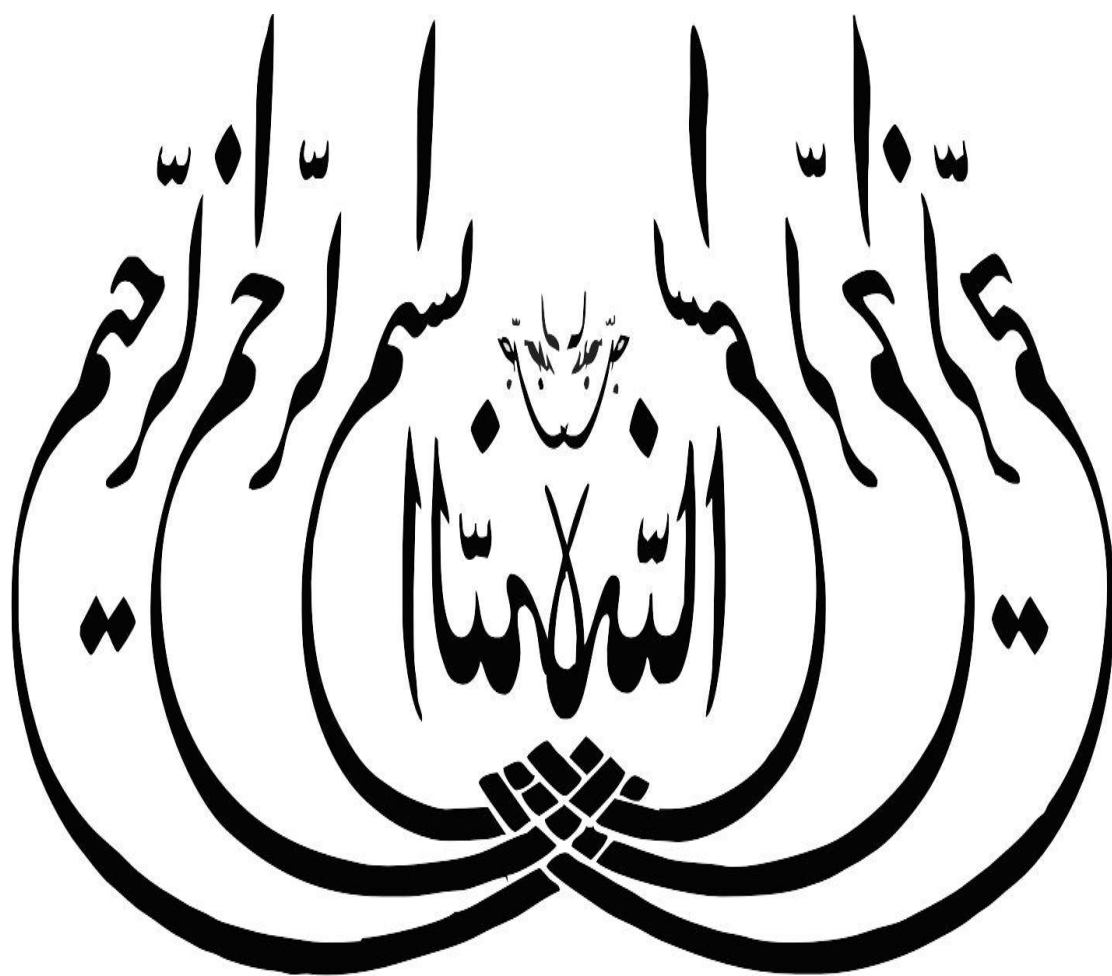






دانشکده شیمی

گروه: شیمی تجزیه

مطالعه کمی ساختار – فعالیت IC_{50} برخی از مشتقات گوانین به عنوان

بازدارنده CDK2 در بدن

دانشجو: پیام کلهر

استاد راهنما:

جناب آقای دکتر گودرزی

استاد مشاور:

جناب آقای دکتر نیکوفرد

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

اردیبهشت ماه ۱۳۹۰

این مجموعه را به پدر و مادر بزرگوالم تقدیم می‌کنم،

در ضمن از زحمات استاد فرهیخته، جناب آقای دکتر گودرزی و سایر اساتید دانشگاه صنعتی شاهرود

کمال تشکر و سپاسگزاری را دارم،

در نهایت از حمایت‌های معنوی سرکار خانم یاری‌وند نیز کمال تشکر را دارم.

چکیده

در بخش اول این تحقیق، ارتباط کمی ساختار - فعالیت (QSAR) ۵۶ ترکیب بازدارنده CDK2 با استفاده از دو روش خطی (رگرسیون خطی چندگانه؛ MLR) و غیر خطی (شبکه عصبی مصنوعی؛ ANN) مدل سازی شد. بازدارنده های CDK2 (کیناز وابسته به حلقه) ترکیباتی هستند که نقش اساسی در تنظیمات سلولی داشته و از تکثیر اضافی سلول ها توسط CDK2 که دچار فعالیت بیش از حد شده و منجر به سرطان می شوند جلوگیری می کنند. توانایی پیش بینی مدل ها با استفاده از سری ارزیابی، سری تست، روش رد مرحله ای تک تک و Y- تصادفی مورد ارزیابی قرار گرفت و خطای مجذور میانگین (MSE) برای سری تست به روش غیرخطی و خطی به ترتیب ۰/۰۶۳ و ۰/۰۶۹ می باشد. نتایج حاصله نشان دهنده ی قدرت پیش بینی بهتر مدل غیر خطی می باشد.

در بخش دوم، ارتباط کمی ساختار - ویژگی (QSPR) مربوط به شاخص بازداری ۱۹۶ ترکیب مونومتیل آلکان با استفاده از دو روش غیر خطی (شبکه عصبی مصنوعی؛ ANN) و خطی (رگرسیون خطی چندگانه؛ MLR) مدل سازی شد. توانایی پیش بینی مدل ها با استفاده از سری ارزیابی، سری تست، روش رد مرحله ای تک تک و Y- تصادفی مورد ارزیابی قرار گرفت و خطای مجذور میانگین (MSE) برای سری تست به روش غیرخطی و خطی به ترتیب ۴/۲۱ و ۲۸/۹۶ می باشد.

کلید واژه ها: ارتباط کمی ساختار - فعالیت، ارتباط کمی ساختار - ویژگی، رگرسیون خطی چندگانه، شبکه عصبی مصنوعی، فعالیت بیولوژیکی، شاخص بازداری

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

1. Linear and nonlinear QSAR study of N2 and O6 substituted Guanine Derivatives as Cyclin-Dependent Kinase 2 (CDK2) Inhibitors
2. Linear and nonlinear QSPR study of temperature-programmed gas chromatography linear retention indices of all C4 – C30 monomethylalkanes on methylsilicone OV-1 stationary phase

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

فهرست مطالب

فصل اول : مقدمه

۲	۱-۱- تعریف تحقیق
۳	۱-۱-۱- کاربرد QSAR در طراحی دارو
۴	۱-۲- شاخص بازداري
۴	۲-۱- ضرورت تحقیق
۵	۳-۱- پیشینه‌ی تحقیق

فصل دوم : کمومتری‌کس

۸	۱-۲- ارتباط کمی ساختار - ویژگی/ فعالیت
۱۰	۱-۱-۲- سری داده‌ها
۱۰	۲-۱-۲- بهینه‌سازی ساختار هندسی
۱۱	۳-۱-۲- محاسبه‌ی توصیف‌کننده‌ها
۱۱	۴-۱-۲- انتخاب بهترین توصیف‌کننده‌ها
۱۲	۵-۱-۲- مدل سازی
۱۳	۶-۱-۲- رگرسیون خطی چندگانه (MLR)
۱۶	۷-۱-۲- شبکه عصبی مصنوعی
۱۸	۱-۷-۱-۲- اساس بیولوژیکی شبکه‌های عصبی مصنوعی
۱۹	۲-۷-۱-۲- ساختار و عملکرد شبکه‌های عصبی مصنوعی (مدل ریاضی)
۲۳	۳-۷-۱-۲- توابع انتقال
۲۵	۴-۷-۱-۲- آموزش شبکه‌های عصبی مصنوعی
۲۵	۵-۷-۱-۲- الگوریتم آموزشی پس انتشار خطا
۲۹	۸-۱-۲- ارزیابی مدل

فصل سوم : مدل سازی QSAR مشتقات گوانین به عنوان بازدارنده های CDK2 با

استفاده از روش رگرسیون خطی و شبکه عصبی مصنوعی

۳۳	۱-۳- معرفی بازدارنده‌های CDK2
۳۴	۲-۳- نرم افزار های استفاده شده
۳۴	۱-۲-۳- نرم افزار HyperChem

۳۵	۳-۲-۲- نرم افزار Dragon
۳۵	۳-۲-۳- نرم افزار SPSS
۳۵	۳-۲-۴- نرم افزار MATLAB
۳۶	۳-۳- انتخاب سری داده‌های استفاده شده در مدل سازی
۴۳	۳-۴- محاسبه ی توصیف کننده‌ها
۴۵	۳-۵- انتخاب توصیف کننده‌های مناسب
۴۶	۳-۵-۱- انتخاب مدل مناسب
۵۱	۳-۵-۲- مدل سازی به روش شبکه عصبی
۵۲	۳-۵-۲-۱- انتخاب اولیه ی مقدار وزن ها و بایاس
۵۲	۳-۵-۲-۲- بهینه سازی پارامترهای شبکه
۵۳	۳-۵-۳- انتخاب تعداد متغیرهای ورودی شبکه، الگوریتم آموزشی، نوع تابع انتقال، تعداد گره لایه ی پنهان و تعداد دوره های آموزشی

۵۹	۳-۵-۲-۴- بهینه سازی پارامتر μ
۶۰	۳-۵-۳- معماری شبکه ی عصبی بهینه شده

فصل چهارم: بحث و نتیجه گیری

۶۲	۴-۱- ارزیابی مدل های خطی و غیر خطی
۶۲	۴-۱-۱- ارزیابی مدل های خطی و غیر خطی با استفاده از سری های ارزیابی و تست
۶۵	۴-۱-۲- ارزیابی مدل های خطی و غیر خطی با استفاده از روش رد مرحله ای تک تک
۷۰	۴-۱-۳- ارزیابی مدل های ارائه شده با استفاده از پارامترهای آماری
۷۱	۴-۱-۴- ارزیابی شبکه با استفاده از Y - تصادفی
۷۲	۴-۲- بررسی ارتباط توصیف کننده ها ساختاری وارد شده در مدل با فعالیت بیولوژیکی ترکیبات مورد بررسی
۷۷	۴-۳- بررسی میزان مشارکت توصیف کننده ها در مدل شبکه ی عصبی
۷۸	۴-۴- نتیجه گیری نهایی

فصل پنجم: مدل سازی QSPR شاخصهای بازدارندگی خطی کروماتوگرافی مربوط به مونومیل آلکان های C4 – C30 بر روی فاز ثابت متیل سلیکون 1 – OV با استفاده از روش رگرسیون خطی و شبکه عصبی مصنوعی

- ۸۰ ----- ۱-۵- معرفی مونومتیل آلکان ها
- ۸۱ ----- ۲-۵- نرم افزارهای استفاده شده
- ۸۱ ----- ۳-۵- انتخاب سری داده ها
- ۸۱ ----- ۴-۵- ویژگیهای سیستم اندازه گیری اندیس بازداري سری داده ها
- ۸۲ ----- ۵-۵- محاسبه ی توصیف کننده ها
- ۸۳ ----- ۶-۵- انتخاب توصیف کننده های مناسب
- ۹۳ ----- ۵-۶-۱- مدل سازی به روش شبکه عصبی
- ۹۴ ----- ۵-۶-۱-۱- انتخاب اولیه ی مقدار وزن ها و بایاس ها
- ۹۴ ----- ۵-۶-۱-۲- بهینه سازی پارامترهای شبکه
- ۹۴ ----- ۵-۶-۱-۳- انتخاب تعداد ورودی های شبکه، الگوریتم آموزشی، نوع تابع انتقال، تعداد گره های لایه ی پنهان و تعداد دوره های آموزشی
- ۱۰۰ ----- ۵-۶-۱-۴- بهینه سازی پارامتر μ
- ۱۰۱ ----- ۵-۶-۳- معماری شبکه ی عصبی بهینه شده

فصل ششم : بحث و نتیجه گیری

- ۱۰۳ ----- ۶-۱- ارزیابی مدل های خطی و غیرخطی
- ۱۰۳ ----- ۶-۱-۱- ارزیابی مدل های خطی و غیر خطی با استفاده از سری های ارزیابی و تست
- ۱۰۸ ----- ۶-۱-۲- ارزیابی مدل های خطی و غیر خطی با استفاده از روش رد مرحله ای تک تک
- ۱۱۷ ----- ۶-۱-۳- ارزیابی مدل های ارائه شده با استفاده از پارامترهای آماری
- ۱۱۸ ----- ۶-۱-۴- ارزیابی شبکه به روش Y - تصادفی
- ۱۱۹ ----- ۶-۲- بررسی ارتباط توصیف کننده های ساختاری وارد شده در مدل با اندیس بازداري ترکیبات مورد بررسی
- ۱۲۲ ----- ۶-۳- بررسی میزان مشارکت توصیف کننده ها در مدل شبکه عصبی
- ۱۲۳ ----- ۶-۴- نتیجه گیری نهایی
- ۱۲۳ ----- آینده نگری
- ۱۲۶ ----- فهرست منابع و مراجع

فهرست جداول

۲۴	جدول ۱-۲- نام و تعریف انواع مختلف توابع در شبکه عصبی
۳۷	جدول ۱-۳- ساختار، نام و فعالیت بیولوژیکی ترکیبات شرکت کننده در مدل سازی
۴۴	جدول ۲-۳- هجده سری از توصیف کننده‌های مورد استفاده در کمومتریکس
۴۶	جدول ۳-۳- کل مدل‌های منتخب از روش رگرسیون مرحله‌ای به همراه پارامترهای مربوطه
۴۸	جدول ۴-۳- توصیف کننده‌های مدل منتخب به همراه کلاس مربوطه
۴۸	جدول ۵-۳- ماتریس همبستگی توصیف کننده‌های انتخاب شده
۴۹	جدول ۶-۳- ضریب رگرسیون و اثر متوسط توصیف کننده‌ها
۵۰	جدول ۷-۳- مقادیر عددی توصیف کننده‌های وارد شده در مدل سازی
۵۴	جدول ۸-۳- پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا
	جدول ۹-۳- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی
۵۵	لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئید
	جدول ۱۰-۳- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی بایزین
۵۶	و تابع تبدیل لگاریتم سیگموئید
	جدول ۱۱-۳- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت
۵۷	و تابع تبدیل تانژانت سیگموئید
	جدول ۱۲-۳- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین
۵۸	و تابع تبدیل تانژانت سیگموئید
۵۹	جدول ۱۳-۳- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ
	جدول ۱-۴- مقادیر تجربی و پیش بینی شده ی pIC_{50} برای ترکیبات سری ارزیابی با استفاده از روش‌های
۶۳	خطی و غیرخطی
	جدول ۲-۴- مقادیر تجربی و پیش بینی شده ی pIC_{50} برای ترکیبات سری تست با استفاده از روش‌های
۶۴	خطی و غیرخطی
	جدول ۳-۴- مقادیر تجربی، پیش بینی شده و خطای حاصله به روش رد مرحله‌ای تک تک با
۶۶	مدل ANN و MLR
۷۰	جدول ۴-۴- مقایسه‌ی آماره‌های بدست آمده از دو مدل ANN و MLR
۷۱	جدول ۵-۴- مقادیر مجذور ضریب همبستگی سری تست بعد از چندین Y- تصادفی
	جدول ۱-۵- هفت مدل اول حاصل از رگرسیون مرحله‌ای به همراه پارامترهای آماری مربوطه
۸۴	جدول ۲-۵- توصیف کننده‌های انتخاب شده توسط رگرسیون خطی مرحله‌ای
	جدول ۳-۵- ماتریس همبستگی توصیف کننده‌های انتخاب شده توسط رگرسیون خطی چندگانه
۸۵	مرحله‌ای
۸۶	جدول ۴-۵- مقادیر R^2 و MSE مربوط به سری ارزیابی
۸۶	جدول ۵-۵- نام و معنای توصیف کننده به همراه اثر متوسط
۸۷	جدول ۶-۵- مقادیر عددی توصیف کننده‌ها و مقادیر تجربی اندیس بازداري برای ترکیبات مورد بررسی

- جدول ۵-۷- پارامترهای شبکه‌های بهینه بدست آمده بر اساس مقادیر میانگین مربعات خطا (MSE) ۹۵
- جدول ۵-۸- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی ۹۶
- جدول ۵-۹- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم سیگموئید ۹۷
- جدول ۵-۱۰- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونبرگ- مارکوارت و تابع تبدیل تانژانت سیگموئید ۹۸
- جدول ۵-۱۱- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت سیگموئید ۹۹
- جدول ۵-۱۲- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ ۱۰۰
- جدول ۶-۱- مقادیر تجربی و پیش بینی شده‌ی RI برای ترکیبات سری ارزیابی با استفاده از روش‌های خطی و غیرخطی ۱۰۴
- جدول ۶-۲- مقادیر تجربی و پیش بینی شده‌ی اندیس بازداري برای ترکیبات سری تست با استفاده از روش‌های خطی و غیرخطی ۱۰۶
- جدول ۶-۳- مقادیر تجربی، پیش بینی شده و خطای حاصله به روش رد مرحله‌ای تک تک با مدل ANN و MLR ۱۰۹
- جدول ۶-۴- مقایسه‌ی آماره‌های بدست آمده از دو مدل ANN و MLR ۱۱۷
- جدول ۶-۵- مقادیر مجذور ضریب همبستگی سری تست بعد از چندین Y - تصادفی ۱۱۸
- جدول ۶-۶- اثر توصیف‌کننده‌ی MW بر اندیس بازداري ۱۲۰

فهرست اشکال

- شکل ۲-۱- ساختار یک نرون بیولوژیکی ۱۸
- شکل ۲-۲- مدل ریاضی یک نرون ۱۹
- شکل ۲-۳- یک ساختار سه لایه ای از نرونها ۲۱
- شکل ۲-۴- توابع تبدیل (الف) خطی، (ب) سیگموئیدی ۲۴
- شکل ۳-۱- تغییرات R^2 بر حسب شماره‌ی مدل ۴۷
- شکل ۳-۲- تغییرات SE بر حسب شماره‌ی مدل ۴۷
- شکل ۳-۳- تغییرات F بر حسب شماره‌ی مدل ۴۷
- شکل ۳-۴- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت و تابع تبدیل لگاریتم سیگموئید ۵۵
- شکل ۳-۵- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل لگاریتم سیگموئید ۵۶
- شکل ۳-۶- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت و تابع تبدیل تانژانت سیگموئید ۵۷
- شکل ۳-۷- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل تانژانت سیگموئید ۵۸
- شکل ۳-۸- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی ۵۸

- ۵۹ ----- بر حسب پارامتر μ -----
- ۶۰ ----- شکل ۳-۹- طرح شماتیک ساختار شبکه عصبی بدست آمده پس از بهینه سازی -----
- ۶۳ ----- شکل ۴-۱- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل ANN -----
- ۶۳ ----- شکل ۴-۲- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل MLR -----
- ۶۴ ----- شکل ۴-۳- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل ANN -----
- ۶۴ ----- شکل ۴-۴- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل MLR -----
- شکل ۴-۵- نمودار مقادیر پیش‌بینی شده pIC_{50} با استفاده از شبکه‌ی عصبی مصنوعی بهینه شده
- ۶۸ ----- بر حسب مقادیر واقعی به روش رد مرحله ای تک تک -----
- شکل ۴-۶- نمودار مقادیر پیش‌بینی شده pIC_{50} با استفاده از روش خطی بر حسب مقادیر واقعی
- ۶۸ ----- به روش رد مرحله‌ای تک تک -----
- شکل ۴-۷- نمودار مقادیر خطای مطلق در ارزیابی مدل شبکه‌ی عصبی مصنوعی به روش
- ۶۹ ----- رد مرحله ای تک‌تک در مقابل مقادیر تجربی -----
- شکل ۴-۸- نمودار مقادیر خطای مطلق در ارزیابی مدل خطی به روش رد مرحله ای تک‌تک
- ۶۹ ----- در مقابل مقادیر تجربی -----
- شکل ۴-۹- نمودار درصد مشارکت توصیف‌کننده‌های بکار گرفته شده در مدل شبکه عصبی
- ۷۸ ----- برای ترکیبات CDK2-----
- ۸۶ ----- شکل ۵-۱- تغییرات R^2 بر حسب تعداد توصیف‌کننده‌ها برای سری ارزیابی -----
- شکل ۵-۲- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت و تابع تبدیل لگاریتم
- ۹۶ ----- سیگموئید -----
- ۹۷ ----- شکل ۵-۳- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل لگاریتم سیگموئید --
- شکل ۵-۴- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت و
- ۹۸ ----- تابع تبدیل تانژانت سیگموئید -----
- ۹۹ ----- شکل ۵-۵- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل تانژانت سیگموئید--
- شکل ۵-۶- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی
- ۱۰۰ ----- بر حسب پارامتر μ -----
- ۱۰۱ ----- شکل ۵-۷- تصویر شماتیک ساختار شبکه عصبی بدست آمده پس از بهینه سازی -----
- ۱۰۵ ----- شکل ۶-۱- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل ANN -----
- ۱۰۵ ----- شکل ۶-۲- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل MLR -----
- ۱۰۷ ----- شکل ۶-۳- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل ANN -----
- ۱۰۷ ----- شکل ۶-۴- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل MLR -----
- شکل ۶-۵- نمودار مقادیر پیش‌بینی شده اندیس بازداري با استفاده از شبکه‌ی عصبی مصنوعی
- ۱۱۵ ----- بهینه شده بر حسب مقادیر واقعی به روش رد مرحله ای تک تک -----
- شکل ۶-۶- نمودار مقادیر پیش‌بینی شده اندیس بازداري به روش خطی بر حسب مقادیر واقعی
- ۱۱۵ ----- به روش رد مرحله ای تک تک -----
- شکل ۶-۷- نمودار مقادیر خطای مطلق در ارزیابی مدل شبکه‌ی عصبی مصنوعی به روش رد مرحله ای
- ۱۱۶ ----- تک‌تک در مقابل مقادیر تجربی -----

- شکل ۶-۸- نمودار مقادیر خطای مطلق در ارزیابی مدل به روش رد مرحله ای تک-تک
در مقابل مقادیر تجربی ----- ۱۱۶
- شکل ۶-۹- نمودار درصد مشارکت توصیف‌کننده‌های بکار گرفته شده در مدل شبکه عصبی برای
مونومتیل آلکان‌ها ----- ۱۲۲

فصل اول

مقدمه

۱-۱- تعریف تحقیق

همان‌طور که در فصل دوم بیشتر توضیح داده خواهد شد، کمومتریکس شاخه‌ای از شیمی است که ابزارهای ریاضی و آمار را برای بهبود بخشیدن به داده‌های شیمیایی و بدست آوردن اطلاعات بیشتر از آنها به کار می‌گیرد. نگرش چند متغیره به مسئله، ساخت و ارزیابی مدل‌هایی با قابلیت پیش‌بینی، مقایسه‌ی نتایج بدست آمده از روش‌های مختلف، قابلیت تعریف و استفاده از شاخص-هایی که قادرند کیفیت اطلاعات استخراج شده و مدل‌های بدست آمده را بسنجند از کاربردهای کمومتریکس می‌باشد.

