کامل آن‌ها در جدول (۲-۱۲) آمده است. در اینجا به بررسی علت اهمیت این توصیف‌گرها و رابطه‌ی احتمالی آن‌ها با فعالیت مهارکنندگی ترکیبات مورد بررسی، می‌پردازیم.

الف- توصیف‌گرهای گروه 2D Autocorrelation [۵۸]

این توصیف‌کننده‌ها براساس تابع خود ارتباطی AC_l تعریف می‌شوند:

$$AC_l = \int_a^b f(x) \cdot f(x+l) \cdot dx \quad \text{رابطه‌ی (۱-۳)}$$

که $f(x)$ هر تابع از متغیر x و l اختلاف فاصله از x می‌باشد؛ a و b فاصله‌ی مطالعه شده را بیان می‌کند. $f(x)$ معمولاً یک تابع وابسته به زمان مانند یک سیگنال الکتریکی یا وابسته به فضا مانند دانسیته جمعیت در فضا می‌باشد. این گروه از توصیف‌گرها به بررسی اطلاعات مختلفی در مورد ترکیبات می‌پردازد که از دایره‌ی بحث این نوشتار خارج است و ما فقط به بررسی دو نوع از آن‌ها که در مدل انتخاب شدند، اکتفا می‌کنیم. نوعی از این توصیف‌گرها به نام Moran Autocorrelation است که به این شکل محاسبه می‌شود:

$$I(d) = \frac{\frac{1}{\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - \bar{w}) \cdot (w_j - \bar{w})}{\frac{1}{A} \sum_{i=1}^A (w_i - \bar{w})} \quad \text{رابطه‌ی (۲-۳)}$$

که w یک ویژگی اتم، $\bar{w}$ میانگین مقدار آن روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیک است و δ_{ij} نیز - که به تابع کرونکر^۱ معروف است - در حالتی که $d_{ij} = d$ باشد یک است، و گرنه صفر است و Δ هم مجموعه δ هاست. این توصیف‌گرها مقادیری بین $+1$ و -1 دارند.

نوع دیگر این گروه از توصیف‌گرها گروه Geary Autorrelation است که به این شکل

محاسبه می‌شود:

$$C(d) = \frac{\frac{1}{2\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - w_j)^2}{\frac{1}{A-1} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad \text{رابطه‌ی (۳-۳)}$$

که پارامترها مشابه رابطه‌ی (۲-۳) هستند. مقدار این توصیف‌گر که از جنس فاصله است از صفر تا بی‌نهایت متغیر است.

از این نوع توصیف‌گرها MATS1e^۲ و GATS6e^۳ توسط مدل رگرسیون خطی چندگانه انتخاب شد. این توصیف‌گرها نقش خاصیت فیزیکولوژی الکترون‌گاتیویته‌ی ساندerson را در ترکیبات نشان می‌دهند. وجود ضریب منفی در مدل برای توصیف‌گر GATS6e نشان می‌دهد که افزایش مقدار این توصیف‌گر سبب کاهش PIC₅₀ ترکیبات مورد بررسی می‌شود.

ب- توصیف‌گرهای گروه *Topological* [۵۸]

این توصیف‌گرها بر اساس نمایش گراف مولکول می‌باشند. در این گراف‌ها هر نقطه نشان‌دهنده یک اتم بوده و خطوط بین نقاط نیز نشان‌دهنده‌ی پیوند شیمیایی بین اتم‌ها می‌باشد. معمولاً در گراف‌های مولکولی اتم هیدروژن را نشان نمی‌دهند. این توصیف‌گرها اطلاعاتی راجع به ساختمان، اندازه، شکل، تقارن، شاخه‌دار شدن، نحوه‌ی اتصال اتم‌ها و نوع اتم‌های موجود در یک

^۱ - Kronecker

^۲ - Moran autocorrelation-lag 1 weighted by Sanderson electronegativities

^۳ - Geary autocorrelation-lag 6 weighted by atomic Sanderson electronegativities

مولکول در اختیار ما قرار می‌دهند. محاسبه‌ی این توصیف‌گرها به سادگی از روی ساختمان دوبعدی مولکول‌ها امکان‌پذیر می‌باشد. از این گروه توصیف‌گرها X5sol^۱ انتخاب شد که اندیس‌های ارتباطی حلال‌پوشی نام دارند. به طور کلی این توصیف‌گرهای مولکولی آنتروپی حلال‌پوشی و برهمکنش‌های پراکندگی در محلول را شرح می‌دهند که به صورت زیر تعریف می‌شود:

$${}^mX^s = \frac{1}{2^{m+1}} \cdot \sum_{k=1}^K \frac{(\prod_{a=1}^n L_a)_k}{(\prod_{a=1}^n \delta_a)_k}^{\frac{1}{2}} \quad \text{رابطه‌ی (۴-۳)}$$

که L_a عدد کوانتومی اصلی (۲ برای اتم C، N و O و ۳ برای Si، S، Cl، ...) مربوط به a امین اتم در k امین زیرگراف و δ_a درجه‌ی رأس؛ K تعداد کل زیرگراف‌های m ام؛ n تعداد یال‌ها در زیرگراف می‌باشد (در این مورد $m=1$ می‌باشد).