از شاخه‌های مهم کمومتریکس، ارتباط کمی ساختار - فعالیت / ساختار - ویژگی^۱ بوده که در این پایان‌نامه هر دو بررسی شده است. در قسمت نخست پروژه، تکنیک ارتباط کمی ساختار - فعالیت برای پیش‌بینی فعالیت بیولوژیک تعدادی ترکیب بازدارنده و در قسمت دوم پیش‌بینی شاخص‌های بازدارندگی از مونو متیل آلکان‌ها بررسی شده است. جهت بررسی ارتباط کمی ساختار - فعالیت / ویژگی، روش‌های مورد استفاده به دو دسته‌ی روش‌های خطی و روش‌های غیرخطی تقسیم می‌شوند. از جمله روش‌های خطی که بدین منظور مورد استفاده قرار می‌گیرد می‌توان به روش‌های رگرسیون خطی چندگانه^۲، رگرسیون اجزای اصلی^۳ و روش حداقل مربعات جزئی^۴ اشاره کرد. از مهمترین روش-های غیرخطی نیز می‌توان به شبکه‌های عصبی مصنوعی^۵ اشاره کرد. در تحقیق حاضر مطالعه‌ی ارتباط کمی ساختار - فعالیت / ویژگی با استفاده از روش‌های خطی و غیرخطی انجام گرفته و نتایج باهم مقایسه شده‌اند. در فصل‌های سوم و چهارم، رگرسیون خطی چندگانه به منظور ایجاد مدل مناسبی برای ایجاد ارتباط خطی بین فعالیت بیولوژیکی و ساختار ۵۶ ترکیب از نوع بازدارنده‌های CDK2 و سپس، شبکه عصبی مصنوعی به منظور بررسی روابط غیرخطی بین ساختار این ترکیبات و

^۱Quantitative Structure – Activity / Property Relationship (QSAR/QSPR)

^۲Multiple Linear Regression (MLR)

^۳Principal Component Regression

^۴Partial Least Square(PLS)

^۵Artificial Neural Network (ANN)

فعالیت بیولوژیکی به کار گرفته شده است. در فصل‌های پنجم و ششم نیز به کمک این روش‌ها ارتباط خطی و غیر خطی بین ساختار ۱۹۶ ترکیب از مونومتیل آلکان‌ها و شاخص بازدارندگی آنها بررسی شده است.

۱-۱-۱- کاربرد QSAR در طراحی دارو

QSAR به راحتی می‌تواند برای کمک به بهینه‌سازی سنتز داروها 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 می‌باشد. پس از شناسایی اصول مولکولی حیات، 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 باعث بروز یکی از آثار دارویی مربوط به آن می‌شود. در علم داروسازی مدرن با تکیه بر روش‌های QSAR، تولید دارو به یک روند تحقیقاتی کاملاً جهت‌دار تبدیل شده است. برای انجام یک مطالعه‌ی QSAR ابتدا باید مقادیر فعالیت دارویی ترکیبات از طریق تجربی بدست آیند و همچنین توصیف‌کننده‌های مناسب از طریق تجربی یا محاسبه توسط کامپیوتر تولید شوند، مرحله‌ی بعد دستیابی به ارتباط ریاضی بین ویژگی‌های ساختاری و مقادیر فعالیت است که معمولاً -ترین تکنیک چندمتغیری نیز رگرسیون خطی چندگانه (MLR) می‌باشد. با این حال این روش تنها زمانی به کار برده می‌شوند که تعداد مولکول‌های مورد بررسی بسیار بیشتر از تعداد توصیف‌کننده‌ها بوده و همچنین متغیرها با هم همبستگی زیادی نداشته باشند. استفاده از شبکه‌های عصبی مصنوعی به جای MLR، خصوصاً هنگامی که ارتباط بین توصیف‌کننده‌ها و فعالیت مورد نظر خطی نبوده و یا اینکه بین آنها برهمکنش‌هایی وجود داشته باشد، باعث بهبود مدل حاصله خواهد شد.

۱-۱-۲- شاخص بازداري

شاخص بازداري پارامتری کمی است که میزان بازداري نسبی هر جزء در نمونه را نسبت به آلکان‌های نرمال بر روی یک فاز ساکن معین و در یک دمای مشخص نشان می‌دهد. شاخص بازداري، متغیرهای مختلف سیستم‌های کروماتوگرافی را نرمالیزه کرده، به گونه‌ای که امکان مقایسه‌ی نتایج سیستم‌های مختلف فراهم می‌شود. مقدار شاخص بازداري از معادله‌ی (۱-۱) محاسبه می‌شود:

$$I = 100Y(Z - Y) \left[\frac{\log t'_r X - \log t'_r Y}{\log t'_r Z - \log t'_r Y} \right] \quad \text{معادله‌ی ۱-۱}$$

در معادله‌ی بالا $t'_r Y$ زمان بازداري اصلاح شده برای آلکان نرمال که قبل از ترکیب X از ستون خارج می‌شود، $t'_r X$ زمان بازداري اصلاح شده برای ترکیب مورد نظر، $t'_r Z$ زمان بازداري اصلاح شده برای آلکان نرمال که بعد از ترکیب X از ستون خارج می‌شود، Y تعداد کربن‌های آلکان نرمال که قبل از ترکیب مورد نظر از ستون خارج می‌شود. Z نیز تعداد کربن‌های آلکان نرمال که بعد از ترکیب مورد نظر از ستون خارج می‌شود، می‌باشد.

۱-۲- ضرورت تحقیق

پیشرفت علوم در زمینه‌های مختلف روش‌های آزمایشگاهی کمک شایانی به بهبود نتایج حاصل از این روش‌ها کرده است. با وجود این، مشکلات و محدودیت‌هایی از جمله وجود عدم قطعیت در نتایج آزمایشگاهی، هزینه‌ی بالا، در دسترس نبودن امکانات و عدم توانایی در اندازه‌گیری همه‌ی ترکیبات باعث کشف و توسعه‌ی روش‌های نظری شده است.

در این تحقیق سعی شده است با به‌کارگیری روش‌های QSAR/QSPR مدلی برای پیش‌بینی فعالیت بیولوژیکی در بخش اول تحقیق و پیش‌بینی شاخص بازداري در بخش دوم تحقیق ارائه شود.

۱-۳- پیشینه‌ی تحقیق

در سال ۲۰۰۵ فرناندز و کامبا^۱ بازدارنده‌های CDK2 را با استفاده از شبکه‌ی عصبی مصنوعی مدل‌سازی نمودند که در آن دقت مدل برای سری آموزش ۸۷٪ بود [۱].

سوما سامانتا^۲ در سال ۲۰۰۶، در مطالعه‌ی QSAR ترکیبات ۳-آمینوپیرازول‌ها به عنوان ترکیبات ضد تومور برای فعالیت بازدارندگی CDK2 بر مبنای پارامترهای شاخص^۳ و بارهای اتمی برای بررسی تاثیر کمی تفاوت‌های ساختاری روی فعالیت ضد تومور به این نتیجه رسیدند که اتم شماره‌ی ۵ برای فعالیت این گروه از ترکیبات نقش مهمی دارد. همچنین مطالعات نشان داد که با افزایش خاصیت الکترون‌گاتیویته‌ی اکسیژن در موقعیت ۸، فعالیت افزایش می‌یابد [۲].

در سال ۲۰۰۷ لی جی و لیو^۴ ارتباط ساختار- فعالیت بازدارنده‌های اکسیندول CDK را با استفاده از روش حداقل مربعات برداری (LSV)^۵ مطالعه کردند [۳].

کاتریتزکی^۶ و همکارانش در سال ۲۰۰۰، مدل‌های QSPR شامل ۱۷۸ ترکیب مونومتیل، دی-متیل، تری‌متیل و تترا متیل آلکان‌ها را توسط توصیف‌کننده‌های کوانتومی و توپولوژیکی برای پیش-بینی شاخص بازداري آنها ارائه کردند که انحراف استاندارد میانگین مدل ۴/۶ بود [۴].

دوو و لیانگ^۷ در سال ۲۰۰۳ با بررسی ارتباط کمی بین ساختار مولکولی و شاخص‌های بازداري ایزومتریکی ۱۳۴ آلکان متفاوت، خطای استاندارد $\pm 3i.u$ را بدست آوردند [۵].

در سال ۲۰۰۵ ژیانگ، لیانگ و هوکیو^۹ با مدل‌سازی QSPR مربوط به ۲۰۷ ترکیب از متیل آلکان‌ها برای پیش‌بینی اندیس بازداري با بکار گیری روش‌های ژنتیک الگوریتم^۱ (GA) و حداقل

^۱Fernandez and Camba^۲Soma Samanta^۳Indicatore parameteres^۴Li. J. and Liu^۵Least Square Vortex^۶Katritzky^۷DU and Liang^۸Index Unit^۹Ziang, Liang and Hu Q

مربعات جزئی^۲(PLS)، مقدار مجذور ضریب همبستگی ۰/۹۹۹۹ و انحراف استاندارد^۳ ۲/۸ را بدست آوردند [۶].

زارعی و همکارانش در سال ۲۰۰۸، در یک مطالعه‌ی QSPR بر مبنای شبکه‌ی عصبی ریز موج^۴ (WNN) برای شاخص بازداري تعداد ۱۷۸ ترکیب از مونومیل آلکانهای تولید شده توسط حشرات، مقدار خطای نسبی میانگین^۵ ۲/۲٪ را بدست آوردند [۷].

^۱Genetic Algorithm

^۲Partial Least Squares

^۳Standard Deviation

^۴Wavelet Neural Network

^۵Average Relative Error

فصل دوم

کمومتریکس

۲-۱- ارتباط کمی ساختار- ویژگی / فعالیت^۱

توانایی حل مسائل مختلف در زمینه درک مکانیسم فرایندهای مختلف شیمیایی، کشف و توسعه-ی مواد جدید و زمینه‌های دیگر شیمی هنوز به طور کامل وجود ندارد. روشهای محاسباتی کمومتریکس^۲ توانسته‌اند در جهت حل این مشکلات به میزان زیادی مفید واقع شوند. روشهای تئوری تکنیکهای کمومتریکس از چند دهه پیش شناخته شده‌اند، اما اخیراً با ظهور کامپیوترهای قدرتمند، کمومتریکس ماهیت واقعی خود را پیدا کرده است. بر طبق تعریف جامعی که انجمن بین المللی کمومتریکس از این روش دارد، کمومتریکس علم برقراری ارتباط بین سنجشهای انجام شده بر روی یک سیستم یا فرایند شیمیایی با استفاده از روشهای آماری و ریاضی می‌باشد. شاخه‌های مختلف شیمی اعم از شیمی فیزیک، شیمی آلی، شیمی معدنی، بیوشیمی و به ویژه شیمی تجزیه به نحوی از این تکنیکها بهره می‌جویند. یکی از شاخه‌های بسیار مهم کمومتریکس که مورد بحث این پایان نامه است، ارتباط کمی ساختار- ویژگی / فعالیت (QSPR/QSAR) است. هرگاه مطالعات به صورت ارتباط بین ساختار مولکولی و خواص مشاهده شده‌ی مولکول انجام گیرد، به آن ارتباط کمی ساختار- ویژگی (QSPR) گفته که می‌توان به یافتن رابطه‌ای بین خاصیتی نظیر نقطه‌ی جوش و ساختار مولکولی اشاره نمود. اما وقتی خواصی از نوع بیولوژیکی (مانند فعالیت دارو) مد نظر باشد، اشاره به ارتباط کمی ساختار - فعالیت (QSAR) خواهد داشت [۸]. QSAR سعی در پیدا کردن رابطه‌ای 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 شفاف سازی نحوه-ی ارتباط بین خواص مولکولها و ویژگیهای ساختمانی آنها، به پژوهشگران در پیش‌بینی رفتار مولکولهای جدید براساس رفتار مولکولهای مشابه کمک می‌کند. مدل‌های QSAR را می‌توان براساس پارامترهایی که بصورت تجربی بدست آمده و یا از طریق تئوری به دست آمده اند، به زبان ریاضی بیان نمود.

^۱Quantitative Structure Activity/Property relationship^۲Computational Chemometrics

از جمله روشهایی که به منظور مطالعه ارتباط خطی ساختار-ویژگی/ فعالیت مورد استفاده قرار می‌گیرد می‌توان به روش رگرسیون خطی چندگانه^۱، رگرسیون اجزای اصلی^۲ و روش حداقل مربعات جزیی^۳ اشاره نمود. از جمله روش‌های دیگر، شبکه‌های عصبی مصنوعی هستند که رابطه‌ی غیرخطی میان ساختار و ویژگی/ فعالیت ترکیبات را علاوه بر روابط خطی آنها مورد مطالعه قرار داده و در این زمینه کاربردهای چشم‌گیری داشته‌اند. در تحقیق حاضر، مطالعه ارتباط کمی ساختار-ویژگی و ارتباط کمی ساختار- فعالیت تعدادی از ترکیبات با استفاده از تکنیکهای رگرسیون خطی چندگانه و شبکه‌های عصبی مصنوعی انجام گرفته و نتایج این دو روش با یکدیگر مقایسه شده‌اند که در ادامه به توضیح هر کدام می‌پردازیم.

کورنهایش^۴ برای اولین بار در سال ۱۹۶۴ مفهوم رابطه‌ی کمی ساختار- فعالیت یا همان QSAR را مطرح کرد. او نشان داد که توصیف‌کننده‌های کمی مولکولی در ترکیبات دارویی با فعالیت بیولوژیکی این ترکیبات ارتباط تنگاتنگی دارند. او و دانشمندان پس از وی، از توصیف‌کننده‌های ساده‌ای برای توصیف ساختار ترکیبات استفاده کردند که توانایی توصیف ویژگیهای الکترونی، هندسی، فضایی و هیدروفوبیک ترکیبات دارویی را داشتند. هانش^۵ و فوجیتا^۶ از روشهای رگرسیون خطی چندگانه در مطالعات QSAR استفاده کردند [۹].

از سال ۱۹۷۰ به بعد تعداد مقالات منتشر شده در QSAR سریعاً افزایش یافت. برای انجام روش QSAR می‌توان شش مرحله‌ی مشخص را از هم تفکیک کرد. این شش مرحله عبارتند از:

۱- جمع‌آوری فهرستی از ترکیبات با مقادیر فعالیت آنها (سری داده‌ها)

۲- تعیین ساختار هندسی

^۱Multiple Linear Regression(MLR)

^۲Principle Component Analysis(PCA)

^۳Partial Least Squares(PLS)

^۴ koronharsh

^۵Hansh

^۶Fujitah

۳- استخراج توصیف کننده‌ها

۴- انتخاب بهترین توصیف کننده‌ها برای ورود به مدل

۵- مدل سازی

۶- ارزیابی مدل

۲-۱-۱- سری داده‌ها

اولین مرحله در توسعه‌ی مطالعات QSAR و QSPR جمع آوری فهرستی از ترکیبات است که کمیت‌های تجربی خاصی برای آنها توسط محققین گزارش شده‌اند. عموماً جمع‌آوری مولکول‌های سری داده‌ها از طریق جستجو در مقالات و مجلات علمی صورت می‌گیرد. از میان مجموعه مولکول‌های جمع آوری شده تعدادی که داده‌های گزارش شده در مورد آنها در شرایط عملی یکسانی به دست آمده باشند، به عنوان سری داده‌ها انتخاب می‌شوند.

برای اینکه QSAR یک روش مطلوب محسوب شود، بهتر است تنوع گونه‌ها زیاد باشد و همچنین هرچه سری داده‌ها بزرگتر باشد، مدل حاصله از قدرت پیش‌بینی بالاتری برخوردار خواهد بود.

۲-۱-۲- بهینه سازی ساختار هندسی

فراهم کردن ساختار هندسی مولکول‌ها و بهینه کردن آنها مرحله‌ی بعدی در مطالعات QSAR می‌باشد. بدین منظور از روش‌های 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 زیادی از مولکول دسترسی پیدا کنند. در این روش‌ها به راحتی می‌توان محاسبات را انجام داد، بدون اینکه از اصول و طریقه‌ی محاسبه آگاهی و دانش کافی را داشت.

^۱Computational Chemistry

شیمی محاسباتی ساختارها و واکنشهای شیمیایی را شبیه سازی می کند و این امر به شیمی دان اجازه می دهد که به مطالعه ی روشهای شیمیایی با اجرای محاسبات توسط کامپیوتر بپردازد بدون اینکه از روشهای تجربی و آزمایشگاهی استفاده کنند [۱۰].

۲-۱-۳- محاسبه ی توصیف کننده ها

پس از بهینه سازی ساختار مولکولها بایستی توصیف کننده های وابسته به ساختار مولکول استخراج شوند. توصیف کننده ها مقادیر عددی هستند که ویژگیهای مختلف ساختاری و الکترونی مولکولها را به طور کمی نشان می دهند. توصیف کننده هایی که در مطالعات QSAR و QSPR مورد استفاده قرار می گیرند شامل دو دسته توصیف کننده های تئوری و تجربی می باشند. توصیف کننده های تجربی از طریق آزمایش بدست آمده و برای گروههای محدودی از ترکیبات قابل دسترس هستند. توصیف کننده های تئوری بدون نیاز به داده های تجربی به کمک نرم افزارهای مناسب برای طیف وسیعی از ترکیبات قابل محاسبه هستند. بنابراین توصیف کننده های تئوری در این مطالعات بسیار با اهمیت می باشند.

۲-۱-۴- انتخاب بهترین توصیف کننده ها

یکی از مهمترین مراحل QSAR و QSPR انتخاب توصیف کننده های مناسب برای ایجاد مدل می باشد، زیرا توصیف کننده های زائد عملیات برازش و مدل سازی را طولانی نموده و تاثیری در بهبود نتایج نخواهند داشت. بدین منظور لازم است در انتخاب توصیف کننده ها دقت لازم به عمل آید و توصیف کننده هایی انتخاب شوند که با پارامتر مورد نظر ارتباط نزدیکی داشته و بتوانند تغییرات آنها به خوبی توجیه نمایند. یکی از روشهای انتخاب توصیف کننده های مناسب روش برازش مرحله ای^۱ است که با محاسبه ی ضرایب همبستگی^۲ مشخص می شوند. ضریب همبستگی میزان نزدیکی اطلاعات بدست آمده به خط رگرسیون را نشان داده که مقدار آن بین ۰ و ۱ تغییر می کند. صفر یعنی هیچ

^۱Stepwise

^۲Correlation coefficient

ارتباطی بین فعالیت یا ویژگی و توصیف‌کننده وجود ندارد و یک یعنی ارتباط کامل است. پس توصیف‌کننده‌هایی با ضرایب همبستگی بزرگ در برازش منحنی برای ساختن مدل استفاده می‌شوند.

۲-۱-۵- مدل سازی

پس از ارزیابی توصیف‌کننده‌ها و انتخاب مناسب‌ترین آنها، باید مدلی ساخته شود که قادر باشد با استفاده از این توصیف‌کننده‌ها ویژگی یا فعالیت مورد نظر را پیش‌بینی نماید. روشهای متفاوتی در مدل سازی به کار گرفته می‌شود که می‌توان آنها را در دو دسته‌ی کلی زیر قرار داد:

الف) استفاده از روشهای آماری چند متغیره: در این نوع مدل سازی، از روشهای کالیراسیون چند متغیره نظیر رگرسیون خطی چندگانه و روش حداقل مربعات جزئی استفاده می‌شود. این روشها خطی بوده و زمانی کارایی مناسبی دارند که داده‌ها دارای رفتار خطی باشند. یک رگرسیون با ایجاد معادله‌ی خطی مطابق معادله‌ی (۱-۲) انجام می‌شود:

$$\text{property / activity} = \text{const.} + (c_1 \cdot p_1) + \dots + (c_n \cdot p_n) \quad \text{معادله‌ی ۱-۲}$$

که پارامترهای p_1 تا p_n توصیف‌کننده‌هایی هستند که به روش محاسباتی برای هر مولکول به دست می‌آیند و ضرایب c_1 تا c_n از طریق برازش توصیف‌کننده‌ها با فعالیت یا ویژگی در مجموعه‌ای از مولکول‌ها که دارای خواص مشترک هستند محاسبه می‌شود.

ب) استفاده از روشهای 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 روشهای غیرخطی روش شبکه‌ی عصبی مصنوعی^۱ بطور گسترده در مطالعات QSAR و QSPR به کار گرفته می‌شود. در دو بخش بعد به بررسی مختصر مدلهای استفاده شده برای بررسی رابطه‌ی ساختار مولکولی و فعالیت بازدارندگی ترکیبات موجود می‌پردازیم.

^۱Artificial Neural Network (ANN)

۲-۱-۶- رگرسیون خطی چندگانه MLR

رگرسیون خطی چندگانه روشی است که برای مدل سازی ارتباط خطی بین متغیر وابسته و یک یا چند متغیر مستقل به کار می رود. متغیرهای مستقل (متغیرهای پیش بینی کننده) آن متغیرهایی هستند که فرض می شود سهمی در تابعی که برای مدل سازی مناسب است، دارند. متغیرهای وابسته (متغیرهای پاسخ) پارامترهایی هستند (اغلب از اندازه گیریهای تجربی به دست می آیند) که یافتن وابستگی آماری میان آنها و تعدادی از متغیرهای مستقل مورد نظر است.

معمولاً روش MLR برای پیش بینی زمان بازداری ترکیبات مختلف که با تکنیکهای مختلف کروماتوگرافی جداسازی شده اند یا پیش بینی فعالیت بیولوژیکی یکسری ترکیبات دارویی مورد استفاده قرار می گیرد. در این صورت فاکتور بازداری یا مقادیر فعالیت بیولوژیکی متغیر وابسته^۱ و توصیف کننده های ساختاری ترکیبات، متغیرهای مستقل^۲ هستند.

برای تعداد i داده تجربی و تعداد p متغیر مستقل، مدل رگرسیون خطی چندگانه توسط معادله ی (۲-۲) نمایش داده می شود:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad \text{معادله ی ۲-۲}$$

که در این معادله $\beta_0, \beta_1, \dots, \beta_p$ ضرایب رگرسیون و $x_1, x_2, \dots, x_p$ متغیرهای مستقل (توصیف کننده های عددی) می باشند و Y_i متغیر وابسته مورد نظر (برای مثال زمان بازداری) می باشد. طی فرآیند بدست آوردن مدل مناسب، فاکتورهای آماری نظیر R^2, F, t و ضریب هم بستگی میان توصیف کننده ها که نشان دهنده ی صحت مدل رگرسیون می باشند نیز محاسبه می شوند.

^۱Dependent variable^۲Independent variable

- آماره F : آزمون F یا آزمون فیشر در واقع آزمون معنی دار بودن آماری در تحلیل رگرسیون ساده و چند متغیره و مشابه آزمونهای معنی داری آماری در تحلیل واریانس است [۴]. نسبت F از رابطه (۳-۲) به دست می آید:

$$F = \frac{SS_{reg}/DF_{reg}}{SS_{res}/DF_{res}} \quad \text{معادله ی ۳-۲}$$

SS_{reg} : مجموع مجذورات رگرسیون DF_{reg} : درجه ی آزادی رگرسیون

SS_{res} : مجموع مجذورات باقی مانده ها DF_{res} : درجه آزادی باقی مانده ها

- آماره R : ضریب هم بستگی بین فاکتور تجربی (یا پیش بینی شده) و متغیرهای مستقل را نشان می دهد. مقدار این آماره بین صفر تا یک متغیر است. مقدار کوچک آن نشان دهنده این است که ارتباط خطی اندکی بین متغیر وابسته و متغیرهای مستقل وجود دارد یا چنین رابطه ای وجود ندارد.

- آماره R^2 : مجذور ضریب هم بستگی چند متغیره یا R^2 بیان گر واریانس مشترک بین متغیرهای وابسته و مستقل است. R^2 فصل مشترک دو زیر مجموعه را به صورت مقداری بیان می کند و به آن ضریب تعیین^۵ نیز می گویند، چون نسبتی از واریانس متغیر وابسته را که توسط متغیرهای مستقل به حساب آمده، معین می کند [۱۱].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{معادله ی ۴-۲}$$

که در آن y_i مقدار پیش بینی و $\hat{y}_i$ مقدار حقیقی و $\bar{y}$ مقدار میانگین می باشد.

^۱Sum of Squares of Regression

^۲Degrees of Freedom of Regression

^۳Sum of squares of Residuals

^۴Degrees of Freedom of Residuals

^۵Determination coefficient

- آماره t : آزمون معنی‌داری حضور هر متغیر مستقل در معادله رگرسیون می‌باشد. مقدار t از تقسیم مقدار ضریب هر متغیر مستقل بر خطای استاندارد آن به دست می‌آید. این مقدار برای کلیه متغیرهای موجود در معادله رگرسیون و خارج از آن محاسبه می‌شود [۱۲].
 - ضریب هم بستگی میان توصیف‌کننده‌ها یا هم‌بستگی پیرسون^۱: اندازه‌ای از ارتباط خطی بین دو متغیر مستقل است. مقادیر ضریب هم‌بستگی بین -1 تا $+1$ متغیر هستند که علامت آن نشان‌دهنده جهت ارتباط خطی (مستقیم $(+)$ یا عکس بودن $(-)$) و قدر مطلق آن بزرگی این هم بستگی را نشان می‌دهد [۱۱].
- یک مدل رگرسیون خطی بر اساس فرضهای زیر استوار است:
- ۱- خطای اندازه‌گیری متغیر مستقل ناچیز و قابل چشم پوشی است.
 - ۲- مقادیر متغیر مستقل از قبل توسط محقق معین شده است و در طول رگرسیون تغییر نخواهد کرد.
 - ۳- به ازای هر مقدار متغیر مستقل زیر مجموعه‌ای از مقادیر متغیر وابسته با پراکندگی نرمال وجود دارد.
 - ۴- میانگین‌های زیر مجموعه‌های متغیر وابسته روی یک خط واقعند. این فرض که فرض خطی بودن نیز نام دارد به شکل زیر بیان می‌شود:
- $$\mu_{y|x} = \alpha + \beta x \quad \text{معادله ۲-۵}$$
- که در آن $\mu_{y|x}$ میانگین زیر مجموعه متغیر وابسته برای مقدار خاصی از متغیر مستقل است. همچنین α و β ضرایب رگرسیون هستند.
- ۵- مقادیر متغیر وابسته از نظر آماری مستقل‌اند.

^۱Pearson correlation

فرضیهایی که ذکر شد را می توان با این رابطه‌ی (۶-۲) بیان کرد که همان مدل رگرسیون خطی است:

$$Y = \alpha + \beta x + e \quad \text{معادله‌ی ۶-۲}$$

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

۲-۱-۷- شبکه عصبی مصنوعی

شبکه‌های عصبی مصنوعی [۱۳] مجموعه‌ای از روشها، مدلها و الگوریتمهای ریاضی برای پردازش اطلاعات به صورت تقلیدشده از مغز انسان می‌باشند. مکانیسم فراگیری و آموزش مغز اساسا بر تجربه استوار است. مدل‌های الکترونیکی شبکه‌های عصبی نیز بر اساس همین الگو بنا شده‌اند. شبکه‌های عصبی به ویژه برای سر و کار داشتن با روابط غیرخطی مناسب بوده [۱۴] و برای رویارویی با تنوع بسیار زیاد از مسائل ریاضی مانند جستجوی داده، تشخیص الگو، مدل‌سازی پاسخهای مداوم و دسته بندی شده، مسائل چند پاسخی و... مناسب هستند.

مفهوم استفاده از شبکه‌های عصبی در حل مسائل در اواسط ۱۹۴۰ به وجود آمد. در سال ۱۹۴۳، مک کولوچ و پیتز^۱ [۱۵] مقاله‌ای درباره‌ی روشی جدید که امروزه از آن تحت عنوان شبکه‌های عصبی یاد می‌شود به چاپ رساندند. پس از آن در ۱۹۴۹ هب^۲ [۱۶]، روانشناسی در دانشگاه مک گیل، یک قاعده یادگیری (که امروزه به قاعده هییان معروف است) را منتشر کرد و بعضی از جنبه‌های تئوری الگوریتم آموزشی برای شبکه عصبی را شکل داد. از آن زمان تا اواسط دهه ۱۹۸۰ روش محاسبات عصبی به دلیل نتایج ضعیف آن در حل مسائل تجربی کمتر مورد توجه قرار گرفت. استفاده از شبکه‌های عصبی از سال ۱۹۸۶ و پس از انتشار کتابی توسط رامله‌هارت^۳ و مک کللاند^۴ به طور چشم‌گیری مورد استفاده قرار گرفت. پس از آن زمان کاربردهای شبکه عصبی مصنوعی در شیمی و علوم وابسته

^۱Mc Culloch & Pitts

^۲Hebb

^۳Rumelhart

^۴McClelland

به آن مانند بیوشیمی، مهندسی شیمی و داروسازی رواج پیدا کرد [۱۷]. از میان شاخه‌های مختلف شیمی، شیمی تجزیه یکی از بخشهایی است که بیشترین استفاده را از شبکه‌های عصبی مصنوعی کرده است. شبکه‌های عصبی چه در بعد آنالیز و چه در بعد پیاده‌سازی سخت‌افزاری از نظر کمی و کیفی و توانایی، در حال رشد و پیشرفت بوده و تکنیکهای مختلف محاسبات عصبی از لحاظ تعداد در حال افزایش است [۱۸].

شبکه‌های عصبی مصنوعی و الگوریتم ژنتیکی همگی جزء دسته‌ای از علوم کامپیوتری به نام هوش مصنوعی قرار می‌گیرند [۱۹]. این روشها تفاوت اساسی با سایر روشهای محاسباتی و همچنین با یکدیگر دارند. در سیستمهای ریاضی متداول، دانسته‌ها به صورت قواعد روشن و از پیش مشخص شده‌ای در سیستم وجود دارد، در حالی که این سیستمها قواعد را توسط آموزش از مثالهایی که به آنها داده می‌شود فرا می‌گیرند. استفاده از شبکه‌های عصبی مصنوعی به جای روش رگرسیون خطی چندگانه در مطالعات ارتباط کمی ساختار-فعالیت (QSAR) و ارتباط کمی ساختار-ویژگی (QSPR) خصوصاً هنگامی که ارتباط بین توصیف کننده و فعالیت/ویژگی موردنظر خطی نبوده و یا اینکه بین آنها برهم کنشهایی وجود داشته باشد باعث بهبود مدل حاصله خواهد شد. شبکه‌های عصبی در صنایع مختلف [۲۰] از جمله: صنایع فضائی هوایی، صنعت وسایل نقلیه موتوری، کار بانکی، رسیدگی فعالیت کارت اعتباری، صنایع دفاع، الکترونیک، سرگرمی، مالی [۲۱]، صنعتی، حق بیمه، فرآیند ساخت، پزشکی، اکتشاف نفت و گاز، رباتیک، گفتار، اوراق بهادار، ارتباط از راه دور و حمل و نقل کاربرد دارند.

به طور کلی تواناییهای شبکه‌های عصبی مصنوعی را می‌توان در موارد زیر خلاصه کرد:

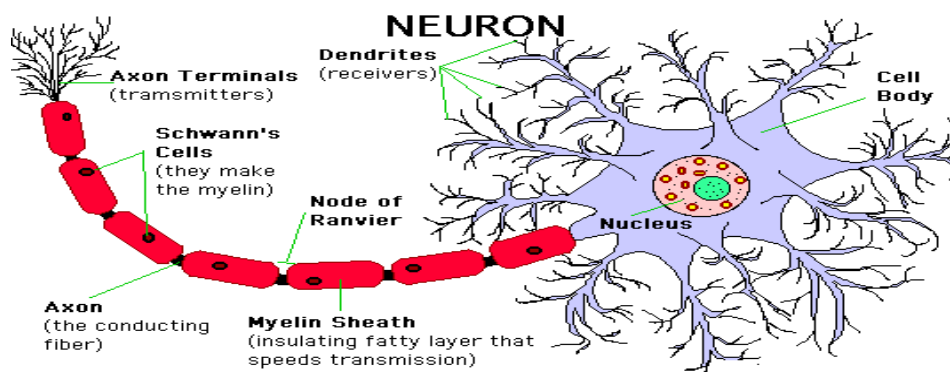
۱- یافتن و آموزش پذیری الگوهای خطی و غیرخطی

۲- پیشگویی و کلاسه بندی

۳- بهینه سازی

۲-۱-۷-۱- 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 اطلاعات با ساختار موازی از ۱۰۰ میلیارد نرون تشکیل شده است. شکل (۱-۲) نمای کلی یک نرون را نشان می‌دهد. هر نرون دارای سه قسمت دندریت^۱، جسم سلولی^۲ و اکسون^۳ است. پیام عصبی به صورت یک سیگنال الکتریکی به طور معمول از دندریت به جسم سلولی و اکسون منتقل می‌شود. بین اکسون یک نرون و دندریت نرون بعدی یک فضای خالی وجود دارد که سیناپس^۴ نامیده می‌شود. سیناپسها با کنترل انتقال پیام عصبی از یک نرون به نرون دیگر نقش مهمی در فعالیت دستگاه عصبی دارند. انتقال‌دهنده‌های عصبی مواد شیمیایی هستند که عمل انتقال پیام را در سیناپسها انجام می‌دهند. اکسون یک نرون با آزاد کردن مقادیر مشخصی از یک انتقال دهنده‌ی عصبی در فضای سیناپسی، باعث تحریک دندریت نرون مجاور شده، پیام عصبی را به آن منتقل می‌کند [۲۲].



شکل ۱-۲- ساختار یک نرون بیولوژیکی [۲۲]

غشای سلولی نقش اصلی را در ایجاد و انتقال پیام عصبی در طول یک نرون به عهده دارد. در حالت عادی یا استراحت، غشای نرونها دارای پتانسیل الکتریکی حدود ۶۵ mV - می‌باشند. وجود و

^۱Dendrite

^۲Soma

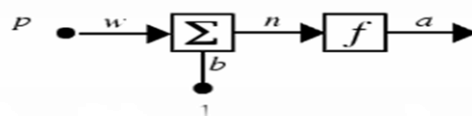
^۳Axon

^۴Synaps

دوام این مقدار از پتانسیل الکتریکی مربوط به اختلاف غلظت یونها در دو طرف غشاست. غلظت یون-های سدیم و کلر در طرف داخلی غشا بسیار بیشتر از طرف بیرونی آن است. از طرف دیگر غلظت یون پتاسیم در طرف خارجی غشا بیشتر است. پمپ سدیم-پتاسیم با مصرف انرژی مسئول نگه داشتن این اختلاف غلظت است. تحریک غشای یک نرون بوسیله انتقال دهنده‌های عصبی باعث نفوذپذیری غشای آن نسبت به یون سدیم می‌شود. تحریک نرون از ناحیه‌ی یک سیناپس موجب باز شدن کانالهای سدیمی و حرکت یونهای سدیم به درون نرون و افزایش پتانسیل غشا به حدود 45mV - می‌شود. تحریک سیناپس‌های متعدد در یک نرون اثر تجمعی روی آن داشته و پتانسیل غشا را تا حدود $20\text{mV}+$ افزایش می‌دهد. این حد از پتانسل را پتانسیل عمل گویند که موجب تحریک کامل نرون و انتقال پیام عصبی می‌شود. پس از این تحریک و انتقال پیام، پتانسیل غشا دوباره به حالت استراحت باز می‌گردد. چون بیان جزئیات فیزیولوژی نرونها از هدف این نوشتار خارج است، به بیان همین کلیات ضروری اکتفا می‌کنیم.

۲-۷-۱-۲- ساختار و عملکرد شبکه‌های عصبی مصنوعی (مدل ریاضی)

در متون فنی برای نمایش مدل ساده‌ای که در بالا تشریح گردید، به‌طور معمول از شکلی مشابه شکل زیر استفاده می‌شود.



شکل ۲-۲- مدل ریاضی یک نرون [۱۷]

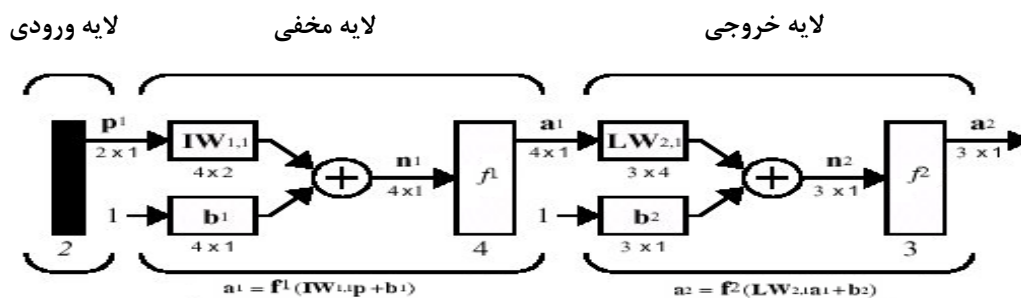
در این شکل کلاسیک، از علامت p برای نشان دادن یک سیگنال ورودی استفاده می‌شود. در واقع در این مدل، یک سیگنال ورودی پس از تقویت شدن (یا تضعیف شدن) به اندازه پارامتر w ، به‌صورت یک سیگنال الکتریکی با اندازه pw وارد نرون می‌شود. به‌جهت ساده‌سازی مدل ریاضی، فرض می‌شود که در هسته سلول عصبی، سیگنال ورودی با سیگنال دیگری به اندازه b جمع می‌گردد. در واقع

سیگنال b خود به معنی آن است که سیگنالی به اندازه واحد در پارامتری مانند b ضرب (تقویت یا تضعیف) می‌شود. ورودی خالص تابع انتقال^۱ (n) ، که یک عدد است، مجموع ورودی وزن دار PW و بایاس b است. در این جا f یک تابع انتقال است، به طور نمونه تابع خطی یا یک تابع سیگموئیدی، که ورودی n را می‌گیرد و خروجی a را تولید می‌کند. پارامتر w یا همان ضریبی که سیگنال ورودی p در آن ضرب می‌شود، در اصطلاح ریاضی به نام پارامتر وزن^۲ نیز گفته می‌شود. زمانی که از کنار هم قرار دادن تعداد بسیار زیادی از سلول‌های فوق یک شبکه عصبی بزرگ ساخته شود، شبکه‌ای در دست خواهیم داشت که رفتار آن علاوه بر تابع خروجی f ، کاملاً به مقادیر w و b وابسته خواهد بود. w و b هر دو پارامترهای عددی قابل تنظیم از نرون هستند و ایده اصلی شبکه عصبی آن است که چنین پارامترهایی می‌توانند طوری تنظیم شوند که شبکه رفتارهای مطلوب و جالبی را نشان دهد. در چنین شبکه بزرگی، تعداد بسیار زیادی از پارامترهای w و b باید توسط طراح شبکه یا خود شبکه مقداردهی شوند. این فرآیند از کار، در اصطلاح دانش شبکه‌های عصبی، به فرآیند یادگیری^۳ (آموزش) معروف است. در واقع در یک آزمایش واقعی، پس از آن که سیگنال‌های ورودی چنین شبکه بزرگی داده شدند، طراح شبکه یا خود شبکه با اندازه‌گیری خروجی و با تنظیم پارامترهای w و b به گونه‌ای که خروجی مطلوب به دست آید، شبکه را آموزش می‌دهد (در نخستین مرحله آموزش، مقادیر w و b به طور تصادفی انتخاب می‌شوند. زیرا تا این پارامترها مقدار نداشته باشند، شبکه عصبی قابل استفاده نخواهد بود)، در حین آموزش دیدن شبکه عصبی (یعنی به تدریج هم زمان با افزایش دفعاتی که مقادیر پارامترها برای رسیدن به خروجی مطلوب‌تر، تنظیم می‌شوند) مقدار پارامترها به مقدار حقیقی و نهایی خود نزدیک‌تر می‌شوند یا هم‌گرایی^۴ رخ می‌دهد. به این ترتیب پس از آن که چنین شبکه‌ای به ازای مجموعه‌ای از ورودی‌ها برای ساختن خروجی‌های مطلوب آموزش دید، می‌توان از آن برای حل مسائلی که شامل ورودیهای متفاوتی است، بهره برد.

^۱Transfer function^۲Weight^۳Learning (Training) process^۴Convergence

سلول‌های عصبی دارای ورودی‌های متعددی هستند و خروجی آنها نیز الزاما محدود به یک خروجی نیست. بر این اساس زمانی که بخواهیم از مدل‌سازی ریاضی برای مدل کردن یک سلول عصبی استفاده کنیم، به جای آن که همانند شکل (۲-۲) از یک ورودی p و یک خروجی a استفاده کنیم، از یک بردار p و یک بردار a سخن می‌گوییم. بدین ترتیب بدون آنکه نیاز به اعمال تغییری در این تصویر داشته باشیم، می‌توانیم از آن برای مدل‌سازی سلولی با n ورودی $(p_1, p_2, p_3 \dots p_n)$ و به همین ترتیب m خروجی $(a_1, a_2, a_3 \dots a_m)$ استفاده کنیم. توجه داشته باشید که در این صورت، تعداد عناصر b و w نیز به تناسب افزایش می‌یابند.

ساختار شبکه‌های عصبی امروزی، از لایه‌های نرونی تشکیل شده است. در چنین ساختاری، نرون‌ها علاوه بر آنکه در لایه خود به شکل محدودی به یکدیگر اتصال داده شده‌اند، از طریق اتصال بین لایه‌ها نیز به نرون‌های طبقات مجاور ارتباط داده می‌شوند. در شکل (۲-۳) نمونه‌ای از ساختار لایه‌ای یک شبکه عصبی مصنوعی نمایش داده شده است.



شکل ۲-۳- یک ساختار سه لایه‌ای از نرون‌ها [۱۷]

در این ساختار، گروهی از نرون‌ها از طریق ورودی‌های خود با جهان واقعی ارتباط دارند. گروه دیگری از نرون‌ها نیز از طریق خروجی‌های خود، جهان خارج را می‌سازند. در توپولوژی فوق، مابقی نرون‌ها از دید پنهان هستند. در بسیاری از شبکه‌های عصبی، اتصالات بین‌نرونی به گونه‌ای است که نرون‌های لایه‌های میانی، ورودی خود را از تمام نرون‌های لایه پایینی خود (به طور معمول لایه

نرون‌های ورودی) دریافت می‌کنند. بدین ترتیب در یک شبکه عصبی، سیگنال‌ها به تدریج از یک لایه نرونی به لایه‌های بالاتر حرکت می‌کنند و در نهایت به لایه آخر و خروجی شبکه می‌رسند. چنین مسیری در اصطلاح فنی پیش‌خور^۱ نامیده می‌شود. نوع دیگری از ارتباط بین نرونی در شبکه‌های عصبی که به پس‌خور^۲ معروف است، خروجی یک لایه نرونی به لایه قبلی (یا لایه‌های پایین‌تر) اتصال داده می‌شود.

به‌طور کلی دو روش برای آموزش دادن شبکه‌های عصبی وجود دارد: روش با سرپرست^۳ و روش بدون سرپرست^۴ [۲۳]. در یادگیری با سرپرست قاعده یادگیری توسط مجموعه‌ای از مثالها (مجموعه آموزشی) برای ایجاد شبکه مناسب فراهم می‌شود. در این شیوه ورودیها و خروجیهای منطبق بر آنها (هدفها)^۵ در دسترس هستند. زمانی که ورودیها برای شبکه به کار می‌روند، خروجیهای شبکه با هدفها مقایسه می‌شوند. سپس قاعده یادگیری برای تنظیم وزن‌ها و بایاسهای شبکه به منظور نزدیک‌تر کردن خروجیهای شبکه به هدفها استفاده می‌شود. اما در یادگیری بدون سرپرست شبکه عصبی باید به تنهایی و بدون کمک داده‌های خروجی و با توجه به شکل سیگنال‌های خروجی خود بتواند درباره درستی و نادرستی آنها تصمیم‌گیری نماید.

در روش معمول آموزش شبکه‌های عصبی، از مجموعه‌ی شناخته‌شده‌ای از داده‌های ورودی و خروجی‌های متناظر آنها (داده‌های سری آموزشی)^۶ برای آموزش دادن شبکه استفاده می‌شود. در چنین فرایندی، پس از اعمال داده‌های آموزشی، پارامترهای شبکه به تدریج به سمت مقادیر نهایی خود همگرا می‌شوند. در مواردی ممکن است که شبکه عصبی اصولاً موفق به فراگیری نشود. بدین معنی که پارامترهای شبکه پس از زمان‌های طولانی به مقدار مشخصی همگرا نشود. چنین مواردی

^۱Feed Forward

^۲Feed Back

^۳Supervised

^۴Unsupervised

^۵Target

^۶Training set

ممکن است بر اثر ناکافی بودن داده‌های آموزشی و یا اصولاً نقص طراحی شبکه ایجاد شوند. از طرف دیگر در مواردی شبکه عصبی دچار آموزش بیش از حد^۱ می‌شود. در صورتی که تعداد داده‌های آموزشی آموزشی یک شبکه عصبی بیش از اندازه زیاد باشد (در واقع از تمامی داده‌های مسئله برای آموزش دادن به شبکه استفاده شود)، شبکه عصبی به جای آنکه آموزش ببیند، به حالتی می‌رسد که به آن حفظ کردن اطلاعات می‌گویند. در واقع به جای آن که یک شبکه عصبی برای حل مسئله از هوش خود کمک بگیرد، از محفوظات خود استفاده می‌کند.

پس از آنکه یک شبکه عصبی به اندازه کافی آموزش دید، طراح یا کاربر شبکه می‌تواند پارامترهای شبکه را قفل کند. در این مرحله شبکه عصبی برای کاربرد واقعی خود و حل مسائل آماده خواهد بود.

۲-۱-۷-۳- توابع انتقال

سه تابع انتقال رایج در شبکه‌های عصبی که در این پایان‌نامه برای بهینه‌سازی شبکه به کار رفته است عبارتند از:

- تابع انتقال خطی (purelin)

نرونی‌هایی که از این تابع انتقال استفاده می‌کنند معمولاً برای تقریب خطی در فیلترهای خطی به کار می‌روند. این تابع همان مقدار ورودی را به عنوان خروجی برمی‌گرداند.

- تابع انتقال لگاریتم سیگموئید (logsig)

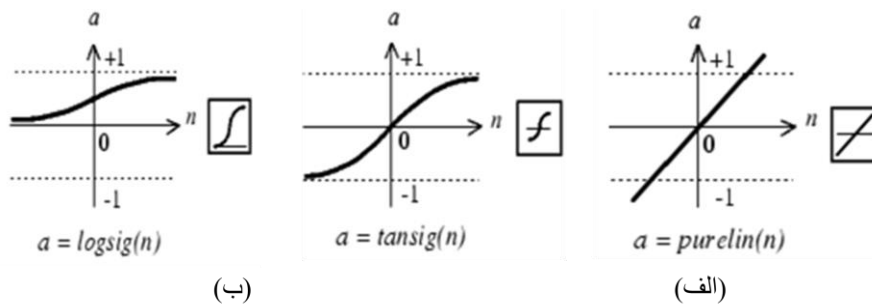
از این تابع انتقال در شبکه‌های پس انتشار استفاده می‌شود. این تابع انتقال مقادیر ورودی را در محدوده $-\infty$ و $+\infty$ دریافت کرده و خروجی بین ۰ و ۱ تولید می‌نماید.