با توجه به اثر متوسط مثبت این توصیف‌گر، با افزایش مقدار این توصیف‌گر، قدرت مهارکنندگی ترکیب افزایش یافته است. همچنین طبق تعریف این توصیف‌گر (رابطه‌ی (۴-۳)) با بزرگتر شدن مولکول‌ها (افزایش تعداد اتم‌ها) بدلیل بزرگ‌شدن صورت کسر، مقدار این توصیف‌گر و به طور مستقیم PIC₅₀ افزایش یابد. ظاهرشدن این توصیف‌گر نشان‌دهنده‌ی اهمیت آنتروپی حلال‌پوشی در فعالیت دارویی این ترکیبات می‌باشد.

ج- توصیف‌گرهای گروه GETAWAY^۲ [۵۹]

این توصیف‌گرها با توجه به مختصات فضایی اتم‌ها در یک مولکول به راحتی قابل محاسبه هستند. این توصیف‌کننده‌ها بر اساس ماتریس قدرت نفوذ^۳ تعریف می‌شوند و بیان‌کننده‌ی

^۱ - Solvation connectivity index chi-5

^۲ - Geometry, Topology, and Atom-Weights Assembly

^۳ - Leverage matrix

ویژگی‌های هندسی و توپولوژیک مولکول‌ها هستند. ماتریس قدرت نفوذ یا ماتریس تأثیر مولکول^۱ (MIM) با رابطه‌ی زیر بیان می‌شود.

$$H = M \cdot (M^T \cdot M)^{-1} \cdot M^T \quad \text{رابطه‌ی (۵-۳)}$$

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

توصیف‌گرهای GETAWAY به دو زیر مجموعه‌ی H و R تقسیم می‌شوند. مدل MLR توصیف‌گر $H8m$ و $H8p$ را از گروه توصیف‌گرهای H-GETAWAY برای بازدارنده‌های تیروزین کیناز انتخاب کرد که اولی بر اساس جرم اتمی و دومی بر اساس قطبش‌پذیری اتمی محاسبه می‌شود. این توصیف‌گرها به اندازه و محل اتم‌ها در مولکول وابسته بوده به گونه‌ای که با افزایش اندازه‌ی اتم و فاصله‌ی اتم از مرکز مولکول، مقدار آن افزایش می‌یابد. افزایش در $H8m$ سبب کاهش PIC_{50} ترکیب می‌شود ولی توصیف‌گر $H8p$ اثر مستقیم روی PIC_{50} ترکیب دارد. گروه دیگر توصیف‌گرهای R-GETAWAY هستند که بر اساس ماتریس تأثیر-فاصله به صورت زیر تعریف می‌شوند.

$$[R]_{ij} = \left[\frac{\sqrt{h_{ii} \cdot h_{jj}}}{r_{ij}} \right]_{ij} \quad i \neq j \quad \text{رابطه‌ی (۶-۳)}$$

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

^۱ - Molecular Influence Matrix

این توصیف‌گرها به ساختار سه‌بعدی مولکول‌ها حساس بوده و در برگیرنده‌ی اطلاعاتی نظیر شکل، سایز، تقارن مولکول و نحوه‌ی توزیع اتم‌ها در مولکول می‌باشند. از این گروه، توصیف‌گر R8e در مدل ظاهر شده است، که محاسبات آن بر اساس الکترونگاتیویته سندرسون می‌باشد و روی PIC₅₀ ترکیبات اثر مستقیم دارد.

د- توصیف‌گرهای گروه BCUT^۱ [۵۸]

این توصیف‌گرهای مولکولی مقادیر ویژه‌ای از یک ماتریس ارتباطی تغییر شکل یافته به نام ماتریس بردن هستند. ماتریس بردن یک گراف مولکولی تهی از هیدروژن است در آن مولفه‌های قطری اعداد اتمی، الکترونگاتیویته و ... و عناصر غیر قطری در ارتباط با مرتبه‌ی پیوند دو اتم پیوندشده، هستند. از این گروه BEHV3 که بر اساس حجم واندروالس اتمی محاسبه شده است بر اساس مدل رگرسیون خطی چندگانه انتخاب گردید. افزایش این توصیف‌گر با اثر متوسط منفی، سبب کاهش PIC₅₀ می‌شود.