- تابع انتقال تانژانت هایپربولیک (tansig)

این تابع انتقال مقادیر ورودی را در محدوده $-\infty$ و $+\infty$ دریافت کرده و خروجی بین ۱ و -۱ تولید می‌نماید.

^۱Over training

دو تابع سیگموئید و خطی در شکل (۲-۴) نشان داده شده‌اند [۲۴]. انتخاب تابع تبدیل بستگی به خروجی مورد نظر دارد. اگر خروجیها کیفی باشند، معمولا تابع انتقال سیگموئیدی استفاده می‌شود و اگر خروجیها کمی باشند، تابع انتقال خطی مورد استفاده قرار می‌گیرد. جدول (۲-۱) معمولترین توابع انتقال را نشان می‌دهد.



شکل ۲-۴- توابع تبدیل (الف) خطی، (ب) سیگموئیدی

جدول ۲-۱- نام و تعریف انواع مختلف توابع در شبکه عصبی

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	$\frac{e^x}{\sum e^{x_i}}$
Unit sum	$\frac{x}{\sum x_i}$
Square root	$\sqrt{x}$
Sine	Sin(x)

۲-۱-۷-۴- آموزش^۱ شبکه‌های عصبی مصنوعی

بارزترین ویژگی شبکه‌های عصبی مصنوعی آموزش پذیری آنهاست. آموزش فرآیندی است که منجر به دستیابی به وزنهای بهینه برای ورودی هر نرون شبکه عصبی می‌گردد. هدف از آموزش آن است که خروجی شبکه در نزدیکترین حد ممکن به خروجی مطلوب قرار گیرد [۲۵]. الگوریتم یادگیری شبکه یک الگوی یادگیری^۲ 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 موثر 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^۳ خطاست که در ادامه توضیح داده خواهد شد.

۲-۱-۷-۵- الگوریتم آموزشی پس انتشار خطا

پس انتشار یا انتشار خطا [۲۶ و ۲۷] یک روش متداول از یادگیری شبکه‌های عصبی مصنوعی برای چگونگی انجام یک کار معین است. این شیوه یک روش یادگیری با سرپرست بوده و برای شبکه های پیش خور بسیار مفید است. این اصطلاح مختصر شده عبارت "انتشار در جهت عقب از خطاها"^۴ است.

روش پس انتشار خطا از قانون کلی دلتا^۵ 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 رفته و در پایان یکسری خروجی بعنوان جواب حقیقی شبکه به دست می‌آید. در این مرحله اوزان اتصال ثابت است. در مرحله‌ی بازگشت، اوزان اتصال بر اساس قانون دلتا تغییر کرده و تفاضل پاسخ حقیقی شبکه و پاسخ مورد انتظار که خطا

^۱Training^۲Learning pattern^۳Backpropagation algorithm^۴Backwards propagation of errors^۵Delta rule

نامیده می‌شود، در جهت مخالف اتصالات در شبکه منتشر می‌شود و اوزان به گونه ای تغییر می‌یابند که پاسخ حقیقی شبکه به پاسخ مطلوب نزدیکتر می‌شود. مراحل الگوریتم پس انتشار را می‌توان به صورت زیر بیان کرد:

۱- شبکه یک مثال آموزشی دریافت نموده و با استفاده از اوزان موجود در شبکه

خروجیها را محاسبه می‌کند.

۲- خطا بعنوان اختلاف بین نتیجه‌ی محاسبه شده و مقدار مورد انتظار محاسبه می‌شود.

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

مهمترین بخش این الگوریتم، تنظیم اوزان نرونهاست که در آن اوزان به گونه‌ای تغییر داده می‌شود که خطا کاهش یابد؛ هر چند این تغییر بسیار کوچک است و 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 نداده و در این صورت خطا نیز کاهش بیشتری نشان نخواهد داد. در چنین حالتی شبکه ورودی را فراگرفته و آموزش پایان یافته است [۲۸]. همانطور که گفته شد در خلال آموزش، وزنها و بایاسها مکرراً برای کمینه کردن تابع عملکرد تعدیل می‌شوند. تابع عملکرد پیش فرض برای شبکه‌های پیش‌خور، میانگین مربعات خطا (MSE)^۱ بین خروجیهای شبکه و خروجیهای هدف، می‌باشد. الگوریتمهای پس انتشار مختلفی برای آموزش مدل‌های شبکه (تطبیق بایاس و وزنها) وجود دارد که به آسانی برای تخمین آماری و تئوری بهینه سازی به کار می‌روند. متداولترین الگوریتم که در الگوی آموزش پس انتشار به کار می‌رود، الگوریتم کاهش گرادیان^۲ است. در ادامه‌ی این فصل در مورد الگوریتمهای مختلف آموزش برای شبکه های پیش خور بحث می‌کنیم.

^۱Mean squared error

^۲Gradient Descent Algorithm

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 شده تا اینکه مینیمم خطا بدست آید. روابط برگشتی این الگوریتم بصورت معادله‌های (۷-۲) و (۸-۲) می‌باشد:

$$w_{ji,n+1} = w_{ji,n} + \alpha \Delta w_{ji,n} \quad \text{معادله‌ی ۷-۲}$$

که در آن $w_{ji,n+1}$ مقدار جدید وزنها و بایاسهاست. سطح خطا با استفاده از دستور زیر کاهش می‌یابد.

$$\Delta w_{ji,n} = - \frac{\partial E}{\partial w_{ji,n}} \quad \text{معادله‌ی ۸-۲}$$

که در آن $w_{ji,n}$ وزن ارتباطی بین نرون i ام و نرون j ام در گام آموزشی n ام، α سرعت یادگیری و E تابع خطاست. عملکرد الگوریتم کاهش گرادیان تا حدودی به انتخاب مقادیر اولیه‌ی وزن بستگی دارد. هر چند خطا در الگوریتم کاهش گرادیان با یک روند خطی (درجه‌ی اول) به کمترین مقدار می‌رسد، سرعت همگرایی آن کم بوده و احتمال به دام افتادن در مینیمم‌های محلی وجود دارد، در نتیجه یادگیری شبکه سخت و کند است.

همچنین روشهای آموزش دیگری مانند الگوریتمهای پس انتشار گوس- نیوتن (GN) و لونبرگ-مارکورت (LM)^۱ نیز گزارش شده‌اند. هر چند همگرایی الگوریتم گوس- نیوتن به مقدار زیادی به انتخاب مقادیر اولیه‌ی وزن بستگی دارد، ولی خواص همگرایی درجه‌ی دوم باعث سرعت بالای آن می‌شود.

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 شده که بعنوان یک الگوریتم بهینه‌سازی حد واسط روش گوس- نیوتن و الگوریتم کاهش گرادیان عمل کند.

^۱Levenberg-Marquardt

یکی از مشکلاتی که در خلال آموزش شبکه رخ می‌دهد، انطباق بیش از حد^۱ داده‌هاست. 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.

یک روش برای ارتقای عمومیت شبکه، بکارگیری شبکه‌های بزرگ برای ایجاد یک انطباق مناسب است. اگر ما از شبکه‌ی 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 باید بزرگ باشد مشکل است.

روش دیگر برای ارتقای عمومیت، تنظیم^۲ است. این روش تابع عملکرد را که عموماً مجموع میانگین مربعات خطای شبکه برای مجموعه آموزش می‌باشد، تعدیل می‌کند:

$$E_D = \frac{1}{N} \sum_{i=1}^N e_i^2 \quad \text{معادله‌ی ۹-۲}$$

که در آن 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:

$$F = \beta E_D + \alpha E_W \quad \text{معادله‌ی ۱۰-۲}$$

$$E_W = \frac{1}{N} \sum_{j,i=1}^N W_{j,i}^2 \quad \text{معادله‌ی ۱۱-۲}$$

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

^۱Overfitting

^۲Regularization

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، انطباق بیش از حد کمتری 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 فرآیند، چارچوب بایزن^۱ بوده که اولین بار توسط دیوید مک کی^۲ پیشنهاد شد. در این چارچوب، فرض می‌شود که وزنها و بایاسهای شبکه متغیرهای تصادفی با توزیع-های 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 بایزن، در قالب تابع trainbr در جعبه ابزار نرم افزار MATLAB به کار برده می‌شود.

۲-۱-۸- ارزیابی مدل

منظور از ارزیابی مدل، ارائه‌ی شاخصهای کمی است که بتوان بوسیله آن‌ها دقت و صحت نتایج ارائه شده توسط مدل را ارزیابی کرد. از این شاخصها ابتدا برای بهینه سازی پارامترهای مدل و سپس برای گزارش میزان اعتمادپذیری آن‌ها استفاده می‌شود. معمولاً برای ارزیابی شبکه داده‌ها به سه دسته‌ی آموزش^۳، ارزیابی^۴ و تست^۵ تقسیم می‌شوند. ابتدا شبکه را توسط نمونه‌های آموزشی تمرین داده و وزنهاي 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 ارزیابی 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. 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.

^۱Bayesian

^۲David Mc key

^۳Training set

^۴Validation set

^۵Test set

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

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad \text{معادله‌ی ۱۲-۲}$$

۲- خطای استاندارد^۱:

$$SE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{معادله‌ی ۱۳-۲}$$

۳- پارامتر F : معادله‌ی (۳-۲)

۴- ضریب همبستگی: معادله‌ی (۴-۲)

۵- خطای مطلق میانگین^۲:

$$MAE(\%) = \frac{100}{n} \times \sqrt{\sum_{i=1}^n |(\hat{y}_i - y_i)|} \quad \text{معادله‌ی ۱۴-۲}$$

۶- خطای نسبی پیش‌بینی^۳:

$$REP(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{معادله‌ی ۱۵-۲}$$

در رابطه‌ی بالا $\hat{y}_i$ مقدار پیش‌گویی شده‌ی نمونه‌ی i ام و y_i مقدار حقیقی یا تجربی آن است، $\bar{y}$

مربوط به میانگین تجربی کل نمونه‌ها و n تعداد کل نمونه‌ها در مدل است.

۷- مجموع مربع باقیمانده‌ها^۴ ($PRESS$): برابر مجموع مربعات تفاوت بین مقدار کمیت مشاهده شده

(y_i) و مقدار تخمین زده شده ($\hat{y}_i$) است.

^۱Standard error

^۲Mean absolute error

^۳Relative error prediction

^۴Predictive Residual Sum of Squares

$$PRESS = \sum (y_i - \hat{y}_i)^2 \quad \text{معادله ی ۱۶-۲}$$

۸- خطای استاندارد پیش بینی^۱ (SEP):

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{معادله ی ۱۷-۲}$$

۹- خطای مطلق میانگین نسبی^۲ (RMAE):

$$RMAE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \times 100 \quad \text{معادله ی ۱۸-۲}$$

۱۰-FIT: به عنوان یک پارامتر آماری برای مقایسه مدل های مختلف با تعداد متفاوت توصیف کننده به کار می رود که هر چه مقدار این پارامتر برای یک مدل بیشتر باشد، آن مدل مناسب تر می باشد. مقدار آن از رابطه زیر به دست می آید.

$$FIT = \frac{R^2(n-p-1)}{(n+p^2)(1-R^2)} \quad \text{معادله ی ۱۹-۲}$$

که R^2 مجذور ضریب همبستگی مدل، p تعداد توصیف کننده های مدل و n تعداد ترکیبات مربوط به مدل است.

۱۱- مجذور ضریب همبستگی:

$$R^2 = 1 - \frac{RSS}{TSS} \quad \text{معادله ی ۲۰-۲}$$

$$RSS = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad \text{معادله ی ۲۱-۲}$$

$$TSS = \sum_{i=1}^N (y_i - \bar{y})^2 \quad \text{معادله ی ۲۲-۲}$$

^۱Standard Error of Prediction

^۲Relative Mean Absolute Error

فصل سوم

مدل سازی QSAR مشتقات گوانین به عنوان بازدارنده‌های CDK2

با استفاده از روش رگرسیون خطی چندگانه و شبکه عصبی مصنوعی

۳-۱- معرفی بازدارنده های CDK2^۱

بازدارنده های CDK2 (کیناز وابسته به حلقه) یک سری مواد شیمیایی بوده که نقش اساسی در تنظیمات سلولی داشته و از تکثیر اضافی سلول ها توسط CDK2 که دچار فعالیت بیش از حد^۲ شده و منجر به سرطان می شوند جلوگیری می کنند [۲۹ و ۳۰]. این بازدارنده ها بطور آزمایشگاهی برای جلوگیری از سرطان بکار می روند.

همانطور که از نام این مواد پیداست، فعالیت CDK2 به طور جزئی به اتصال گروه دیگری از پروتئینها تحت عنوان سیکلینها^۳ وابسته می باشد. مثلاً سیکلینهای خانواده ی D در فاز G1 با CDK4 و CDK6 کمپلکس می دهند، هم چنین سیکلین از خانواده ی E در انتهای فاز G1 با CDK2 و سیکلین از خانواده ی A در فاز S با CDK2 کمپلکس می دهد. پس، عملکرد نادرست CDKها منجر به برهم خوردن چرخه ی سلولی شده که به طور مستقیم مرتبط با مطالعات مولکولی سرطان می باشد [۳۱].

دانشمندان در حال بررسی انواع جدیدتر بازدارنده های گزینشی و مناسب تر برای بکارگیری و جایگزینی هستند. بدین منظور روش های کمومتریکی شامل ارتباط کمی ساختار- فعالیت (QSAR) به کار گرفته شده است.

ارتباط کمی ساختار- فعالیت برای ایجاد همبستگی بین فعالیت و خصوصیات ساختارهای مولکولی استفاده می شود. در سالهای اخیر، چندین مدل QSAR بر پایه ی روشهای خطی و غیرخطی با هدف پیش بینی فعالیت ترکیبات استفاده شده است. پیش بینی مطمئن بازدارندگی CDK2 نقش بسار مهمی در تحقیقات پزشکی دارد. وظیفه ی نهایی روابط مختلف در تئوری QSAR پیشنهاد مدل های ریاضی برای تخمین پارامترهای مورد نظر، بویژه هنگامیکه این پارامترها بنا به دلایلی قابل

^۱Cyclin Dependent Kinase^۲Over active^۳Cyclins

تعیین نیستند، می باشد. این مطالعات بر این فرض استوار است که ساختار یک ترکیب، فعالیت ترکیب مرتبط را تعیین می کند. بنابراین ساختار مولکولی از طریق فرمولهای ریاضی بدست آمده از تئوریهای مختلف مانند تئوری گراف شیمیایی، تئوری اطلاعات، مکانیک کوانتوم و... به توصیف کننده های مولکولی تبدیل می شوند.

در اینجا، یک بررسی QSAR با استفاده از اطلاعات دو بعدی توپولوژیکی با پیش بینی فعالیت بازدارندگی یک سری از مشتقات گوانین به کمک روشهای رگرسیون خطی چند گانه (MLR) و شبکه نرون مصنوعی (ANN) انجام شده است.

در این تحقیق از نرم افزار Hyper Chem. 8 برای رسم ساختار ترکیبات مورد نظر، نرم افزار Dragon برای محاسبه ی توصیف کننده های مولکولی، نرم افزار SPSS 17 برای انجام آنالیزهای آماری رگرسیون خطی چند گانه و نرم افزار MATLAB 2008 به منظور برنامه نویسی و اجرای شبکه عصبی مصنوعی استفاده شد که در ادامه به معرفی آنها پرداخته می شود.

۳-۲- نرم افزارهای استفاده شده

۳-۲-۱- نرم افزار HyperChem

در ابتدا با کمک این نرم افزار، ساختارهای ترکیبات به طور تقریبی رسم شد، سپس با اضافه کردن اتمهای هیدروژن و انتخاب گزینه ی بهینه سازی هندسی از قسمت محاسبه، با کمک الگوریتم گرادیان مزدوج^۱ شکل هندسی مولکول بهینه شد. برای مقدار بیشینه ی چرخش، ۱۰۰۰۰ و برای مقدار گرادیان میانگین مجذور خطا برای تعیین پایان محاسبات، ۰/۰۰۱ انتخاب شد. این نرم افزار قابلیت های مختلفی نظیر ترسیم ساختار کلاسترها و ترکیبات پیچیده را دارد. با کمک این نرم افزار می توان اطلاعاتی مانند بار اتمها، انرژی تشکیل ترکیبات، زوایای چرخشی و طول پیوندها را بدست آورد.

^۱Conjugated Gradient Algorithm

۳-۲-۲- نرم افزار Dragon

محاسبه‌ی توصیف کننده‌های مولکولی به کمک نرم افزار Dragon انجام شد که توسط گروه کمومتریکس دانشگاه میلاندنودر ایتالیا طراحی شده است. این توصیف کننده‌های مولکولی برای ارزیابی روابط ساختار- فعالیت یا ساختار- خصوصیت استفاده می‌شوند. این نرم افزار قادر به محاسبه‌ی بیش از ۱۴۹۷ توصیف کننده بوده که به ۱۸ دسته‌ی اصلی تقسیم می‌شوند. برای استفاده‌ی بهتر از محاسبات این نرم افزار باید بهینه سازی سه بعدی ساختارها با در نظر گرفتن هیدروژنهاشان انجام شود. نرم افزار Dragon دارای این خصوصیت است که می‌تواند توصیف کننده‌های مولکولی و خصوصیات تعیین شده توسط کاربر برای یک سری از مولکولها را ادغام نموده و یک فایل خروجی کامل که به سادگی با هر نرم افزار آنالیز همبستگی قابل کاربرد است فراهم کند.

۳-۲-۳- نرم افزار SPSS

این بسته‌ی نرم افزاری، توانایی تحلیل کلی اطلاعات را دارا بوده و عمومی‌ترین نرم افزار آماری است. امروزه، آمار مجموعه‌ای مفاهیم و روشهایی است که در هر زمینه‌ی پژوهشی برای گردآوری و تغییر اطلاعات مربوط به آن و انجام نتیجه گیریها، در شرایطی که عدم قطعیتها وجود دارد بکار می‌رود. در اینجا از این نرم افزار برای بدست آوردن بهترین مدل، حذف متغیرهایی با همبستگی بالا، ارزیابی و اعتبار مدل رگرسیون خطی چند گانه استفاده شده است.

۳-۲-۴- نرم افزار MATLAB

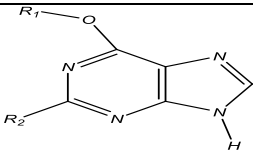
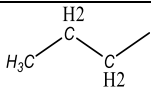
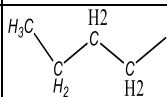
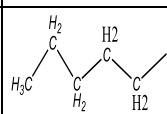
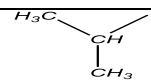
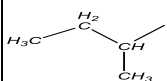
زبان برنامه‌نویسی MATLAB نرم افزاری با کارایی بالا برای محاسبات تخصصی است. از قابلیت‌های این نرم افزار می‌توان به قابلیت تجسم، شمارش و برنامه نویسی در محیطی نسبتا ساده که مسائل و راه حلها با علائم ریاضی مشابهی تعریف می‌شوند اشاره کرد. کاربردهای نوعی آن شامل ریاضیات و الگوریتم محاسباتی، مدل سازی، شبیه سازی و آنالیز اطلاعات و ... می‌باشد. این نرم افزار که امکان حل مسائل تخصصی، به خصوص مسائل ماتریسی را دارد، عناصر اولیه‌ی اطلاعات آن آرائه‌های

بدون نیاز به کادربندی یا اندازه گیری می باشند. یکی از کاربردهای مهم این نرم افزار که مورد نظر شیمی دانان قرار گرفته است، امکان استفاده از آن برای آنالیز آماری و مدل سازی داده های شیمیایی است. مدل سازی به روش شبکه عصبی مصنوعی که در تحقیق حاضر مورد استفاده قرار گرفته است از کاربردهای مهم دیگر آن است. روند کار به این صورت است که اطلاعات شیمیایی ترکیبات به صورت ماتریسی به عنوان ورودی به نرم افزار ارائه شده و در محیط MATLAB با استفاده از فرامین و آرایه های ویژه آن مدل سازی غیرخطی فعالیت شیمیایی ترکیبات امکان پذیر است.

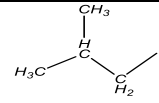
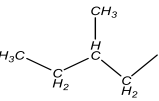
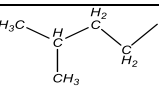
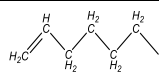
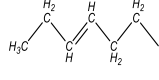
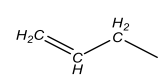
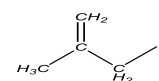
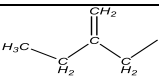
۳-۳- انتخاب سری داده های استفاده شده در مدل سازی

داده های فعالیت بیولوژیکی آزمایشگاهی به عنوان مقادیر IC_{50} برای بازدارندگی تولید شده توسط مشتقات گوانین استخلاف شده با N2 و O6 استفاده شدند [۳۲]. کل سری داده ها شامل ۵۶ ترکیب بود که نام، ساختار و مقادیر تجربی فعالیت بیولوژیکی شان (IC_{50}) برای بازدارندگی CDK2 در جدول (۱-۳) آمده است. لازم به ذکر است که سری داده ها به سه سری آموزش (۶۰٪)، ارزیابی (۲۰٪) و تست (۲۰٪) تقسیم شدند، همچنین برای مدل سازی، مقادیر IC_{50} به صورت $-\log(IC_{50} \times 10^{-6})$ وارد شدند که دارای گستره ۴/۱۱ تا ۷/۲۲ می باشد.