ه- توصیف‌گرهای گروه RDF^۲ [۵۸]

می‌توان گفت RDF یا تابع توزیع شعاعی مربوط به یک دسته اتم‌ها، معادل توزیع احتمال یافتن یک اتم در یک فضای کروی به شعاع R است. رابطه‌ی (۷-۳) نحوه‌ی محاسبه‌ی کد RDF را نشان می‌دهد:

$$\sum_i^{N-1} \sum_{j>i}^N A_i \cdot A_j \cdot e^{-\beta \cdot (R-r_{ij})^2} g(R) = f. \quad \text{رابطه‌ی (۷-۳)}$$

که f یک فاکتور مقیاس و N تعداد اتم‌های مولکول است. همچنین، r_{ij} فاصله‌ی بین دو اتم i و j و A یک ویژگی اتمی (جرم اتمی) است. β یک فاکتور تسهیل کننده است که توزیع احتمال فاصله بین

^۱ - Burden-CAS-university of texas eigenvalues

^۲ - Radial Distribution Function

اتمی را مشخص می‌کند و می‌توان از آن به فاکتور دما برای تعریف جنبش اتمی تعبیر کرد. $g(R)$ در نقاط گسسته‌ای با فواصل معین حساب می‌شود با مجموعه‌ای از کدهای RDF با ویژگی‌های مختلف اتمی می‌توان ساختار سه بعدی یک مولکول را به طور واضح توصیف کرد. این توصیف‌گرها اهمیت توزیع اتم‌ها در مولکول روی فعالیت مهارکنندگی آن در مقابل مرگ سلولی نکروپتوسیس را نشان می‌دهد. مدل برای بازدارنده‌های تیروزین کیناز توصیف‌گرهای RDF105m و RDF095m را انتخاب کرد که بر اساس جرمهای اتمی محاسبه شده‌اند. با توجه به رابطه‌ی (۳-۶) هر چه جرم اتم بیشتر و فاصله‌ی بین دو اتم کمتر شود مقدار این توصیف‌گر بیشتر می‌شود. این توصیف‌گرها دارای اثر منفی روی PIC₅₀ ترکیبات دارند.

۳-۱-۲- نتیجه‌گیری نهایی

روند نتایج در این تحقیق، همانند پژوهش‌های مشابه نشان داد که روابط ساختار و فعالیت (QSAR) در مولکول‌های آلی با فعالیت زیستی، روابطی غیرخطی و نسبتاً پیچیده است که با مدل‌های خطی ساده نمی‌توان آن‌ها را به خوبی توصیف کرد. لذا در این پروژه از مدل شبکه‌ی عصبی مصنوعی با الگوریتم آموزشی تنظیم بایرین برای مطالعات QSAR استفاده شده است که بر اساس اطلاعات موجود این اولین گزارش در مورد بکارگیری روش فوق برای پیش‌بینی فعالیت مهارکنندگی ترکیبات شیمیایی است. نتایج حاکی از این است که شبکه قادر است یک رابطه‌ی رضایت بخش بین ساختار مولکولی و فعالیت مهارکنندگی ترکیبات برقرار کند. نتایج حاصل از این مدل می‌تواند کمک بسیار زیادی به محققان در طراحی و آزمایش ترکیبات جدید با ویژگی بازدارندگی قویتر نماید.

QSAR بازدارنده‌های تیروزین کیناز شامل مشتقات ۴- (۳- برموانیلینو)-۶ و ۷ دی متوکسی کینازولین توسط شی و همکارانش در سال ۲۰۰۷ با استفاده از مدل رگرسیون خطی چند گانه و بهینه سازی کلونی مورچه اصلاح شده (ACO) مورد بررسی قرار گرفته است. ارزیابی مدل ارائه شده توسط شی از روش cross-validation انجام شد. بهترین مدل با بالاترین تطابق، مدل با ۶ و ۷

توصیف‌گر می‌باشد، که مقدار R^2 برای سری تست و روش رد تک تک با ۶ توصیف‌گر بترتیب ۰/۷۲۲۳ و ۰/۶۳۴۷ و با ۷ توصیف‌گر بترتیب ۰/۶۴۲۲ و ۰/۶۲۴۷ می‌باشد.

با توجه به جدول (۸-۲) مشاهده می‌شود استفاده از مدل رگرسیون خطی چندگانه در تحقیق حاضر با ۷ توصیف‌گر نتایج بهتری نسبت به مدل شی و همکارانش دارد. همچنین با توجه به نتایج جداول (۲۱-۲) و (۳۸-۲) نیز مشاهده می‌شود که مدل شبکه عصبی مصنوعی با الگوریتم تنظیم بایزین نسبت به هر دو مدل رگرسیون خطی چندگانه (مطالعه‌ی حاضر و مطالعه‌ی شی و همکارانش) نتایج بهتری دارد که بیانگر قدرت پیش‌بینی بالای شبکه عصبی مصنوعی نسبت به مدل خطی می‌باشد. لذا همانطور که ملاحظه می‌شود و در پژوهش‌های مشابه نیز گزارش شده ([۳۶]، [۴۰] و [۴۹])، بهترین مدل‌های پیش‌گو با استفاده از روابط ساختار - فعالیت از نوع شبکه عصبی مصنوعی‌اند. در صورتی که این مدل‌ها با توصیف‌گرهای مناسبی تغذیه شوند، بهترین نتایج را در پی دارند.