جدول ۳-۱- ساختار، نام و فعالیت بیولوژیکی ترکیبات شرکت کننده در مدل سازی

				
No.	Substituents		pIC50 (μM)	Chemical name of compounds
	R1	R2		
۱			۴/۱۷	6-propoxy-9H-purin-2-amine
۲			۴/۳۲	6-butoxy-9H-purin-2-amine
۳			۴/۳۱	6-pentyloxy-9H-purin-2-amine
۴			۴/۱۲	6-isppropoxy-9H-purin-2-amine
۵			۴/۶۰	6-sec-butoxy-9H-purin-2-amine

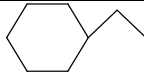
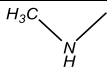
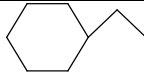
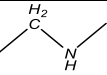
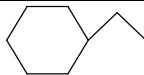
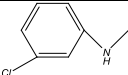
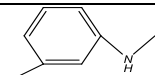
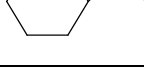
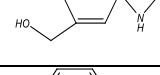
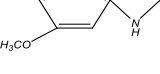
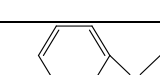
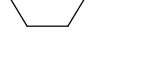
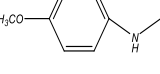
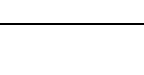
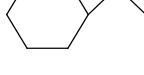
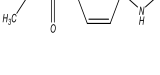
ادامه‌ی جدول ۱-۳

۶			۴/۳۸	6-isobutoxy-9H-purin-2-amine
۷			۴/۸۲	6-(2-methyl butoxy)-9H-purin-2 amine
۸			۴/۵۹	6-(isopenthyloxy)-9H-purin-2-amin
۹			۴/۳۳	6-(hex-5-enyloxy)-9H-purin-2-amine
۱۰			۴/۱۶	6-((E)- hex-3-enyloxy)-9H-purin-2-amine
۱۱			۴/۱۱	6-(allyloxy)-9H-purin-2-amine
۱۲			۴/۴۶	6-(2-methyl allyloxy)-9Hpurin-2-amine
۱۳			۴/۶۸	6-(2-methyl enebutoxy)-9H-purin-2-amine

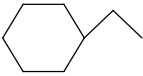
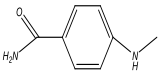
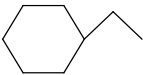
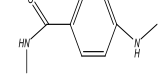
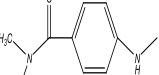
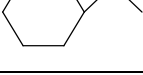
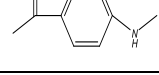
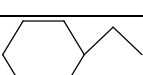
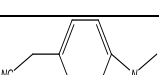
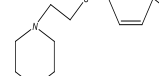
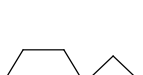
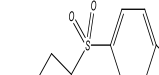
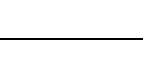
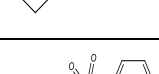
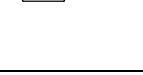
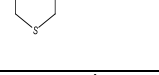
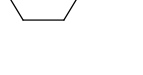
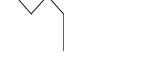
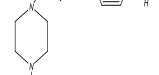
ادامه‌ی جدول ۱-۳

۱۴			۴/۸۰	6-(3-methyl-2-methylenebutoxy)-9H-purin-amine
۱۵			۴/۶۸	6-(cyclo pentylmethoxy)-9H-purin-2-amine
۱۶			۴/۵۱	6-(cyclopentenylmethoxy)-9H-purin-2-amine
۱۷			۴/۶۶	6-(cyclohexylmethoxy)-9H-purin-2-amine
۱۸			۴/۸۰	6-((cyclohex-3-enyl) methoxy)-9H-purin-2-amine
۱۹			۴/۳۶	6-(2-cyclo hexylethoxy)-9H-purin-2-amine
۲۰			۴/۴۶	6-(benzyloxy)-9H-purin-2-amine
۲۱			۴/۱۹	6-(phenethyl oxy)-9H-purin-2-amine
۲۲			۴/۷۰	6-(2,2-di ethoxypropoxy)-9H-purin-2-amine
۲۳			۴/۱۹	6-((2-isopropyl-1,3-dioxolan-2-yl)methoxy)-9H-purin-2-amine
۲۴			۴/۷۷	6-(cyclo hexylmethoxy)9H-purin-2-amine
۲۵			۶/۰۱	6-(cyclohexylmethoxy)-N-phenyl-9H-purin-2-amine

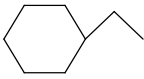
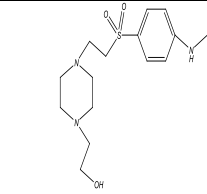
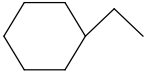
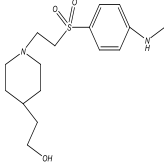
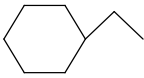
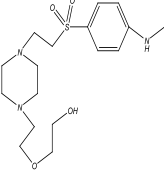
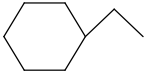
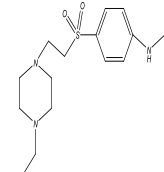
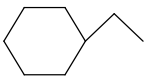
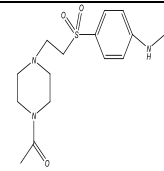
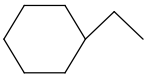
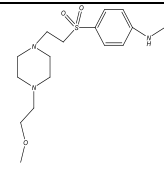
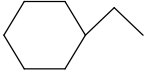
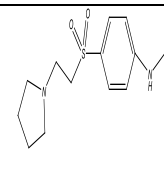
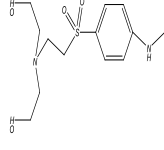
ادامه‌ی جدول ۱-۳

۲۶			۵/۳۰	6-(cyclohexylmethoxy)-N-methyl-9H-purin-2-amine
۲۷			۵/۵۵	6-(cyclohexylmethoxy)-N-ethyl-9H-purin-2-amine
۲۸			۵/۶۴	N-(3-chlorophenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۲۹			۵/۱۷	N-(3-bromophenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۳۰			۶/۴۰	(3-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)phenyl)methanol
۳۱			۵/۷۴	6-(cyclohexylmethoxy)-N-(3-methoxyphenyl)-9H-purin-2-amine
۳۲			۵/۷۷	6-(cyclohexylmethoxy)-N-(3-methylthiophenyl)-9H-purin-2-amine
۳۳			۶/۱۹	6-(cyclohexylmethoxy)-N-(4-methoxyphenyl)-9H-purin-2-amine
۳۴			۷/۲۲	6-(cyclohexylmethoxy)-N-(4-(dimethylsulfonyl)phenyl)-N,N-dimethyl-9H-purin-2-amine
۳۵			۷/۲۲	6-(cyclohexylmethoxy)-N-(4-(methylsulfonyl)phenyl)-9H-purin-2-amine
۳۶			۷/۰۰	6-(cyclohexylmethoxy)-N-(4-(methylsulfonyl)phenyl)-9H-purin-2-amine
۳۷			۶/۶۸	6-(cyclohexylmethoxy)-N-(4-(ethylsulfonyl)phenyl)-9H-purin-2-amine

ادامه‌ی جدول ۱-۳

۳۸			۶/۱۹	4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)benzamide
۳۹			۶/۷۰	4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)-N-methylbenzamide
۴۰			۶/۷۰	4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)-N,N-dimethylbenzamide
۴۱			۶/۵۲	6-(cyclohexylmethoxy)-N-(4-(prop-1-en-2-yl)phenyl)-9H-purin-2-amine
۴۲			۶/۱۰	1-(4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)phenyl)ethanol
۴۳			۶/۵۲	2-(4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)phenyl)acetonitril
۴۴			۶/۹۲	N-(4-(2-(4-methylpiperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۴۵			۶/۴۷	N-(4-(2-(piperidin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۴۶			۶/۲۷	N-(4-(2-(thiomorpholinoethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۴۷			۶/۳۵	N-(4-(2-(diethylamino)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۴۸			۶/۴۷	N-(4-(2-(4-isopropylpiperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine

ادامه‌ی جدول ۱-۳

۴۹			۶/۵۹	2-(4-(2-(4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)phenylsulfonyl)ethyl)piperazin-1-yl)ethanol
۵۰			۶/۶۲	2-(1-(2-(4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)phenylsulfonyl)ethyl)piperidin-4-yl)ethanol
۵۱			۶/۵۱	2-(2-(4-(2-(4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)phenylsulfonyl)ethyl)piperazin-1-yl)ethoxy)ethanol
۵۲			۶/۵۹	N-(4-(2-(4-(ethylpiperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۵۳			۶/۶۴	1-(4-(2-(4-(6-(cyclohexylmethoxy)-9H-purin-2-ylamino)phenylsulfonyl)ethyl)piperazin-1-yl)ethanone
۵۴			۶/۵۴	N-(4-(2-(4-(2-(methoxyethyl)piperazin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۵۵			۶/۷۴	N-(4-(2-(pyrrolidin-1-yl)ethylsulfonyl)phenyl)-6-(cyclohexylmethoxy)-9H-purin-2-amine
۵۶			۶/۹۶	6-(cyclohexylethoxy)-9H-purin

با توجه به مشکلات موجود در اندازه‌گیری فعالیت بیولوژیکی به روشهای آزمایشگاهی از جمله وجود عدم قطعیت در مقادیر فعالیت نمونه‌ی سنتز شده، بالا بودن هزینه‌ی استانداردها و بسیاری مشکلات دیگر از این قبیل، مقادیر فعالیت بیولوژیکی با استفاده از مقادیر اندازه‌گیری شده‌ی این پارامترها برای این ترکیبات مدل‌سازی شد تا بتوان با مدل طراحی شده، این پارامتر را برای سایر ترکیبات پیش‌بینی کرد. بدین منظور ابتدا مقادیر توصیف‌کننده‌ها محاسبه و سپس با استفاده از روشهای آنالیز رگرسیون خطی (MLR)، ارتباط خطی بین توصیف‌کننده‌ها و پارامتر مورد نظر برقرار گردید. سپس با استفاده از شبکه عصبی مصنوعی (ANN) یک مدل غیرخطی برای پیش‌بینی فعالیت بیولوژیکی ارائه شد.

۳-۴- محاسبه‌ی توصیف‌کننده‌ها

ساختارهای مولکولی با استفاده از الگوریتم گرادیان مزدوج بهینه شدند که این کار تا جایی ادامه می‌یابد که گرادیان RMS به $0.001 \text{ kcal mol}^{-1}$ برسد. ساختارهای هندسی حاصله به نرم افزار Dragon منتقل و تعداد ۱۴۸۱ توصیف‌کننده تولید شد. نرم افزار Dragon قادر است ۱۸ نوع توصیف‌کننده را برای ترکیبات مختلف محاسبه کند که در جدول (۳-۲) لیست این توصیف‌کننده‌ها آورده شده است.

جدول ۳-۲- هجده سری از توصیف کننده های مورد استفاده در کمومتریکیس

نام سری توصیف کننده ها	
۱- توصیف کننده های زیر ساختاری ^۱	۲- توصیف کننده های هندسی ^۲
۳- توصیف کننده های توپولوژی ^۳	۴- توصیف کننده های RDF ^۴
۵- شمارنده های مولکول ^۵	۶- توصیف کننده های سه بعدی موریس ^۶
۷- توصیف کننده های BCUT ^۷	۸- توصیف کننده های WHIM ^۸
۹- شاخصهای بار ^۹	۱۰- توصیف کننده های GETAWAY ^{۱۰}
۱۱- خودارتباطی دو بعدی ^{۱۱}	۱۲- گروه های عاملی ^{۱۲}
۱۳- توصیف کننده های بار ^{۱۳}	۱۴- اجزا میان اتمی ^{۱۴}
۱۵- شاخصهای آروماتیسیت ^{۱۵}	۱۶- توصیف کننده های تجربی ^{۱۶}
۱۷- پروفیل های مولکولی راندیک ^{۱۷}	۱۸- خصوصیات ^{۱۸}

^۱Constitutional descriptors

^۲Geometrical descriptors

^۳Topological descriptors

^۴RDF

^۵Molecular Walk Count

^۶Molecule Representation of Structure backed on Electron diffraction

^۷BCUT descriptors

^۸Weighted Holistic Invariant Molecular descriptors

^۹Charge indices

^{۱۰}Geometry, Topology, and atom weights assembly

^{۱۱}2-D autocorrelation

^{۱۲}Functional groups

^{۱۳}Charge descriptors

^{۱۴}Atom-centered fragments

^{۱۵}Aromatic indices

^{۱۶}Empirical descriptors

^{۱۷}Randic Molecular profiles

^{۱۸}Properties

۳-۵- انتخاب توصیف‌کننده‌های مناسب

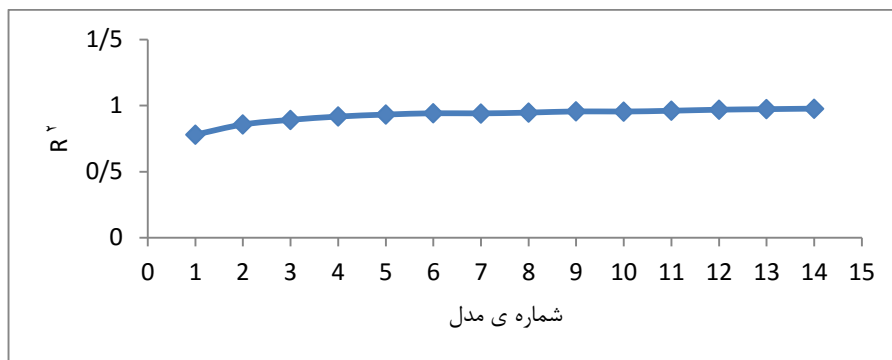
برای ساختن مدلی که بیانگر ارتباط بین ساختار ترکیبات مورد بررسی با فعالیت بیولوژیکی آنها باشد از روش رگرسیون خطی چندگانه استفاده شد. توصیف‌کننده‌های محاسبه شده توسط نرم افزار Dragon، به عنوان متغیرهای مستقل و فعالیت بیولوژیکی به عنوان متغیر وابسته وارد نرم افزار SPSS گردید. از آنجایی که تعداد زیاد توصیف‌کننده‌ها باعث پیچیدگی محاسبات شده و همچنین احتمال وجود برهم‌کنش بین توصیف‌کننده‌ها افزایش می‌یابد، پس قبل از مدل‌سازی، تعدادی از توصیف‌کننده‌ها به خاطر داشتن تعداد زیاد مقادیر صفر یا مقادیر ثابت دیگر حذف شدند، چرا که حاوی اطلاعات کافی در مورد ساختار ترکیبات نبودند. همچنین از بین توصیف‌کننده‌هایی که با یکدیگر همبستگی بیش از ۰/۹ داشتند، توصیف‌کننده‌ای که همبستگی کمتری با متغیر وابسته داشت حذف شد. در پایان جهت بدست آوردن توصیف‌کننده‌هایی که بیشترین ارتباط را با مقادیر فعالیت بیولوژیکی داشتند، روش رگرسیون مرحله‌ای بر روی تمام توصیف‌کننده‌های باقی‌مانده از مراحل قبل انجام گردید. در روش رگرسیون مرحله‌ای ابتدا متغیری که بالاترین میزان همبستگی را با متغیر وابسته دارد وارد مدل می‌شود، سپس با ورود هر متغیر جدید کلیه متغیرهای موجود در معادله مورد بررسی قرار می‌گیرند و اگر متغیری سطح معنی‌داری خود را از دست بدهد قبل از ورود متغیر جدید از مدل کنار گذاشته می‌شود. بدین ترتیب از میان تمام توصیف‌کننده‌های ارائه شده، تعداد ۱۴ مدل بعنوان مدل‌های برتر انتخاب شدند. این مدل‌ها به همراه پارامترهای آماری مربوطه در جدول (۳-۳) آورده شده‌اند.

جدول ۳-۳- کل مدل های منتخب از روش رگرسیون مرحله ای به همراه پارامترهای مربوطه

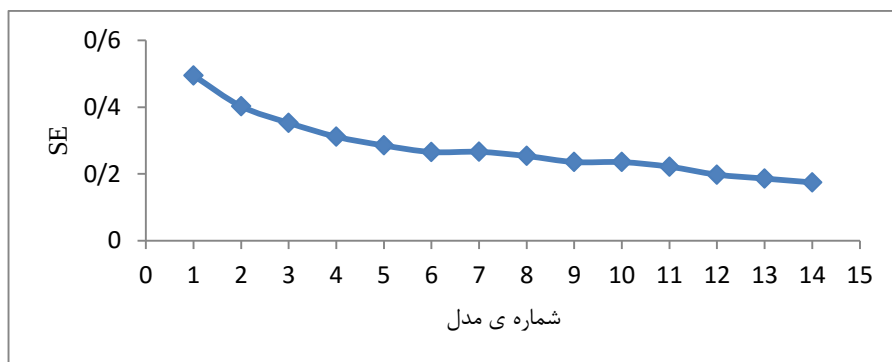
مدل	توصیف کننده	R ²	SE	F
۱	LP1	۰/۷۸۰	۰/۴۹۴۷	۱۹۱
۲	LP1, HATS6u	۰/۸۵۷	۰/۴۰۲۶	۱۵۸/۵
۳	LP1, HATS6u, Mor20u	۰/۸۹۲	۰/۳۵۲۹	۱۴۳/۲
۴	LP1, HATS6u, Mor20u, GATS2e	۰/۹۱۷	۰/۳۱۱۵	۱۴۱/۷
۵	LP1, HATS6u, Mor20u, GATS2e, R5p	۰/۹۳۲	۰/۲۸۵۳	۱۳۷/۳
۶	LP1, HATS6u, Mor20u, GATS2e, R5p, MATS2e	۰/۰۴۲	۰/۲۶۵۶	۱۳۳/۵
۷	Lp1, Mor20u, GATS2e, R5p, MATS2e	۰/۹۴۱	۰/۲۶۶۲	۱۵۹/۲
۸	Lp1, Mor20u, GATS2e, R5p, MATS2e, JGI6	۰/۹۴۷	۰/۲۵۳۶	۱۴۷/۳
۹	Lp1, Mor20u, GATS2e, R5p, MATS2e, JGI6, JGI4	۰/۹۵۶	۰/۲۳۵۵	۱۴۷/۷
۱۰	Mor20u, GATS2e, R5p, MATS2e, JGI6, JGI4	۰/۹۵۵	۰/۲۳۵۴	۱۷۲/۱
۱۱	Mor20u, GATS2e, R5p, MATS2e, JGI6, JGI4, R1m	۰/۹۶۱	۰/۲۲۱۶	۱۶۷/۶
۱۲	Mor20u, GATS2e, R5p, MATS2e, JGI6, JGI4, R1m, HATS0m	۰/۹۶۹	۰/۱۹۷۹	۱۸۵/۵
۱۳	Mor20u, GATS2e, R5p, MATS2e, JGI6, JGI4, R1m, HATS0m, Mor06v	۰/۹۷۳	۰/۱۸۶۱	۱۸۷/۲
۱۴	Mor20u, GATS2e, R5p, MATS2e, JGI6, JGI4, R1m, HATS0m, Mor06v, E1s	۰/۹۷۷	۰/۱۷۴۶	۱۹۲/۲

۳-۵-۱- انتخاب مدل مناسب

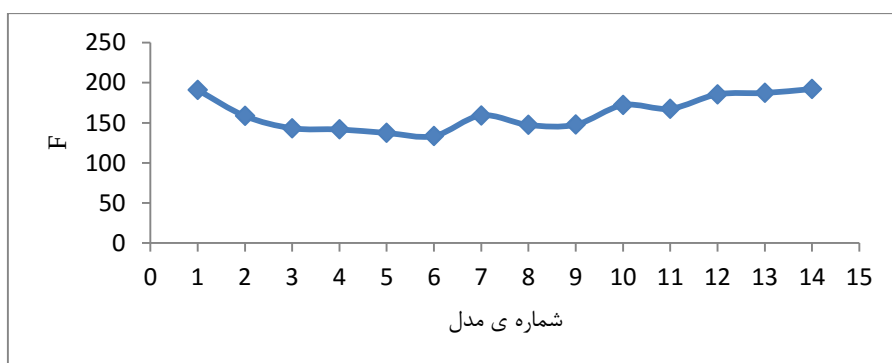
در این مرحله با توجه به پارامترهای آماری ۱۴ مدل برتر حاصل از رگرسیون مرحله ای، باید بهترین مدل انتخاب شود. شکلهای (۱-۳)، (۲-۳) و (۳-۳)، مقادیر پارامترهای آماری R²، SE و F را برحسب مدل مربوطه نشان می دهند. لازم به ذکر است که این پارامترها مربوط به کل سری داده ها می باشد.



شکل ۳-۱- تغییرات R^2 بر حسب شماره ی مدل



شکل ۳-۲- تغییرات SE بر حسب شماره ی مدل



شکل ۳-۳- تغییرات F بر حسب شماره ی مدل

با توجه به این شکلها مشاهده می شود که بهترین مدل مربوط به مدل شماره ی ۱۴ بوده، چرا که دارای بالاترین مقدار مجذور ضریب همبستگی، کمترین مقدار خطای استاندارد و در نهایت بالاترین مقدار پارامتر آماری فیشر می باشد. لازم به ذکر است که این مدل دارای ۱۰ توصیف کننده می باشد.

جدول (۳-۴)، توصیف کننده های مدل منتخب به همراه کلاس مربوط به آنها را نشان می دهد.

جدول ۳-۴- توصیف کننده های مدل منتخب به همراه کلاس مربوطه

ردیف	علامت اختصاری	کلاس	توضیحات
۱	Mor20u	3D-MoRSE	3D-MoRSE-signal 20/unweighted
۲	GATS2e	2D autocorrelation	Geary autocorrelation-lag 2/weighted by atomic Sanderson electronegativities
۳	R5p	GETAWAY	R autocorrelation of lag 5/weighted by atomic polarizabilities
۴	MATS2e	2D autocorrelation	Moran autocorrelation-lag 2/weighted by atomic Sanderson electronegativities
۵	JGI6	Galvez topol.Charge indices	mean topological charge index of order 6
۶	JGI4	Galvez topol.Charge indices	mean topological charge index of order 4
۷	R1m	GETAWAY	R autocorrelation of lag 5/weighted by atomic masses
۸	HATS0m	GETAWAY	leverage-weighted autocorrelation of lag 0/weighted by atomic masses
۹	Mor06v	3D-MoRSE	3D-MoRSE-signal 20/weighted by atomic van der waals volumes
۱۰	E1s	WHIM	1st component accessibility directional WHIM index/weighted by atomic electrotopological states

جدول (۳-۵) ماتریس همبستگی این توصیف کننده ها را نشان می دهد. همانطور که مشاهده می شود

همبستگی معناداری میان این توصیف کننده ها وجود ندارد.

جدول ۳-۵- ماتریس همبستگی برای توصیف کننده های انتخاب شده

	Mor20u	GATS2e	R5p	MATS2e	JGI6	JGI4	R1m	HATS0m	Mor06v	E1s
Mor20u	۱									
GATS2e	۰/۱۲	۱								
R5p	۰/۱۲۱	۰/۱۶۳	۱							
MATS2e	-۰/۵۸	-۰/۵۹۹	۰/۰۸۸	۱						
JGI6	۰/۶۱۴	۰/۱۱۷	-۰/۲۲	-۰/۵۲۸	۱					
JGI4	-۰/۴۵۹	۰/۳۰۴	۰/۱۰۶	۰/۱۹۹	-۰/۱۰۲	۱				
R1m	۰/۲۲۶	-۰/۲۲۵	۰/۲۳۸	۰/۰۹۷	۰/۲۴۸	-۰/۰۴۳	۱			
HATS0m	-۰/۴۵۲	-۰/۰۷۹	۰/۱۵۷	۰/۳۳۳	۰/۱۲۶	۰/۱۴۸	۰/۴۱۹	۱		
Mor06v	۰/۶۰۱	-۰/۰۴۷	-۰/۱۴۱	-۰/۵۱۲	۰/۵۵۸	-۰/۴۶۲	۰/۳۳۲	-۰/۱۳۸	۱	
E1s	۰/۳۵۴	-۰/۱۹۵	-۰/۰۷۱	-۰/۲۴	۰/۳۵۲	-۰/۲۰۵	۰/۳۷۳	-۰/۱۱۲	۰/۴۳۱	۱

برای مدل سازی سری داده ها به طور تصادفی به سه سری آموزش، ارزیابی و تست با نسبت های به ترتیب ۶۰٪، ۲۰٪ و ۲۰٪ تقسیم شدند. بدین ترتیب مدل رگرسیون خطی چندگانه با استفاده از سری آموزش طبق معادله ی (۳-۱) بدست آمد:

$$pIC50 = 11.476 + 0.684 \times Mor20u - 2.228 \times GATS2e - 18.175 \times R5p - 5.114 \times MATS2e + 65.07 \times JGI6 - 40.405 \times JGI4 + 2.155 \times R1m - 4.343 \times HATS0m - 0.149 \times Mor06v - 0.715 \times E1s$$

معادله ی ۳-۱

جدول (۳-۶) علامت توصیف کننده ها، ضرائب رگرسیون مربوطه در معادله ی MLR و اثر متوسط^۱

آنها را نشان می دهد. اثر متوسط یک متغیر مستقل با استفاده از رابطه ی (۳-۲) محاسبه می شود:

$$MeanEffect = \frac{\beta_x \times \sum x_n}{\sum S_p}$$

معادله ی ۳-۲

که در آن β_x ضریب متغیر مستقل x در مدل MLR، x_n مقدار متغیر مستقل x مورد نظر برای

ترکیب n ام و S_p مقدار فعالیت بیولوژیکی می باشد.