۳-۲- بررسی نتایج مدل سازی بازدارنده های نکروپتوزیس، نکروستاتین - ۵

مدل های پیش گوی مورد استفاده در این تحقیق با شاخص MSE ارزیابی شد، این شاخص میانگین انحراف از جواب صحیح را بدست می دهد. مقدار این شاخص نشان می دهد که مدل های غیر خطی در پیش گویی متغیر وابسته بهتر از مدل های خطی عمل می کنند. این نتیجه که معمولاً در تحقیقات مشابه نیز دیده می شود نمایانگر این واقعیت است که روابط ساختار - فعالیت در ترکیبات فعال زیستی از پیچیدگی خاصی برخوردار است و با مدل های خطی ساده نمی توان آن ها را به راحتی بیان کرد.

۳-۲-۱- توصیف گرهای انتخاب شده [۵۸]

الف- توصیف گرهای گروه GETAWAY

همانطور که در بخش قبل (۳-۱-۱ ج)) بیان شد این توصیف گرها بر اساس ماتریس نفوذ مولکولی تعریف شده اند، برای مشتقات نکروستاتین-۵ توصیف گر $H7u$ توسط مدل انتخاب شد. u در این توصیف گر نشان دهنده ی آن است که آن ها با ویژگی خاصی از اتم ها محاسبه نشده اند. ضریب اثر این توصیف گر منفی می باشد.

ب- توصیف گرهای گروه WHIM^۱

این توصیف گرها به وسیله ی گروه تحقیقی QSAR تودسچینی^۲ و همکارانش توسعه یافتند. حاوی اطلاعات سه بعدی هستند و بر اساس تجزیه ی اجزا اصلی ماتریس کوواریانس وزن دار شده حاصل از کوردینه های کارترین مولکول ها می باشند. اهمیت این توصیف گرها به خاطر تأثیر کنفورماسیون ترکیب در طول برهمکنش با هدف، می باشد. این توصیف گرها اطلاعات سه بعدی در

^۱ - Weighted Holistic Invariant Molecular descriptors

^۲ - Todeschini

مورد اندازه، شکل، تقارن و توزیع اتمی در مولکول‌ها را در اختیار ما قرار می‌دهند. رابطه‌ی (۸-۳) روش کلی محاسبه این توصیف‌گرها را بیان می‌کند:

$$S_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_i)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad \text{رابطه‌ی (۸-۳)}$$

که S_{jk} کوواریانس وزن‌دار بین کوئوردینه‌های اتمی j ام و k ام، A تعداد اتم‌ها، w_i وزن i امین اتم، q_{ij} و q_{ik} به ترتیب j امین و k امین کوئوردینه‌های i امین اتم، $\bar{q}$ مقدار میانگین مربوطه را نشان می‌دهند. توصیف‌گرهای WHIM به دو دسته‌ی مهم تقسیم می‌شوند: توصیف‌گرهای جهت‌دار^۱ و توصیف‌گرهای کروی^۲، گروه توصیف‌گرهای جهت‌دار از نظر اندازه، شکل و تقارن مولکولی در جهات محورهای اصلی توصیف‌هایی از مولکول ارائه می‌دهند. ولی اطلاعات توصیف‌گرهای کروی فقط با دید کروی از مولکول محاسبه می‌شود.

مدل رگرسیون خطی چندگانه در تحقیق حاضر توصیف‌گر 3ke را از دسته توصیف‌گرهای WHIM کروی انتخاب کرد که بر اساس ویژگی الکترونگاتیویته‌ی سندرسون اتم‌ها وزن‌دار شده است. بر اساس نتایج، تجسم کروی مولکول‌ها از نظر الکترونگاتیوی برای خاصیت بازدارندگی ترکیبات در مقابل مرگ سلولی نکروپتوزیس مهم است. اثر متوسط این توصیف‌گر مثبت می‌باشد یعنی با افزایش آن PEC_{50} ترکیب افزایش می‌یابد.

ج- توصیف‌گرهای گروه 2D autocorrelation

. این توصیف‌گرها نقش خاصیت فیزیکولوژی (مانند جرم، حجم و اندروالس و ...) در فعالیت مهارکنندگی ترکیبات در مقابل نکروپتوزیس را نشان می‌دهد. توصیف‌گر MATS7e که جزء توصیف‌گرهای Moran Autocorrelation می‌باشد، در مدل ظاهر شده است که بر اساس

¹ - Directional WHIM descriptors

² - Global WHIM descriptors

³ - K global shap index

الکترون گاتیویته‌ی سندرسون وزن دار شده است. وجود ضریب مثبت در مدل برای این توصیف‌گر نشان می‌دهد که افزایش مقدار آن سبب کاهش PEC_{50} ترکیبات مورد بررسی می‌شود.