جدول ۳-۶- ضریب رگرسیون و اثر متوسط توصیف کننده ها

اثر متوسط	ضریب	علامت
۰/۱۴۵۱	۰/۶۸۴	Mor20u
-۰/۶۴۸۱	-۲/۲۲۸	GATS2e
-۰/۱۵۰۶	-۱۸/۱۷۵	R5p
-۰/۱۸۷۴	-۵/۱۱۴	MATS2e
۰/۲۰۰۱	۶۵/۰۷	JGI6
-۰/۳۳۶۹	-۴۰/۴۰۵	JGI4
۰/۰۸۳۷	۲/۱۵۵	R1m
-۰/۰۷۴۰	-۴/۳۴۳	HATS0m
-۰/۰۰۳۳	-۰/۱۴۹	Mor06v
-۰/۰۵۴۶	-۰/۷۱۵	E1s

جدول (۳-۷) نیز مقادیر عددی توصیف کننده های وارد شده در مدل را نشان می دهد.

^۱Mean effect

جدول ۳-۷- مقادیر عددی توصیف کننده های وارد شده در مدل برتر

No.	Mor20u	GATS2e	R5p	MATS2e	JGI6	JGI4	R1m	HATS0m	Mor06v	E1s
۱	۰/۱۷۷	۱/۷۱۳	۰/۰۳۸	۰/۲۳۴	۰/۰۱۲	۰/۰۵۶	۰/۱۸۶	۰/۰۸۶	-۱/۱۵	۰/۲۶۸
۲	۰/۰۸۴	۱/۶۴۸	۰/۰۳۵	۰/۲۵۹	۰/۰۱۱	۰/۰۵۲	۰/۱۸۸	۰/۰۸۴	-۱/۰۷۵	۰/۳۴۴
۳	-۰/۰۸۸	۱/۶۰۱	۰/۰۳۳	۰/۲۷۵	۰/۰۱۰	۰/۰۴۸	۰/۱۸۱	۰/۰۸۱	-۰/۸۸۲	۰/۴۱۱
۴	۰/۵۳۷	۱/۶۲۲	۰/۰۴۷	۰/۲۸۲	۰/۰۱۵	۰/۰۶۲	۰/۱۹۵	۰/۱۱۷	-۰/۵۱۳	۰/۳۱۰
۵	۰/۴۲۵	۱/۵۶۱	۰/۰۴۶	۰/۳۰۳	۰/۰۱۸	۰/۰۵۷	۰/۱۹۴	۰/۱۱۵	-۰/۴۹۴	۰/۳۵۷
۶	۰/۶۰۴	۱/۶۳۵	۰/۰۴۸	۰/۲۶۵	۰/۰۱۹	۰/۰۵۷	۰/۱۹۰	۰/۱۰۶	-۰/۳۲۸	۰/۳۶۴
۷	۰/۷۳۳	۱/۵۸۹	۰/۰۴۶	۰/۲۸۱	۰/۰۱۸	۰/۰۵۴	۰/۱۹۴	۰/۱۱۰	-۰/۲۱۹	۰/۴۱۲
۸	۰/۶۴۷	۱/۵۸۹	۰/۰۱۵	۰/۲۸۲	۰/۰۱۳	۰/۰۴۸	۰/۱۸۴	۰/۱۰۴	-۰/۴۹۳	۰/۴۳۱
۹	۰/۹۲۴	۱/۶۶۰	۰/۰۴۹	۰/۲۶۸	۰/۰۱۰	۰/۰۴۵	۰/۱۸۷	۰/۱۱۲	-۰/۲۹۲	۰/۵۰۰
۱۰	۱/۰۴۹	۱/۶۵۴	۰/۰۸۰	۰/۲۷۵	۰/۰۱۰	۰/۰۴۵	۰/۱۹۰	۰/۱۱۲	-۰/۵۵۹	۰/۳۵۳
۱۱	۰/۵۷۳	۱/۸۹۷	۰/۰۵۹	۰/۱۸۰	۰/۰۱۲	۰/۰۵۶	۰/۱۹۰	۰/۱۳۸	-۰/۴۶۰	۰/۲۶۲
۱۲	۰/۵۸۹	۱/۷۶۸	۰/۰۵۰	۰/۲۴۰	۰/۰۱۹	۰/۰۵۷	۰/۲۰۲	۰/۱۱۶	-۰/۶۴۹	۰/۳۶۰
۱۳	۰/۹۶۷	۱/۶۹۴	۰/۰۴۰	۰/۲۶۵	۰/۰۱۸	۰/۰۵۴	۰/۱۷۳	۰/۰۹۳	-۰/۳۳۱	۰/۴۶۴
۱۴	۰/۵۳۴	۱/۶۲۹	۰/۰۴۶	۰/۲۸۸	۰/۰۲۰	۰/۰۵۰	۰/۱۸۸	۰/۱۰۷	-۰/۵۰۸	۰/۴۳۶
۱۵	۰/۴۷۳	۱/۵۱۳	۰/۰۴۵	۰/۲۴۶	۰/۰۱۲	۰/۰۴۸	۰/۱۹۸	۰/۱۱۲	-۰/۵۹۲	۰/۴۲۵
۱۶	۰/۷۹۲	۱/۵۹۲	۰/۰۵۵	۰/۲۳۷	۰/۰۱۲	۰/۰۴۸	۰/۱۸۹	۰/۱۱۳	-۰/۵۰۵	۰/۳۹۳
۱۷	۰/۶۹۶	۱/۵۵۶	۰/۰۵۴	۰/۲۵۳	۰/۰۱۱	۰/۰۴۷	۰/۱۸۴	۰/۱۱۲	۰/۱۳۳	۰/۴۳۲
۱۸	۰/۶۸۱	۱/۵۶۲	۰/۰۳۳	۰/۲۴۶	۰/۰۱۱	۰/۰۴۷	۰/۱۷۸	۰/۰۹۱	۰/۰۵۵	۰/۵۲۶
۱۹	۰/۶۷۱	۱/۴۷۵	۰/۰۶۸	۰/۲۶۵	۰/۰۱۰	۰/۰۴۰	۰/۱۷۰	۰/۰۹۳	-۰/۰۲۸	۰/۵۷۰
۲۰	۰/۸۶۶	۱/۷۶۴	۰/۰۴۸	۰/۲۱۰	۰/۰۱۱	۰/۰۴۷	۰/۱۷۹	۰/۱۲۸	-۰/۱۹۸	۰/۳۶۵
۲۱	۰/۹۵۸	۱/۷۰۰	۰/۰۸۳	۰/۲۳۶	۰/۰۱۰	۰/۰۴۰	۰/۱۸۴	۰/۱۲۰	-۰/۳۸۴	۰/۳۴۵
۲۲	۱/۳۵۶	۲/۴۱۷	۰/۰۵۸	۰/۰۷۲	۰/۰۲۵	۰/۰۵۴	۰/۱۸۷	۰/۱۱۱	۰/۳۰۷	۰/۴۵۹
۲۳	۰/۷۶۸	۲/۲۰۳	۰/۰۵۰	۰/۱۰۷	۰/۰۲۲	۰/۰۶۲	۰/۱۷۴	۰/۱۰۵	-۰/۲۸۷	۰/۵۰۴
۲۴	۰/۶۶۷	۱/۴۹۱	۰/۰۵۳	۰/۲۵۷	۰/۰۱۱	۰/۰۴۷	۰/۱۹۷	۰/۱۱۱	-۰/۰۵۲	۰/۵۲۵
۲۵	۱/۱۷۷	۱/۵۹۶	۰/۰۳۲	۰/۲۶۱	۰/۰۱۸	۰/۰۳۶	۰/۱۴۸	۰/۰۸۰	-۰/۳۷۱	۰/۴۲۵
۲۶	۱/۰۲۸	۱/۶۹۶	۰/۰۳۲	۰/۲۲۱	۰/۰۱۴	۰/۰۴۶	۰/۱۵۹	۰/۰۶۹	۰/۱۸۴	۰/۵۰۰
۲۷	۱/۰۷۵	۱/۶۴۹	۰/۰۴۴	۰/۲۴۶	۰/۰۱۵	۰/۰۴۴	۰/۱۸۱	۰/۰۸۴	-۰/۳۷۰	۰/۴۷۸
۲۸	۰/۸۹۳	۱/۶۲۷	۰/۰۴۰	۰/۱۷۱	۰/۰۲۱	۰/۰۴۲	۰/۲۱۱	۰/۲۰۴	۰/۰۱۴	۰/۴۱۹
۲۹	۰/۴۲۹	۱/۶۱۷	۰/۰۴۰	۰/۲۰۷	۰/۰۲۱	۰/۰۴۲	۰/۳۰۱	۰/۲۲۷	۰/۱۱۰	۰/۴۱۱
۳۰	۱/۲۲۰	۱/۶۷۹	۰/۰۳۲	۰/۱۳۶	۰/۰۱۷	۰/۰۳۹	۰/۱۳۷	۰/۰۸۰	۰/۱۹۴	۰/۴۲۸
۳۱	۱/۵۱۶	۱/۹۸۷	۰/۰۳۸	۰/۱۱۸	۰/۰۱۷	۰/۰۳۹	۰/۱۳۵	۰/۰۷۵	۰/۲۳۹	۰/۳۰۸
۳۲	۱/۲۸۸	۱/۶۴۳	۰/۰۳۵	۰/۲۴۹	۰/۰۱۷	۰/۰۳۹	۰/۲۰۹	۰/۱۱۴	۰/۵۳۴	۰/۲۵۹
۳۳	۱/۴۳۳	۱/۹۸۷	۰/۰۳۸	۰/۱۱۸	۰/۰۲۰	۰/۰۴۲	۰/۱۲۶	۰/۰۷۰	۰/۲۴۴	۰/۳۷۱

ادامه‌ی جدول ۳-۷

No.	Mor20u	GATS2e	R5p	MATS2e	JGI6	JGI4	R1m	HATS0m	Mor06v	E1s
۳۴	۱/۸۹۹	۱/۴۲۰	۰/۰۳۷	۰/۲۷۳	۰/۰۲۳	۰/۰۴۵	۰/۳۳۹	۰/۰۸۷	۰/۷۱۱	۰/۴۸۹
۳۵	۱/۲۹۷	۱/۴۰۱	۰/۰۳۷	۰/۲۲۲	۰/۰۲۲	۰/۰۴۲	۰/۴۱۸	۰/۰۹۷	۰/۳۹۸	۰/۴۵۴
۳۶	۱/۲۸۵	۱/۴۹۴	۰/۰۴۱	۰/۱۶۹	۰/۰۱۹	۰/۰۴۲	۰/۳۲۸	۰/۱۰۳	۰/۲۲۱	۰/۴۵۱
۳۷	۱/۳۴۲	۱/۲۲۷	۰/۰۴۳	۰/۳۰۴	۰/۰۲۲	۰/۰۴۶	۰/۶۶۱	۰/۱۵۶	۰/۱۵۲	۰/۷۶۲
۳۸	۱/۳۱۳	۱/۳۲۴	۰/۰۴۰	۰/۲۳۲	۰/۰۱۹	۰/۰۴۱	۰/۱۴۷	۰/۰۷۷	۰/۲۸۷	۰/۴۳۵
۳۹	۱/۴۵۸	۱/۴۶۴	۰/۰۳۹	۰/۲۰۶	۰/۰۲۰	۰/۰۴۲	۰/۱۱۷	۰/۰۶۵	۰/۶۶۲	۰/۳۳۳
۴۰	۱/۸۴۴	۱/۵۹۴	۰/۰۴۱	۰/۱۸۶	۰/۰۲۲	۰/۰۴۵	۰/۱۲۲	۰/۰۷۳	-۰/۱۹۰	۰/۴۳۴
۴۱	۱/۴۴۴	۱/۴۴۴	۰/۰۴۰	۰/۱۹۳	۰/۰۱۹	۰/۰۴۲	۰/۱۳۴	۰/۰۷۳	۰/۴۸۵	۰/۳۸۳
۴۲	۰/۹۲۸	۱/۱۸۷	۰/۰۴۰	۰/۲۴۲	۰/۰۱۹	۰/۰۴۲	۰/۱۹۰	۰/۰۹۳	۰/۳۰۹	۰/۵۳۶
۴۳	۱/۸۳۴	۱/۶۲۱	۰/۰۴۰	۰/۲۲۲	۰/۰۲۱	۰/۰۴۱	۰/۱۲۸	۰/۰۶۸	-۰/۲۲۷	۰/۴۰۱
۴۴	۲/۲۶۲	۱/۷۸۵	۰/۰۴۶	۰/۱۱۸	۰/۰۲۱	۰/۰۴۷	۰/۴۲۵	۰/۰۹۵	۰/۴۵۱	۰/۴۹۹
۴۵	۱/۴۲۰	۱/۵۵۴	۰/۰۵۴	۰/۱۷۹	۰/۰۲۲	۰/۰۴۳	۰/۴۱۶	۰/۰۹۰	۱/۹۲۹	۰/۴۸۳
۴۶	۱/۴۸۱	۱/۶۱۹	۰/۰۶۲	۰/۱۵۴	۰/۰۲۲	۰/۰۴۳	۰/۴۶۵	۰/۱۵۰	۲/۲۸۴	۰/۵۲۴
۴۷	۱/۸۵۴	۱/۵۸۶	۰/۰۶۳	۰/۱۹۳	۰/۰۲۱	۰/۰۴۴	۰/۴۹۱	۰/۱۱۰	۱/۳۷۸	۰/۴۷۳
۴۸	۲/۴۴۶	۱/۷۰۱	۰/۰۸۸	۰/۱۵۰	۰/۰۲۱	۰/۰۴۶	۰/۲۲۶	۰/۱۱۷	-۰/۵۵۸	۰/۴۷۳
۴۹	۱/۸۴۱	۱/۷۴۶	۰/۰۳۲	۰/۰۷۹	۰/۰۲۱	۰/۰۴۲	۰/۱۰۴	۰/۰۴۹	۰/۷۱۱	۰/۵۵۰
۵۰	۱/۴۳۸	۱/۵۶۸	۰/۰۲۸	۰/۱۲۴	۰/۰۲۱	۰/۰۴۵	۰/۱۰۳	۰/۰۴۶	۲/۰۲۰	۰/۵۱۵
۵۱	۱/۷۲۳	۱/۹۷۸	۰/۰۴۰	۰/۰۲۶	۰/۰۲۰	۰/۰۴۲	۰/۱۸۱	۰/۰۵۱	۰/۸۶۱	۰/۵۲۶
۵۲	۱/۷۴۴	۱/۷۲۳	۰/۰۴۶	۰/۱۳۶	۰/۰۲۱	۰/۰۴۷	۰/۲۴۷	۰/۰۶۲	۰/۶۸۴	۰/۵۳۲
۵۳	۱/۵۳۱	۱/۵۲۰	۰/۰۳۴	۰/۱۵۰	۰/۰۲۱	۰/۰۴۶	۰/۱۰۷	۰/۰۵۱	۰/۵۷۱	۰/۵۶۹
۵۴	۲/۲۷۵	۲/۰۱۵	۰/۰۵۰	۰/۰۵۲	۰/۰۲۱	۰/۰۴۴	۰/۲۷۴	۰/۰۶۳	۱/۰۶۸	۰/۵۵۱
۵۵	۱/۱۹۲	۱/۵۳۲	۰/۰۳۴	۰/۱۷۳	۰/۰۲۲	۰/۰۴۳	۰/۱۰۳	۰/۰۵۱	۰/۱۳۴	۰/۴۳۵
۵۶	۱/۰۱۶	۱/۶۵۷	۰/۰۷۱	۰/۰۲۴	۰/۰۴۴	۰/۱۱۹	۰/۰۸۵	۰/۶۶۵	۰/۶۰۵	۶/۹۶۰

۳-۵-۲- مدل سازی به روش شبکه‌ی عصبی

در این پروژه، شبکه‌ی عصبی مصنوعی برای مدل سازی غیرخطی جهت پیش بینی فعالیت بیولوژیکی سری داده‌ها مورد استفاده قرار گرفت. در این تحقیق، برنامه نویسی و اجرای یک شبکه‌ی عصبی پیش خور سه لایه با الگوی آموزش پس انتشار در محیط نرم افزار MATLAB انجام شد. در این روش تابع کارایی شبکه، میانگین مجموع مربعات خطا (MSE) می باشد. شبکه‌ی طراحی شده از سه لایه تشکیل شده است که لایه‌ی اول شامل همان توصیف کننده‌های منتخب در روش MLR بوده که

به عنوان ورودی به شبکه داده شد. تعداد گره لایه‌ی دوم که در حین آموزش شبکه بهینه شد و در نهایت تعداد گره لایه سوم یا لایه‌ی خروجی به طور پیش فرض شامل یک گره بود. مقدار تجربی فعالیت بیولوژیکی نیز به عنوان هدف وارد شبکه شد. به منظور افزایش تعمیم‌پذیری شبکه، سری داده‌ها به سه دسته‌ی آموزش، ارزیابی و تست تقسیم شدند. شبکه‌ی پیش‌خور طراحی شده مقادیر ابتدایی وزن‌ها و بایاس‌ها را به صورت خودکار تعیین می‌کند. مطابق تابع اجرای مربوطه مقدار خطای شبکه از تفاضل مقدار خروجی محاسبه شده و مقدار تجربی فعالیت بیولوژیکی محاسبه می‌گردد. مقادیر وزن‌ها طی آموزش شبکه آنقدر تغییر می‌کند تا مقدار خطای شبکه به کمترین مقدار ممکن برسد. برای بدست آوردن بهترین مدل و کمترین خطا، باید ابتدا مقادیر بهینه‌ی پارامترهای موثر بر شبکه تعیین گردد که این پارامترها عبارتند از:

- تعداد ورودی شبکه
- تعداد گره لایه‌ی پنهان
- نوع تابع انتقال
- نوع تابع آموزش
- تعداد دوره‌های آموزش
- پارامتر μ (مقدار اولیه برای μ)

۳-۵-۲-۱- انتخاب اولیه‌ی مقدار وزن‌ها و بایاس

قبل از آموزش یک شبکه‌ی پیش‌خور، مقادیر ابتدایی وزن‌ها و بایاس‌ها باید تعیین گردند. تابع newff به صورت خودکار به وزن‌ها مقدار می‌دهد ولی ممکن است با دستور init آنرا به صورت غیر خودکار نیز انجام داد.

۳-۵-۲-۲- بهینه‌سازی پارامترهای شبکه

برای بهینه‌سازی پارامترهای شبکه، شبکه‌هایی با تعداد ورودی‌های از ۲ تا ۱۰ توصیف‌کننده ایجاد شدند. هر شبکه با الگوریتم آموزشی لونبرگ-مارکوارت و تنظیم بایزین با تعداد گره متفاوت از ۲ تا ۱۰ گره در لایه‌ی پنهان آموزش داده شد. این در حالی بود که برای بدست آوردن بهترین تابع انتقال در لایه‌ی پنهان مدل‌های شبکه‌ی عصبی طراحی شده، از توابع لگاریتم سیگموئید و تانژانت هایپربولیک به‌عنوان تابع فعالیت استفاده گردید و تعداد دوره‌های آموزش نیز از ۲ تا ۲۰ تغییر داده شد. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطا (MSE) به‌عنوان معیار انتخاب شد. با توجه به اینکه خروجی‌های شبکه داده‌های کمی می‌باشند، لذا تابع انتقال خطی برای لایه‌ی خروجی برگزیده شد.

۳-۵-۲-۳- انتخاب تعداد متغیرهای ورودی شبکه، الگوریتم آموزشی، نوع تابع انتقال، تعداد گره‌های لایه‌ی پنهان و تعداد دوره‌های آموزشی

در این پروژه بهینه‌سازی و انتخاب بهترین تعداد متغیرهای ورودی شبکه، الگوریتم آموزشی، نوع تابع تبدیل، تعداد نرون‌های لایه‌ی پنهان و تعداد دوره‌های آموزش به طور همزمان انجام گرفت (در این مرحله، مقدار پارامتر μ ۰/۰۰۱ در نظر گرفته شد). این کار موجب می‌شود بتوانیم نقش تمام پارامترها را به طور همزمان و یکجا بررسی نماییم. سپس در مقادیر بهینه‌ی این پارامترها، مقدار پارامتر μ از ۰/۰۰۰۵ تا ۰/۰۱ با گام‌های ۰/۰۰۰۵ تغییر داده شد تا با توجه به مقدار خطای شبکه، مقدار بهینه انتخاب گردد. بدین منظور به شبکه‌ای کاملاً تکرار پذیر نیازمندیم که بتواند به‌طور کاملاً خودکار پارامترهای شبکه را تغییر داده و مقادیر MSE را که معیار انتخاب بهترین شبکه در این پروژه می‌باشد، برای هر شبکه‌ی خاص محاسبه نماید.

با توجه به نتایج حاصله برای تعداد متغیرهای ورودی شبکه، تعداد مختلف گره‌ها، دوره‌های آموزش و همچنین توابع متفاوت آموزش و تبدیل، بهترین شبکه‌های بدست آمده بر اساس کمترین مقدار MSE در جدول (۳-۸) خلاصه شده‌اند.

جدول ۳-۸- پارامترهای شبکه‌های بهینه بدست آمده بر اساس مقادیر میانگین مربعات خطا (MSE)

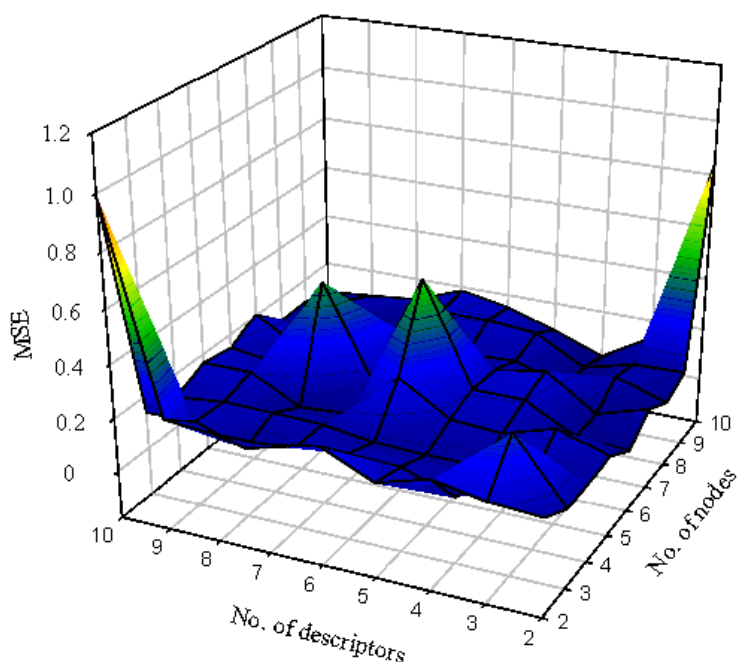
MSE	تعداد دوره‌های آموزش	تعداد گره‌ها	تابع تبدیل	تابع آموزش	تعداد توصیف‌کننده‌ها
۰/۰۱۷۱	۵	۴	لگاریتم سیگموئید	لونبرگ-مارکوارت	۱۰
۰/۰۱۸۶	۵	۶	لگاریتم سیگموئید	بایزین	۱۰
۰/۰۲۸۵	۳	۷	تانژانت هایپربولیک	لونبرگ-مارکوارت	۶
۰/۰۲۱۸	۴	۳	تانژانت هایپربولیک	بایزین	۱۰

با توجه به نتایج بدست آمده، الگوریتم آموزشی لونبرگ - مارکوارت نسبت به تنظیم بایزین دارای میانگین مربعات خطای کمتری است. همچنین تابع لگاریتم سیگموئیدی دارای خطای کمتری نسبت به تانژانت سیگموئیدی است. بنابراین تابع انتقال لگاریتم سیگموئیدی به عنوان تابع انتقال لایه پنهان استفاده شد. طبق نتایج به دست آمده تمام توصیف‌کننده‌های ورودی به شبکه، به عنوان توصیف‌کننده‌هایی که بیشترین ارتباط را با فعالیت بازدارندگی دارند، انتخاب شدند. نتایج محاسبات در جداول (۳-۹)، (۳-۱۰)، (۳-۱۱) و (۳-۱۲) و شکل‌های سه بعدی این جداول نیز در شکل‌های (۳-۴)، (۳-۵)، (۳-۶) و (۳-۷) ارائه شده است.