د- توصیف‌گرهای گروه Geometrical

این توصیف‌گرها از ساختار سه بعدی مولکول مشتق می‌شوند. لازمه‌ی محاسبه این توصیف‌گر بهینه‌سازی کامل ساختار ترکیبات می‌باشد. توصیف‌گرهای هندسی نسبت به توصیف‌گرهای توپولوژیکی اطلاعات بیشتر و قدرت تمایز بالاتری برای ساختارهای مشابه مولکول و صورت‌بندی‌های مختلف مولکول تأمین می‌کنند و ویژگی‌های مولکول را به طور کاملاً دقیق مشخص می‌کند. معمولاً برای نمایش هندسی مولکول، اتم‌ها به صورت کره‌هایی سخت با شعاع واندوالس در نظر گرفته می‌شوند که در محل پیوندها با یکدیگر همپوشانی دارند. سپس در فضای سه بعدی با استفاده از مختصات x ، y ، z مراکز این کره‌ها و شعاع واندروالس آن‌ها، این توصیف‌گرها محاسبه می‌شوند. توصیف‌گرهای هندسی شامل حجم واندروالس مولکول، مساحت سطح، مساحت سایه‌ی مولکول، فاکتور شکل و گشتاورهای اینرسی مولکول می‌باشند. مدل رگرسیون خطی چندگانه از این گروه توصیف‌گرها، توصیف‌گر $PJ13$ را انتخاب کرد. این توصیف‌گر برای تمایز مولکول‌هایی با ساختار مشابه اهمیت دارد.

ه- توصیف‌گرهای گروه RDF

توصیف‌گرهای RDF یا تابع توزیع شعاعی بر اساس تابع توزیع فاصله در نمایش سه بعدی مولکول می‌باشند اهمیت این توصیف‌گرها به دلیل اختلاف در توزیع اتم‌ها در مولکول‌ها و اثر این توزیع بر روی فعالیت مهارکنندگی ترکیبات می‌باشد. مدل، ۳ نوع توصیف‌گر از این گروه شامل $RDF140u$ ، $RDF020u$ و $RDF090u$ را برگزید.

این گروه توصیف‌گرها ساده‌ترین و معمول‌ترین توصیف‌گرهای استفاده شده می‌باشند که خصوصیات مولکول را بدون داشتن اطلاعاتی زیادی در مورد آن منعکس می‌کنند. تعداد اتم‌ها، تعداد پیوندها، تعداد مطلق و نسبی یک نوع اتم خاص در مولکول، تعداد پیوندهای ساده، دوگانه، سه‌گانه، آروماتیک، تعداد حلقه بر اساس تعداد اتم‌های آن‌ها و وزن مولکولی، متداول‌ترین توصیف‌گرهای ساختاری می‌باشند. یکی از این توصیف‌گرها (nBnz) در مدل ظاهر شده است و بیانگر تعداد حلقه‌های بنزن در مولکول می‌باشد.

۳-۲-۱- نتیجه‌گیری نهایی

با توجه به اینکه تاکنون مطالعات QSAR ترکیبات بازدارنده‌ی نکرپتوزیس، نکروستاتین-۵، گزارش نشده است و این اولین گزارش در مورد بکارگیری شبکه عصبی با الگوریتم تنظیم بایزین در مورد ترکیبات فوق می‌باشد، این مدل می‌تواند به محققان در طراحی ترکیبات جدید از این مشتقات با استخلاف‌های مختلف کمک زیادی نماید. روند نتایج در این تحقیق، همانند پژوهش‌های مشابه، نشان داد که روابط ساختار - فعالیت در مولکول‌های آلی با فعالیت زیستی، روابطی غیرخطی و نسبتاً پیچیده است که با مدل‌های خطی ساده نمی‌توان آن‌ها را به خوبی توصیف کرد. از این رو بهترین مدل‌های ساختار - فعالیت برای این ترکیبات مدل‌هایی هستند که ارتباط بین این توصیف‌گرها و عملکرد ترکیبات را با مدل‌های غیرخطی جست‌وجو کنند.

صحت و دقت نتایج مدل‌سازی QSAR به دقت داده‌های تجربی در مورد ساختار و فعالیت مولکول‌های مورد استفاده در مجموعه‌ی ترکیبات وابستگی زیادی دارد. هم‌چنین، بزرگی مجموعه‌ی این ترکیبات در میزان درستی نتایج مؤثر است. از این رو می‌باید تعادلی بین کمیت و کیفیت ترکیبات مورد استفاده برقرار کرد.

۳-۳- آینده‌نگری

۱- محققان می‌توانند از نتایج حاصل از این پروژه در طراحی و آزمایش ترکیبات جدید با ویژگی بازدارندگی قوی‌تر برای آنزیم تیروزین کیناز و همچنین نکروپتوسیس استفاده کنند. این تحقیقات را می‌توان با پژوهش روی خواص دارویی ترکیبات جدید ادامه داد تا نهایتاً منجر به ساخت داروهای جدیدتر با اثربخشی بهتر گردد.

۲- پژوهش‌گران می‌توانند با الگوگیری از روش مورد استفاده در این تحقیق، به بررسی روابط ساختار - فعالیت در گروه‌های دیگر مولکولی بپردازند. از این گروه‌ها می‌توان به بازدارنده‌ها یا فعال‌کننده‌های آنزیم‌های دیگر و پیام‌رسان‌های شیمیایی اشاره کرد.

۳- از آنجا که در این تحقیق، انتخاب توصیف‌گرهای مناسب برای ورود به شبکه‌ی عصبی با استفاده از روش رگرسیون خطی چندگانه انجام شده است و چون در روش رگرسیون خطی چندگانه تنها برهمکنش‌های خطی لحاظ می‌شود، لذا می‌توان از روش‌های دیگر برای انتخاب توصیف‌گر مانند الگوریتم ژنتیک استفاده کرد و نتایج حاصل را با نتایج این پروژه مقایسه نمود.

فهرست منابع و مندرجات

- [۱] - منصور عرب چم‌جنگلی (۱۳۸۶)، "پیش‌بینی فعالیت دارویی ضد ایدز (سیتوتوکسیته) مشتقات ۵- فنیل-۱- فنیل‌آمینو H - ایמידازول به وسیله شبکه عصبی مصنوعی"، دانشگاه صنعتی شاهرود، گزارش طرح پژوهشی
- [2] - M. Arab Chamjangali, M. Beglari, G. Bagherian, (2007), "Prediction of cytotoxicity data (CC₅₀) of anti-HIV 5-phenyl-1-phenylamino-1H-imidazole derivatives by artificial neural network trained with Levenberg-Marquardt algorithm", *Journal of Molecular Graphics and Modelling*, 26, pp 360-367
- [3] - M. Arab Chamjangali, (2009), "Modelling of Cytotoxicity Data (CC₅₀) of Anti-HIV 1-[5-Chlorophenyl] Sulfonyl-1H-Pyrrole Derivatives Using Calculated Molecular Descriptors and Levenberg-Marquardt Artificial Neural Network", *Chem Biol Drug Des* 2009
- [4] - 26. C. Hansch, T. Fujita, (1964), "p-σ-π Analysis: a method for the correlation of biological activity and chemical structure", *J. Am. Chem. Soc.*, 86, 1616
- [5] - T. Clark, (1985) "A Handbook of Computational Chemistry", John Wiley & Sons, New York
- [6] - J. A. Pople, D. Beveridge, (1970), "Approximate Molecular Orbital Theory", McGraw-Hill,
- [7] - Wayne W. Daniel, (1999), "Biostatistics A foundation for Analysis in the Health Science", John WILEY.
- [8] - Wasserman, Philip. D, (1989), "Neural Computing", Theory and Practice,
- [9] - W.W. McCulloch, W. Pitts, (1943), "A Logical Calculus of the Ideas Imminent in Nervous Activity", *Bull. Math. Biophys.*, 5, pp 115-133
- [10] - M. L. Minsky, S. Papert, (1969), "Perceptrons: An Introduction to Computational Gemerty", MIT Press, Cambridge, Mass.,.
- [۱۱] - اسماعیلی ا. (۱۳۸۵) پایان نامه فوق لیسانس "پیش‌بینی کوتاه مدت قیمت مصرف انرژی الکتریکی در بازار برق ایران"، دانشگاه صنعتی شاهرود،
- [12] - L. Stryer, (1999), "Biochemistry", W.H. Freeman and Company
- [13] - F. Rosenblatt, (1958), "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychol. Rev.*, 65, pp 386-408

[14] - F. Rosenblatt, (1962), "Principle of Neurodynamics", Spartan Books, Washington, D. C.

[۱۵] - راهنمای نرم افزار مطلب ۲۰۰۷

[۱۶] - ع.ا. حسنی پاک، م. شرفالدین، (۱۳۸۱)، "تحلیل داده های اکتشافی"، موسسه ی انتشارات دانشگاه تهران

[17] - D. Rummelhurt, G. Hinton, R. Williams, (1986), "Learning Representations by Back propagating Error", *Nature*, 323, pp 533-536

[18] - D. Parker, (1982), "Learning Logic, Invention Report", S81-64, File 1, Office of Technology Licensing, Stanford University, Stanford, Calif

[19] - P. Werbos, (1974) "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Science", Ph. D. dissertation, Harvard University (Wasserman, Philip D, (1998), "Neural Computing, Theory and Practice")

[20] - Michael Fernandez and Julio Caballero, (2006), "Modeling of activity of cyclic urea HIV-1 protease inhibitors using regularized-artificial neural networks" *Bioorganic & Medicinal Chemistry*, 14, pp 280-294

[21] - K. Hirschen, M. Schafer, (2006), "Bayesian regularization neural networks for optimizing fluid flow processes", *Comput. Methods Appl. Mech. Engrg*, 195, pp 481-500

[22] - Karelson, M., 2000, "Molecular Descriptors in QSAR/QSPR", Wiley- VCH, Newyork.

[23] - Hajmeer M., Basheer I., (2003), "Comparison of logistic regression and neural network-based classifiers for bacterial growth", *Food Microbiol*, 20, pp 43-45

[24] - Y.C. Martin, K.H. Kim, G.M. Bures, (1992) "Computer assisted drug design in 21st century", *Medicinal Chemistry for the 21st Century*, Wermuth, C.G. et al., Eds, Blackwell Scientific, London, chap. 20.