جدول ۳-۹ - مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

تعداد ورودی‌های شبکه

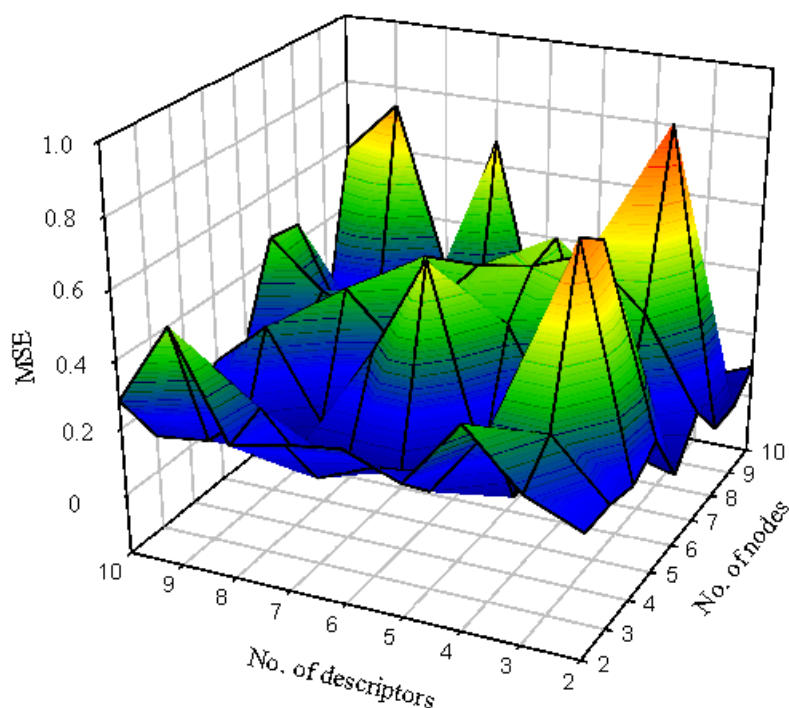
	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	
تعداد گره	۲	۰/۱۸۱۹	۰/۱۱۶۸	۰/۱۱۸۷	۰/۱۲۱۳	۰/۰۵۳۵	۰/۱۲۹۰	۰/۰۶۷۳	۰/۱۰۰۵	۰/۸۳۸۴
	۳	۰/۱۹۴۲	۰/۳۴۸۷	۰/۰۹۵۷	۰/۱۸۵۸	۰/۰۶۶۹	۰/۰۶۸۱	۰/۱۳۱۲	۰/۱۰۰۰	۰/۱۳۰۳
	۴	۰/۲۳۰۷	۰/۰۷۰۲	۰/۰۹۲۵	۰/۱۲۷۳	۰/۱۶۸۶	۰/۱۹۸۰	۰/۰۵۳۵	۰/۰۵۰۳	۰/۰۱۷۱
	۵	۰/۱۷۴۰	۰/۱۳۷۹	۰/۰۹۲۶	۰/۱۴۰۷	۰/۱۷۰۰	۰/۱۶۵۴	۰/۱۱۹۵	۰/۱۵۷۲	۰/۰۸۸۱
	۶	۰/۲۶۲۰	۰/۱۴۴۸	۰/۰۹۸۷	۰/۲۶۰۴	۰/۵۵۴۹	۰/۰۶۹۷	۰/۱۱۰۳	۰/۱۷۰۳	۰/۱۴۸۸
	۷	۰/۲۲۷۰	۰/۱۳۲۶	۰/۱۲۶۰	۰/۰۸۱۷	۰/۱۲۳۱	۰/۲۶۵۸	۰/۰۹۴۲	۰/۱۷۰۱	۰/۱۸۶۸
	۸	۰/۲۲۴۲	۰/۰۸۱۷	۰/۱۳۰۲	۰/۰۶۷۹	۰/۴۷۵۴	۰/۰۹۹۳	۰/۱۰۲۹	۰/۰۳۴۷	۰/۱۱۴۲
	۹	۰/۲۳۸۵	۰/۱۴۲۲	۰/۱۳۳۱	۰/۱۶۸۷	۰/۱۲۲۸	۰/۰۹۴۶	۰/۱۱۱۱	۰/۱۵۱۷	۰/۰۹۴۶
	۱۰	۰/۹۸۸۸	۰/۱۳۸۴	۰/۱۱۲۶	۰/۱۶۶۱	۰/۱۴۶۶	۰/۲۰۴۰	۰/۱۰۵۰	۰/۱۰۹۶	۰/۰۷۸۶



شکل ۳-۴ - نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش لونبرگ - مارکوارت و تابع تبدیل لگاریتم سیگموئیدی

جدول ۳-۱۰- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

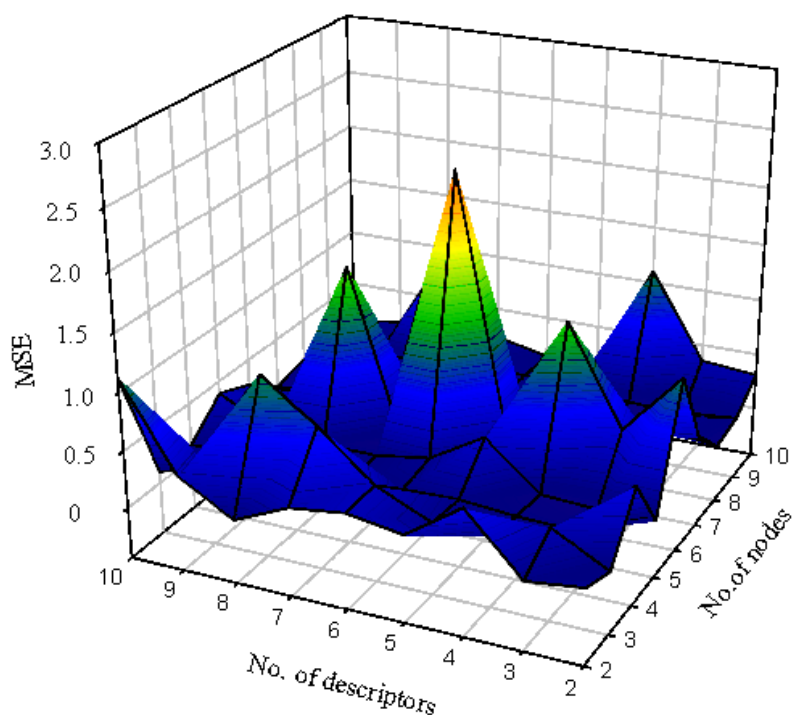
		تعداد ورودی‌های شبکه								
تعداد گره		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
	۲	۰/۱۸۵۹	۰/۱۸۵۵	۰/۱۶۱۱	۰/۲۱۷۳	۰/۰۴۱۲	۰/۱۵۷۱	۰/۰۳۷۳	۰/۰۳۱۹	۰/۰۹۵۶
	۳	۰/۲۵۵۵	۰/۳۴۸۷	۰/۸۱۲۷	۰/۷۵۶۵	۰/۰۸۸۱	۰/۴۲۷۸	۰/۹۱۶۰	۰/۰۲۵۳	۰/۰۲۷۵
	۴	۰/۴۱۶۸	۰/۳۲۷۴	۰/۰۵۶۹	۰/۰۷۰۸	۰/۰۸۸۲	۰/۵۷۲۱	۰/۰۴۹۶	۰/۲۸۰۷	۰/۰۴۴۷
	۵	۰/۲۱۴۱	۰/۱۱۴۴	۰/۰۸۵۶	۰/۰۵۲۹	۰/۳۹۷۱	۰/۵۱۰۵	۰/۵۳۵۱	۰/۰۵۰۱	۰/۱۱۸۱
	۶	۰/۲۷۸۳	۰/۱۵۱۴	۰/۶۵۱۵	۰/۶۱۷۱	۰/۴۵۸۵	۰/۴۸۵۸	۰/۰۷۳۴	۰/۳۱۳۵	۰/۰۱۸۶
	۷	۰/۲۵۵۳	۰/۰۸۱۵	۰/۳۶۷۳	۰/۳۱۷۷	۰/۲۶۴۹	۰/۴۸۵۸	۰/۰۵۰۰	۰/۰۴۱۰	۰/۶۷۶۲
	۸	۰/۲۱۹۶	۰/۲۵۷۵	۰/۳۳۹۳	۰/۰۶۷۹	۰/۴۱۸۲	۰/۲۴۴۹	۰/۳۰۰۰	۰/۱۴۹۲	۰/۲۴۳۳
	۹	۰/۵۲۷۵	۰/۱۲۳۴	۰/۴۳۹۰	۰/۳۲۹۴	۰/۲۰۹۳	۰/۳۲۱۸	۰/۱۱۸۶	۰/۰۲۸۸	۰/۷۴۳۶
	۱۰	۰/۲۸۲۸	۰/۱۰۵۸	۰/۱۰۹۷	۰/۱۹۸۷	۰/۱۹۹۴	۰/۴۶۱۱	۰/۴۴۶۷	۰/۰۹۷۴	۰/۵۸۶۹



شکل ۳-۵- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش بایزین و تابع تبدیل لگاریتم سیگموئید

جدول ۳-۱۱- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

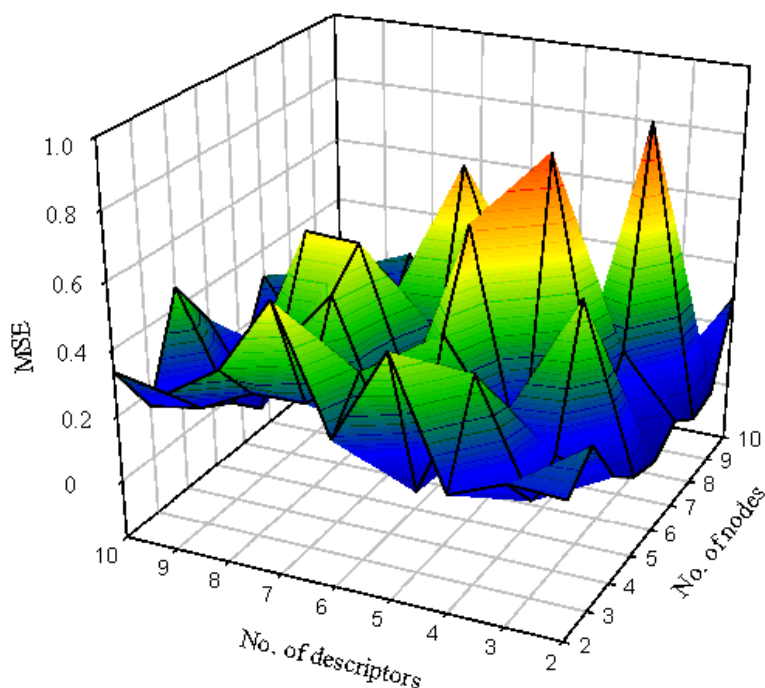
تعداد ورودی‌های شبکه										
تعداد گره		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
	۲	۰/۲۱۰۵	۰/۰۹۵۳	۰/۵۸۴۳	۰/۰۵۵۳	۱/۰۷۴۹	۰/۳۳۱۶	۰/۰۵۶۲	۰/۱۰۲۱	۰/۳۰۷۶
	۳	۰/۱۴۹۳	۰/۳۵۷۳	۰/۰۹۸۵	۰/۰۴۷۹	۰/۵۲۹۲	۰/۳۳۷۷	۰/۲۱۸۲	۰/۰۵۷۵	۰/۳۵۱۳
	۴	۰/۶۴۶۲	۰/۱۷۷۹	۰/۱۰۹۵	۰/۰۷۰۶	۱/۳۷۴۴	۰/۱۱۱۶	۰/۰۵۶۳	۰/۰۶۵۰	۱/۱۰۵۳
	۵	۰/۳۱۳۰	۰/۱۷۹۹	۰/۱۴۸۵	۰/۴۷۶۸	۰/۱۶۰۶	۰/۱۲۷۹	۰/۳۷۵۴	۰/۱۱۸۸	۰/۱۲۵۲
	۶	۰/۳۸۶۵	۰/۳۹۳۰	۰/۱۳۳۱	۰/۱۹۱۳	۲/۴۸۱۲	۰/۱۱۷۹	۰/۱۵۲۹	۰/۱۳۱۳	۰/۱۲۶۲
	۷	۰/۳۲۳۶	۰/۷۹۹۴	۰/۱۰۳۳	۰/۱۱۴۰	۰/۰۲۸۵	۰/۱۲۶۶	۰/۰۵۸۱	۰/۱۴۵۹	۰/۲۲۹۱
	۸	۰/۱۱۰۰	۱/۱۴۵۸	۰/۳۱۹۳	۰/۱۱۷۰	۱/۵۱۱۲	۰/۲۹۳۶	۰/۱۸۱۵	۰/۰۹۶۱	۰/۹۸۵۳
	۹	۰/۳۸۳۲	۰/۲۵۷۲	۰/۱۳۱۹	۰/۱۶۰۶	۰/۲۹۸۳	۰/۰۹۱۵	۰/۰۹۹۵	۰/۰۵۷۵	۰/۲۰۱۹
	۱۰	۱/۱۲۸۸	۰/۱۰۲۷	۰/۱۲۶۵	۰/۳۹۱۲	۰/۲۷۲۶	۰/۰۵۱۲	۰/۱۲۳۲	۰/۰۵۸۱	۰/۱۲۶۳



شکل ۳-۶- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش لونبرگ - مارکوارت و تابع تبدیل تانژانت سیگموئید

جدول ۳-۱۲- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

		تعداد ورودی‌های شبکه								
تعداد گره		۲	۳	۴	۵	۶	۷	۸	۹	۱۰
	۲	۰/۲۳۷۲	۰/۳۱۶۳	۰/۱۴۶۸	۰/۰۶۷۶	۰/۰۳۸۰	۰/۰۹۳۰	۰/۰۲۷۶	۰/۰۴۰۱	۰/۲۸۷۱
	۳	۰/۲۴۴۷	۰/۱۱۸۳	۰/۱۷۴۹	۰/۵۶۷۷	۰/۱۳۳۶	۰/۲۸۱۰	۰/۹۱۲۰	۰/۰۳۶۸	۰/۰۲۱۸
	۴	۰/۱۷۸۴	۰/۴۵۵۰	۰/۲۳۵۹	۰/۰۳۰۹	۰/۸۹۹۰	۰/۰۳۶۱	۰/۱۰۶۹	۰/۰۵۵	۰/۶۹۴۶
	۵	۰/۵۵۲۴	۰/۰۷۰۱	۰/۴۶۹۷	۰/۷۲۲۵	۰/۲۶۲۲	۰/۳۳۰۰	۰/۳۶۰۹	۰/۲۱۹۵	۰/۲۷۲۶
	۶	۰/۲۶۷۰	۰/۳۹۵۵	۰/۱۳۲۹	۰/۱۶۲۴	۰/۳۱۳۰	۰/۷۶۳۲	۰/۰۹۰۴	۰/۲۵۶۰	۰/۰۲۵۰
	۷	۰/۶۳۵۱	۰/۲۶۷۴	۰/۵۲۲۹	۰/۶۱۷۵	۰/۳۶۰۶	۰/۳۵۳۳	۰/۰۳۲۱	۰/۳۲۱۳	۰/۳۸۷۴
	۸	۰/۴۰۵۶	۰/۲۶۶۵	۰/۳۳۹۳	۰/۶۲۶۱	۰/۳۳۳۵	۰/۰۷۷۵	۰/۱۵۲۹	۰/۳۱۴۶	۰/۲۸۶۲
	۹	۰/۳۰۳۴	۰/۱۸۸۲	۰/۱۳۲۰	۰/۰۴۲۶	۰/۱۳۶۸	۰/۰۸۱۱	۰/۳۰۰۹	۰/۰۵۵۹	۰/۲۷۷۴
	۱۰	۰/۳۳۵۰	۰/۱۶۰۳	۰/۴۶۱۹	۰/۲۹۹۲	۰/۰۶۶۱	۰/۳۱۴۱	۰/۱۷۷۵	۰/۰۲۷۳	۰/۰۳۰۸



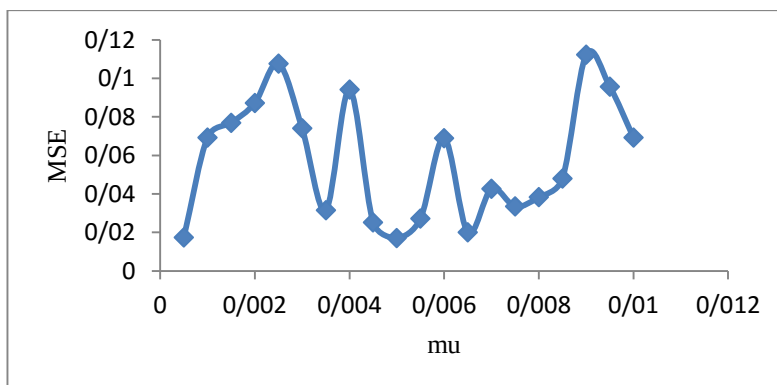
شکل ۳-۷- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش بایزینو تابع تبدیل تانژانت سیگموئید

۳-۵-۲-۴- بهینه سازی پارامتر μ

جهت بهینه کردن مقدار μ ، ساختار شبکه با ۱۰ متغیر ورودی، ۴ گره در لایه پنهان و الگوریتم آموزشی لونیگ-مارکوارت در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۵ تا ۰/۰۱ با گام‌های ۰/۰۰۵ تغییر داده شد و آنگاه برای هر مورد مقدار میانگین مربع خطای سری ارزیابی محاسبه گردید. در نهایت مقدار میانگین مربع خطا بر حسب μ رسم گردید. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد که طبق جدول (۳-۱۳) و شکل (۳-۸) مقدار بهینه ۰/۰۰۵ به دست آمد.

جدول ۳-۱۳- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ

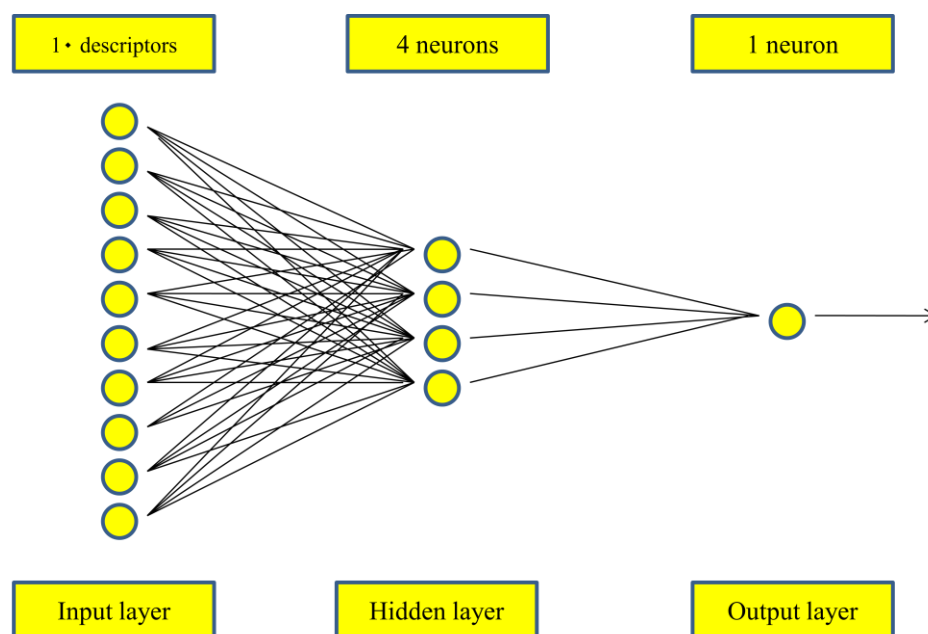
میانگین مربع خطا	μ	میانگین مربع خطا	μ
۰/۰۲۷۲	۰/۰۰۵۵	۰/۰۱۷۴	۰/۰۰۰۵
۰/۰۶۹	۰/۰۰۶	۰/۰۶۹۳	۰/۰۰۱
۰/۰۲۰۰	۰/۰۰۶۵	۰/۰۷۶۹	۰/۰۰۱۵
۰/۰۴۲۶	۰/۰۰۷	۰/۰۸۷۲	۰/۰۰۲
۰/۰۳۳۵	۰/۰۰۷۵	۰/۱۰۷۶	۰/۰۰۲۵
۰/۰۳۸۳	۰/۰۰۸	۰/۰۷۴۰	۰/۰۰۳
۰/۰۴۸	۰/۰۰۸۵	۰/۰۳۱۵	۰/۰۰۳۵
۰/۱۱۲۲	۰/۰۰۹	۰/۰۹۴۲	۰/۰۰۴
۰/۰۹۵۷	۰/۰۰۹۵	۰/۰۲۵۲	۰/۰۰۴۵
۰/۰۶۹۳	۰/۰۱	۰/۰۱۷۱	۰/۰۰۵



شکل ۳-۸- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب پارامتر μ

۳-۵-۳- معماری شبکه‌ی عصبی بهینه شده

با توجه به نتایج حاصل در مرحله‌ی بهینه‌سازی، طرح شماتیک یک شبکه‌ی عصبی پیش‌خور سه لایه با الگوی آموزش پس انتشار و الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متغیر ورودی ۱۰ و تعداد نرون لایه‌ی پنهان ۴ در شکل (۳-۱۰) نشان داده شده است.



شکل ۳-۹- طرح شماتیک ساختار هندسی شبکه‌ی عصبی مصنوعی به دست آمده پس از بهینه سازی

فصل چهارم

بحث و نتیجه گیری

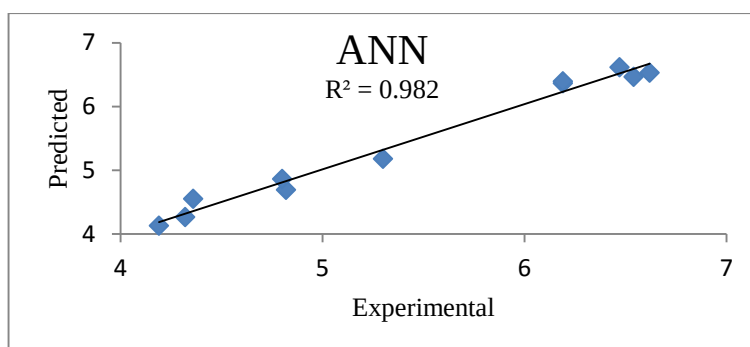
۴-۱- ارزیابی مدل‌های خطی و غیر خطی

۴-۱-۱- ارزیابی مدل‌های خطی و غیر خطی با استفاده از سری‌های ارزیابی و تست به منظور مطالعه‌ی قدرت پیش‌بینی مدل‌های منتخب خطی و غیرخطی به دست آمده، سری‌های ارزیابی و تست مورد استفاده قرار گرفتند، بدین ترتیب که کل داده‌ها به سه قسمت با نسبت‌های ۲۰٪، ۲۰٪ و ۶۰٪ تحت عنوان سری‌های ارزیابی، تست و آموزش تقسیم شدند. در این راستا در مدل خطی با استفاده از ضرائب مدل خطی مربوط به سری آموزش و در مدل غیرخطی با استفاده از شبکه عصبی بهینه شده، مقادیر pIC_{50} مربوط به سری‌های ارزیابی و تست (هر سری شامل ۱۱ ترکیب) که در ساخت مدل استفاده نشده بودند، پیش‌بینی شد (جداول (۴-۱) و (۴-۲)). با نگاهی به مقادیر مجذور ضریب همبستگی در شکل‌های (۴-۱) و (۴-۲) مربوط به سری ارزیابی و شکل‌های (۴-۳) و (۴-۴) مربوط به سری تست، توانایی بالای مدل خطی و غیرخطی در پیش‌بینی فعالیت دارویی این ترکیبات تصدیق می‌شود.