[25] - Kubinyi, H., (1993), "QSAR: Hansch Analysis and Related Approaches", VCH, Weinheim

[26] - Lakshmi M.S., Sherbet G.V., (1997) "Genetics of Cancer: Genes Associated With Cancer Invasion, Metastasis and Cell Proliferation". Academic Press, London

- [27] - McDonald F., Ford CH.G., (1997), "Molecular Biology of Cancer.Springer Verlag", New York, Incorporated
- [28] - Manash K. Paul and Anup K. Mukhopadhyay, (2004), "Tyrosine kinase – Role and significance in Cancer", *Int. J. Med. Sci*, 1(2), pp 101-115
- [29] – Wyllie AH., (1980), "Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation", *Nature*, 284, pp 555-566
- [۳۰] – هاشمی محمد، قوامی سعید، کرمی تهرانی فاطمه، (۱۳۸۲) "آپتوسیس، مرگ برنامه‌ریزی شده سلول"، ویژه‌نامه طبیب شرق، سال پنجم، شماره ۱، ص ۷۱-۷۷
- [31] – Hardy k., (1992)"Apoptosis in the human embryo", Report, 4, pp 125-134
- [32] - Michael Knudson, (2001), "Overview of apoptosis", Cell biology and cancer. http://www.medicine.uiowa.edu/pathology/path_folder/faculty/knudson/knodson.com pp 12-22
- [33] - Weihong Zheng, Alexei Degterev, Emily Hsu Junying Yuan, Chengye Yuan, (2008), "Structure–activity relationship study of a novel necroptosis inhibitor, necrostatin-7", *Bioorganic & Medicinal Chemistry Letters*, 18, pp 4932-4935
- [34].- J.M.J. Tronchet, M. Grigorov, N. Dolatshahi, F. Moriaud, J. Weber, (1997), "A QSAR study confirming the heterogeneity of the HEPT derivative series regarding their interaction with HIV reverse transcriptase", *Europ. J. Med. Chem.*, 32, pp279
- [35] - Y.D. Cai, H. Yu, K.C. Chou,(1998), "Artificial neural network method for predicting HIV protease cleavage sites in protein", *J. Protein Chem*, 17, pp 607-615
- [36] – Jalali-Heravi, M. and Parastar, F.,(2000), "Use of Artificial Neural Networks in a QSAR Study of Anti-HIV Activity for a Large Group of HEPT Derivatives", *J. Chem. Inf. Comput. Sci.*, 40, pp 147-154
- [37] - Frank R. Burden, (2000), "Use of Automatic Relevance Determination in QSAR Studies Using Bayesian Neural", *J. Chem. Inf. Comput. Sci*, 40, pp 1423-1430
- [38] - Frank R. Burden, (2001), "Quantitative Structure–Activity Relationship Studies Using Gaussian Processes", *J. Chem. Inf. Comput. Sci*, 41, pp 830-835
- [39] – G. Hatzakis, M., Boulaajaj, S., Sebti, S. and Zakarya, D., (2001), SAR QSAR *Environ. Res.*, 13, pp 567

- [40] - Ashish Garg, Krishna L. Bhat, Charles W. Bock, (2002), "Mutagenicity of aminoazobenzene dyes and related structures: a QSAR/QPAR investigation", *Dyes and Pigments*, 55, pp 35-52
- [41] - D. Gonza'lez-Arjona, G. Lo'pez-Pe'rez, A. Gustavo Gonza'lez, (2002), "Non-linear QSAR modeling by using multilayer perceptron feedforward neural networks trained by back-propagation", *Talanta*, 56, pp 79-90
- [42] - Les Sztandera, Ashish Garg, Seth Hayik, Krishna L. Bhat, Charles W. Bock, (2003), "Mutagenicity of aminoazo dyes and their reductive-cleavage metabolites: a QSAR/QPAR investigation", *Dyes and Pigments*, 59, pp 117-133
- [43] - Dana Weekes, Gary B. Fogel, (2003), "Evolutionary optimization, backpropagation, and data preparation issues in QSAR modeling of HIV inhibition by HEPT derivatives", *BioSystems*, 72, pp 149-158
- [44] - I. Valkova, M. Vra'cko, S.C. Basak, (2004), "Modeling of structure-mutagenicity relationships: counter propagation neural network approach using calculated structural descriptors", *Analytica Chimica Acta*, 509, pp 179-186
- [45] - Mukesh Doble, S. Karthikeyan, P. A. Padmawar and K. G. Akamanchi, (2005), "QSAR studies of paeonol analogues for inhibition of platelet aggregation", *Bioorganic & Medicinal Chemistry*, 13, pp 5996-6001
- [46] - Sayyed Hamed Sadat Hayatshahi, Parviz Abdolmaleki, Shahrokh Safarian, Khosro Khajeh, (2005), "Non-linear quantitative structure-activity relationship for adenine derivatives as competitive inhibitors of adenosine deaminase", *Biochemical and Biophysical Research Communications*, 338, pp 1138- 1142
- [47] - Fang Zheng, Ersin Bayram, Sangeetha P. Sumithran, Joshua T. Ayers, Chang-Guo Zhan, Jeffrey D. Schmitt, Linda P. Dwoskin and Peter A. Crooks, (2006), "QSAR modeling of mono- and bis-quaternary ammonium salts that act as antagonists at neuronal nicotinic acetylcholine receptors mediating dopamine release" *Bioorganic & Medicinal Chemistry*, 14, pp 3017-3037
- [48] - Alan R. Katritzky, Minati Kuanar, Dimitar A. Dobchev, Barbara W. A. Vanhoecke, Mati Karelson, Virinder S. Parmar, Christian V. Stevense and Marc E. Bracke, (2006), "QSAR modeling of anti-invasive activity of organic compounds using structural descriptors", *Bioorganic & Medicinal Chemistry*, 14, pp 6933-6939
- [49] - Qing Su. Lu Zhou, (2006), "QSAR modeling of AT1 receptor antagonists using ANN", *J Mol Model*, 12, pp 869-875

- [50] - Apilak Worachartcheewan, Chanin Nantasenamat, Thanakorn Naenna, Chartchalerm Isarankura-Na-Ayudhya, Virapong Prachayasittikul, (2008), "Modeling the activity of furin inhibitors using artificial neural network", *European Journal of Medicinal Chemistry*, 44, pp 1664-1673
- [51] - Adam Buciński, Agnieszka Socha, Małgorzata Wnuk, Tomasz Bączek, Alicja Nowaczyk, Jerzy Krysiński, Krzysztof Goryński, Marcin Koba, (2009), "Artificial neural networks in prediction of antifungal activity of a series of pyridine derivatives against *Candida albicans*", *Journal of Microbiological Methods*, 76, pp 25-29
- [52] - Wei-min Shi, Qi Shen, Wei Kong, Bao-xian Ye, (2007), "QSAR analysis of tyrosine kinase inhibitor using modified ant colony optimization and multiple linear regression", *European Journal of Medicinal Chemistry*, 42, pp 81-86
- [53] – (2002), HyperChem 7.0. Toronto, Canada: HyperCube Inc
- [54] - Todeschini R., Milano Chemometrics and QSAR Group.
<http://www.disat.unimib.it/vhml>
- [55] - SPSS for windows, Statistical Package for IBM PC, SPSS Inc.,
<http://WWW.SPSS.com>
- [56] - MATLAB 7.5, The Math Works, Inc., Natick, MA, USA
- [57] - Ke Wang, Jinfeng Li, Alexei Degterev, Emily Hsu, Junying Yuan and Chengye Yuan, (2007), "Structure–activity relationship analysis of a novel necroptosis inhibitor, Necrostatin-5", 17, pp 1455-1465
- [58] - Roberto Todeschini, Viviana Consonni,(2000), "Hanbook of molecular Descriptors ", WILEY-VCH
- [59] - Consonni V, Todeschini R., Pavan M, (2002), "Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors", *J. Chem. Inf. Comput. Sci*, 42, pp 682-692.

Abstract

Tyrosine kinases are important mediators of the signaling cascade, which play key roles in diverse biological processes. Researches show that inhibition of this enzyme has curative effects on some cancers. In the first section, quantitative structure-activity relationship (QSAR) models for 60 analogues of 4-(3-bromoanilino)-6,7-dimethoxyquinazoline on epidermal growth factor receptor tyrosine kinase were constructed using Bayesian Regularized Artificial Neural Networks (BR-ANN). For comparison purpose, the multiple linear regression (MLR) was also tested. Among a large number of calculated descriptors, only nine significant molecular descriptors were obtained by stepwise regression, as the most feasible descriptors, and then they were used as inputs for neural network. The data set was randomly divided into 45 training and 10 validation and 5 test sets and the neural network architecture and its parameters were optimized. The prediction ability of the model was evaluated using the validation, test data sets, and leave-one-out (LOO) cross-validation method. The mean square errors for the validation, test data sets and leave-one-out method were 0.1329, 0.1937 and 0.3367, respectively. The results obtained showed the excellent prediction ability and stability of the proposed model in the prediction of inhibitory activity data of the corresponding analogues.

In the second section, some derivatives of necroptosis inhibitors (Necrostatin-5) were studied (Necroptosis is the third cell death pathway apart from apoptosis and necrosis). In this study, A data set including 161 compounds were used to design predictive models. The models that were established using Bayesian Regularized Artificial Neural Networks, could divide the compounds into two active and inactive groups with the accuracy 100%. In the after stage, were established models with data set for 52 active compounds each described with a diverse set of descriptors. The most effective descriptors were selected using stepwise regression and the predictive models were established using artificial neural networks. This model could predict the $p(EC_{50})(M)$ of the compounds with the mean square errors for the validation, test data sets and leave-one-out method 0.0651, 0.0767 and 0.0598, respectively. The prediction results are in good agreement with the experimental values.

Keywords: QSAR, molecular descriptors, drug activity, inhibitors, artificial neural networks



Shahrood University of Technology

Faculty of chemistry

Application of QSAR modeling for the prediction of drug activity of some new synthesized drug chemicals

Elahe Shirpour

Supervisor:
Dr M. Arab Chamjangali

Advisor:
Dr Gh. Bagherian

Jun 2009