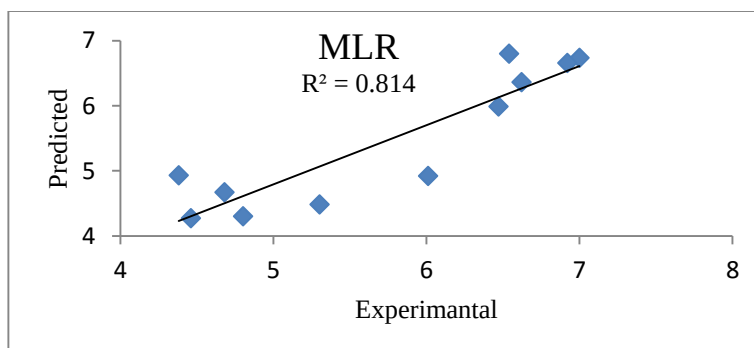
همانطور که مشاهده می‌شود، مدل غیر خطی در پیش‌بینی داده‌های سری ارزیابی موفق‌تر از مدل خطی می‌باشد که این موضوع در رابطه با سری تست صادق نیست.

جدول ۴-۱- مقادیر تجربی و پیش بینی شده ی pIC_{50} برای ترکیبات سری ارزیابی با استفاده از روش های خطی و غیرخطی

شماره ی ترکیب	pIC_{50}		مقدار پیش بینی شده		درصد خطای نسبی	
	مقدار واقعی		MLR	ANN	MLR	ANN
۲	۴/۳۲		۴/۳۰	۴/۲۶	-۰/۴۶	-۱/۳۹
۷	۴/۸۲		۴/۹۳	۴/۶۹	۲/۲۸	-۲/۷۰
۱۴	۴/۸۰		۴/۶۷	۴/۸۶	-۲/۷۱	۱/۲۵
۱۹	۴/۳۶		۴/۴۸	۴/۵۵	۲/۷۵	۴/۳۶
۲۱	۴/۱۹		۴/۲۸	۴/۱۳	۲/۱۵	-۱/۴۳
۲۶	۵/۳۰		۴/۹۲	۵/۱۸	-۷/۱۷	-۲/۲۶
۳۳	۶/۱۹		۵/۹۹	۶/۳۷	-۳/۲۳	۲/۹۱
۳۸	۶/۱۹		۶/۷۴	۶/۴۰	۸/۸۹	۳/۳۹
۴۵	۶/۴۷		۶/۶۶	۶/۶۱	۲/۹۴	۲/۱۶
۵۰	۶/۶۲		۶/۳۶	۶/۵۳	-۳/۹۳	-۱/۳۶
۵۴	۶/۵۴		۶/۸۰	۶/۴۷	۳/۹۸	-۱/۰۷



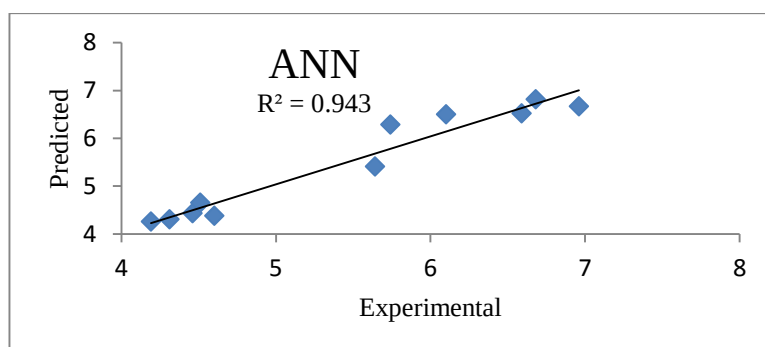
شکل ۴-۱- نمودار مقادیر پیش بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل ANN



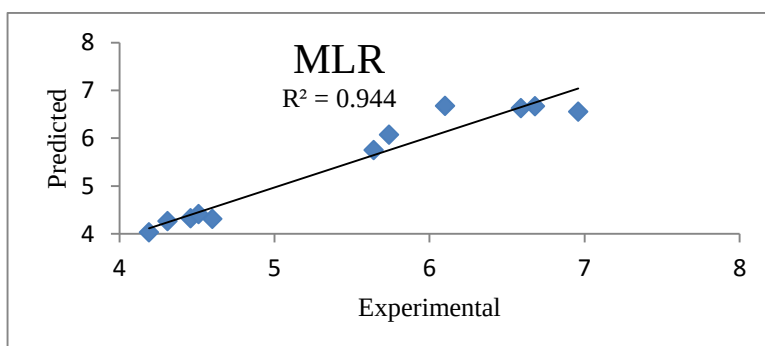
شکل ۴-۲- نمودار مقادیر پیش بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل MLR

جدول ۴-۲- مقادیر تجربی و پیش بینی شده ی pIC_{50} برای ترکیبات سری تست با استفاده از روش های خطی و غیرخطی

شماره ی ترکیب	pIC_{50}		مقدار پیش بینی شده		درصد خطای نسبی	
	مقدار واقعی		MLR	ANN	MLR	ANN
۳	۴/۳۱		۴/۲۷	۴/۳۱	-۰/۹۳	۰/۰
۵	۴/۶۰		۴/۳۱	۴/۳۸	-۶/۳۰	-۴/۷۸
۱۲	۴/۴۶		۴/۳۳	۴/۴۳	-۲/۹۱	-۰/۶۷
۱۶	۴/۵۱		۴/۴۲	۴/۶۵	-۲/۰۰	۳/۱۰
۲۳	۴/۱۹		۴/۰۴	۴/۲۶	-۳/۵۸	۱/۶۷
۲۸	۵/۶۴		۵/۷۶	۵/۴۱	۲/۱۳	-۴/۰۸
۳۱	۵/۷۴		۶/۰۸	۶/۲۹	۵/۹۲	۹/۵۸
۳۷	۶/۶۸		۶/۶۷	۶/۸۲	-۰/۱۵	۲/۱۰
۴۲	۶/۱۰		۶/۶۸	۶/۵۰	۹/۵۱	۶/۵۶
۴۹	۶/۵۹		۶/۶۳	۶/۵۳	۰/۶۱	-۰/۹۱
۵۶	۶/۹۶		۶/۵۶	۶/۶۷	-۵/۷۵	-۴/۱۷



شکل ۴-۳- نمودار مقادیر پیش بینی شده بر حسب مقادیر واقعی برای سری تست با مدل ANN



شکل ۴-۴- نمودار مقادیر پیش بینی شده بر حسب مقادیر واقعی برای سری تست با مدل MLR

۴-۱-۲- ارزیابی مدل‌های خطی و غیرخطی با استفاده از روش رد مرحله‌ای تک‌تک^۱

به منظور بررسی بیشتر مدل‌های خطی و غیرخطی، با استفاده از روش رد مرحله‌ای تک‌تک مقدار فعالیت بیولوژیکی محاسبه شد که نتایج حاصله در جدول (۳-۴) ارائه شده است. این نتایج تعمیم پذیری بالای مدل‌های طراحی شده توسط هر دو روش خطی و غیرخطی را بیان می‌کند. در روش رگرسیون خطی، یک ترکیب از سایر ترکیبات جدا شده و با بقیه‌ی ترکیبات با روش اجباری^۲ مدل ساخته و مقدار فعالیت بیولوژیکی برای ترکیب کنار گذاشته شده محاسبه می‌شود. در روش غیرخطی یک ترکیب به عنوان سری تست و بقیه‌ی ترکیبات به عنوان سری آموزش در نظر گرفته شده و شبکه‌ی بهینه مقدار فعالیت بیولوژیکی برای سری تست را پیش‌بینی می‌کند. با رسم منحنی مقدار تجربی فعالیت بیولوژیکی بر حسب مقدار پیش‌بینی شده‌ی آن و محاسبه‌ی ضریب همبستگی خطی می‌توان این دو مدل را مقایسه کرد.

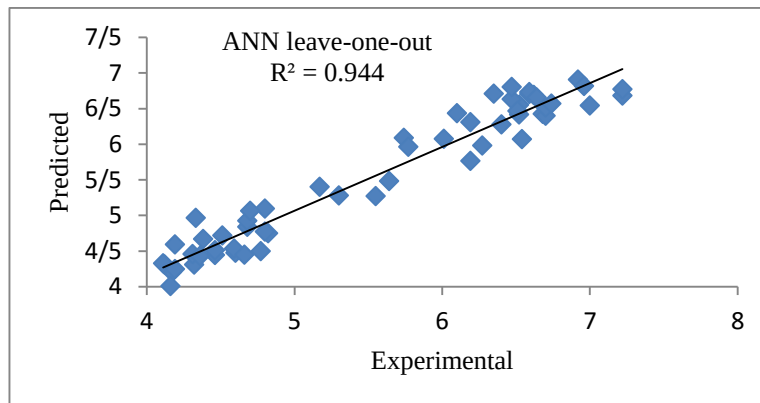
^۱Leave one out^۲Enter

جدول ۴-۳- مقادیر تجربی، پیش بینی شده و خطای حاصله به روش رد مرحله‌ای تک تک با مدل ANN و MLR

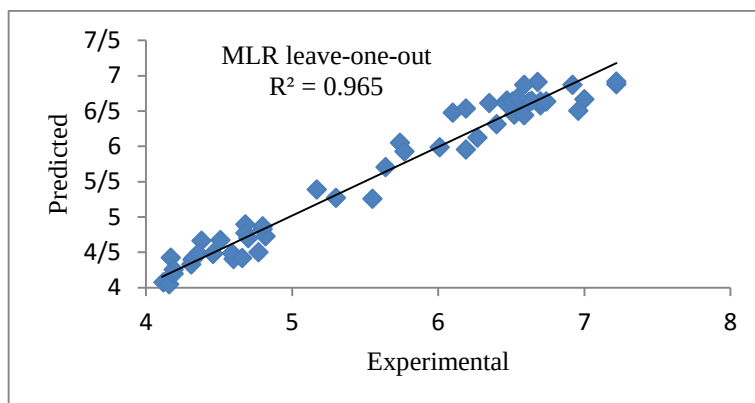
No.	pIC ₅₀	ANN	Absolute Error	MLR	Absolute Error
۱	۴/۱۷	۴/۲۱	۰/۰۴	۴/۴۲	۰/۲۵
۲	۴/۳۲	۴/۳۱	-۰/۰۱	۴/۴۰	۰/۰۸
۳	۴/۳۱	۴/۴۶	۰/۱۵	۴/۳۳	۰/۰۲
۴	۴/۱۲	۴/۳۱	۰/۱۹	۴/۰۸	-۰/۰۴
۵	۴/۶۰	۴/۴۸	-۰/۱۲	۴/۴۱	-۰/۱۹
۶	۴/۳۸	۴/۶۷	۰/۲۹	۴/۶۷	۰/۲۹
۷	۴/۸۲	۴/۷۵	-۰/۰۷	۴/۷۳	-۰/۰۹
۸	۴/۵۹	۴/۵۳	-۰/۰۶	۴/۴۹	-۰/۱۰
۹	۴/۳۳	۴/۹۶	۰/۶۳	۴/۴۲	۰/۰۹
۱۰	۴/۱۶	۴/۰۱	-۰/۱۵	۴/۰۵	-۰/۱۱
۱۱	۴/۱۱	۴/۳۳	۰/۲۲	۳/۷۶	-۰/۳۵
۱۲	۴/۴۶	۴/۴۴	-۰/۰۲	۴/۴۸	۰/۰۲
۱۳	۴/۶۸	۴/۸۴	۰/۱۶	۴/۹۰	۰/۲۲
۱۴	۴/۸۰	۵/۱۰	۰/۳۰	۴/۸۳	۰/۰۳
۱۵	۴/۶۸	۴/۹۲	۰/۲۴	۴/۷۷	۰/۰۹
۱۶	۴/۵۱	۴/۷۲	۰/۲۱	۴/۶۷	۰/۱۶
۱۷	۴/۶۶	۴/۴۵	-۰/۲۱	۴/۴۲	-۰/۲۴
۱۸	۴/۸۰	۴/۷۷	-۰/۰۳	۴/۸۶	۰/۰۶
۱۹	۴/۳۶	۴/۴۳	۰/۰۷	۴/۴۸	۰/۱۲
۲۰	۴/۴۶	۴/۵۱	۰/۰۵	۴/۴۸	۰/۰۲
۲۱	۴/۱۹	۴/۲۵	۰/۰۶	۴/۲۰	۰/۰۱
۲۲	۴/۷۰	۵/۰۶	۰/۳۶	۴/۷۰	۰/۰۰
۲۳	۴/۱۹	۴/۵۹	۰/۴۰	۴/۲۶	۰/۰۷
۲۴	۴/۷۷	۴/۵۰	-۰/۲۷	۴/۵۰	-۰/۲۷
۲۵	۶/۰۱	۶/۰۸	۰/۰۷	۶/۰۰	-۰/۰۱
۲۶	۵/۳۰	۵/۲۸	-۰/۰۲	۵/۲۷	-۰/۰۳
۲۷	۵/۵۵	۵/۲۷	-۰/۲۸	۵/۲۶	-۰/۲۹
۲۸	۵/۶۴	۵/۴۸	-۰/۱۶	۵/۷۱	۰/۰۷
۲۹	۵/۱۷	۵/۴۰	۰/۲۳	۵/۳۹	۰/۲۲
۳۰	۶/۴۰	۶/۲۸	-۰/۱۲	۶/۳۲	-۰/۰۸
۳۱	۵/۷۴	۶/۰۹	۰/۳۵	۶/۰۵	۰/۳۱
۳۲	۵/۷۷	۵/۹۶	۰/۱۹	۵/۹۳	۰/۱۶
۳۳	۶/۱۹	۵/۷۶	-۰/۴۳	۵/۹۶	-۰/۲۳

ادامه‌ی جدول ۳-۴

No.	pIC ₅₀	ANN	Absolute Error	MLR	Absolute Error
۳۴	۷/۲۲	۶/۴۸	-۰/۷۴	۶/۸۸	-۰/۳۴
۳۵	۷/۲۲	۶/۷۷	-۰/۴۵	۶/۹۲	-۰/۳۰
۳۶	۷/۰۰	۶/۵۴	-۰/۴۶	۶/۶۷	-۰/۳۳
۳۷	۶/۶۸	۶/۴۳	-۰/۲۵	۶/۹۱	۰/۲۳
۳۸	۶/۱۹	۶/۳۱	۰/۱۲	۶/۵۴	۰/۳۵
۳۹	۶/۷۰	۶/۴۰	-۰/۳۰	۶/۵۸	-۰/۱۲
۴۰	۶/۷۰	۶/۵۲	-۰/۱۸	۶/۶۳	-۰/۰۷
۴۱	۶/۵۲	۶/۴۲	-۰/۱۰	۶/۵۵	۰/۰۳
۴۲	۶/۱۰	۶/۴۴	۰/۳۴	۶/۴۸	۰/۳۸
۴۳	۶/۵۲	۶/۵۵	۰/۰۳	۶/۴۵	-۰/۰۷
۴۴	۶/۹۲	۶/۹۱	-۰/۰۱	۶/۸۷	-۰/۰۵
۴۵	۶/۴۷	۶/۶۳	۰/۱۶	۶/۶۵	۰/۱۸
۴۶	۶/۲۷	۵/۹۸	-۰/۲۹	۶/۱۲	-۰/۱۵
۴۷	۶/۳۵	۶/۷۱	۰/۳۶	۶/۶۱	۰/۲۶
۴۸	۶/۴۷	۶/۸۰	۰/۳۳	۶/۶۰	۰/۱۳
۴۹	۶/۵۹	۶/۷۱	۰/۱۲	۶/۸۷	۰/۲۸
۵۰	۶/۶۲	۶/۶۹	۰/۰۷	۶/۶۲	۰/۰
۵۱	۶/۵۱	۶/۴۷	-۰/۰۴	۶/۶۴	۰/۱۳
۵۲	۶/۵۹	۶/۷۳	۰/۱۴	۶/۴۴	-۰/۱۵
۵۳	۶/۶۴	۶/۶۶	۰/۰۲	۶/۶۵	۰/۰۱
۵۴	۶/۵۴	۶/۰۷	-۰/۴۷	۶/۶۸	۰/۱۴
۵۵	۶/۷۴	۶/۵۷	-۰/۱۷	۶/۶۴	-۰/۱۰
۵۶	۶/۹۶	۶/۸۲	-۰/۱۴	۶/۵۰	-۰/۴۶

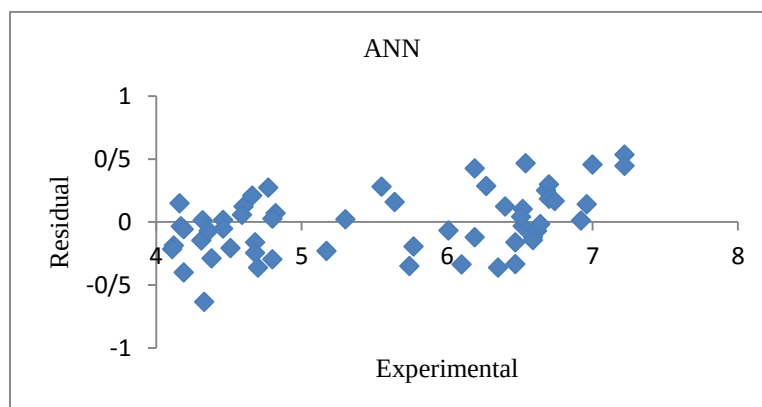


شکل ۴-۵- نمودار مقادیر پیش بینی شده pIC_{50} با استفاده از شبکه‌ی عصبی مصنوعی بهینه شده برحسب مقادیر واقعی به روش رد مرحله ای تک تک

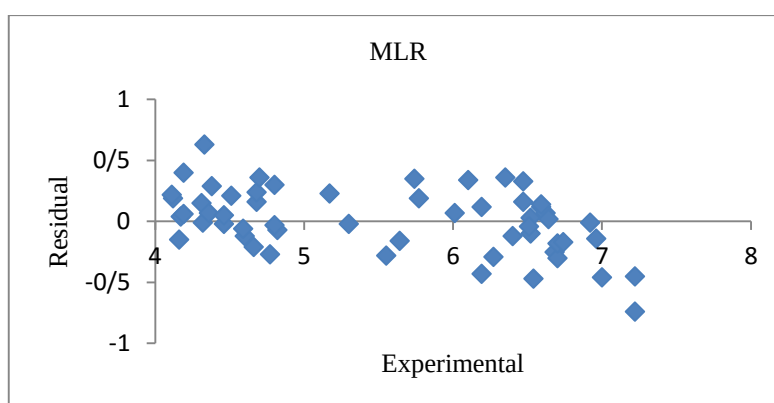


شکل ۴-۶- نمودار مقادیر پیش بینی شده pIC_{50} با استفاده از روش خطی برحسب مقادیر واقعی به روش رد مرحله‌ای تک تک

همچنین در شکل‌های (۷-۴) و (۸-۴) مقادیر باقیمانده‌ها بر حسب مقادیر واقعی pIC_{50} ترکیبات مورد بحث، ترسیم شده است که عدم وجود خطای سیستماتیک را نشان می‌دهد. تمرکز نقاط به سمت محورهای افقی نشان‌دهنده‌ی نزدیک بودن مقادیر پیش‌بینی شده به مقادیر واقعی pIC_{50} است.



شکل ۴-۷- نمودار مقادیر خطای مطلق در ارزیابی مدل شبکه‌ی عصبی مصنوعی به روش رد مرحله‌ای تک‌تک در مقابل مقادیر تجربی



شکل ۴-۸- نمودار مقادیر خطای مطلق در ارزیابی مدل خطی به روش رد مرحله‌ای تک‌تک در مقابل مقادیر تجربی

۴-۱-۳- ارزیابی مدل‌های ارائه شده با استفاده از پارامترهای آماری

جهت بررسی و مقایسه‌ی میزان کارایی دو روش خطی و غیرخطی در پیش‌بینی فعالیت بیولوژیکی چندین پارامتر آماری محاسبه گردید. آماره‌های این دو روش در جدول (۴-۴) آمده است.

جدول ۴-۴- مقایسه‌ی آماره‌های بدست آمده از دو مدل ANN و MLR

پارامتر	سری تست (N=11)		سری ارزیابی (N=11)	
	MLR	ANN	MLR	ANN
MAE	۱۳/۴۳	۱۳/۲۹	۱۳/۸۲	۱۰/۴۰۵
MSE	۰/۰۶۹	۰/۰۶۳	۰/۰۶۵	۰/۰۱۷۱
PRESS	۰/۰۰۱۴	۰/۱۹۶۹	۰/۱۰۸۹	۰/۰۶۲۵
SEP	۰/۱۳۹۴	۰/۱۳۳۳	۰/۱۳۴۹	۰/۰۶۹۳
R ²	۰/۹۴۴	۰/۹۴۳	۰/۸۱۴	۰/۹۸۲
REP (%)	۰/۰۶۲۲	۰/۷۴۲۲	۰/۵۵۱۸	۰/۴۱۸۱
RMAE	۳/۶۰۷	۳/۴۳۳	۳/۶۸	۲/۲۲۱

تمام روابط مربوط به آماره‌های محاسبه شده در این قسمت در بخش (۲-۱-۸) آورده شده است.

مقادیر این پارامترها نشان می‌دهد که مدل‌های ارائه شده از قدرت پیشگویی مناسبی برخوردار

هستند.

۴-۱-۴- ارزیابی شبکه با استفاده از Y - تصادفی^۱

به منظور اجتناب از همبستگی‌های تصادفی و تضمین قدرت شبکه، تست Y - تصادفی انجام شد. لذا داده‌های pIC_{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 بار اجرای این آزمون در جدول (۴-۵) نشان داده شده است که مجذور ضریب همبستگی پایین، نشان-دهنده‌ی عدم وابستگی شانسی در مدل توسعه یافته توسط شبکه عصبی مصنوعی است.

جدول ۴-۵- مقادیر مجذور ضریب همبستگی سری تست بعد از چندین Y - تصادفی

تکرار	R^2 سری تست
۱	۰/۲۹
۲	۰/۰۲
۳	۰/۰۰
۴	۰/۰۹
۵	۰/۰۰
۶	۰/۰۰
۷	۰/۴۶
۸	۰/۰۱
۹	۰/۱۳
۱۰	۰/۰۱

^۱Y - Randomization

۴-۲- بررسی ارتباط توصیف‌کننده‌های ساختاری وارد شده در مدل با فعالیت

بیولوژیکی ترکیبات مورد بررسی

الف) توصیف‌کننده‌های 3D - MoRSE

توصیف‌کننده‌های 3D - MoRSE (نمایش سه بعدی ساختار مولکول براساس تفرق الکترون) از طریق معادله تبدیلی که در پراش الکترون استفاده می‌شود، محاسبه می‌گردند:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad \text{معادله ی ۴-۱}$$

I شدت الکترون پراکنده شده، A_i و A_j خاصیت اتمی اتم i و j ، s زاویه پراکندگی، r_{ij} فاصله بین اتم‌های i و j ، N تعداد کل اتم‌ها را نشان می‌دهد. این روش باعث می‌شود که ساختار سه بعدی مولکول به یک کد ثابت تبدیل شود.

این توصیف‌کننده‌ها قادرند ارتباط بین ساختار سه بعدی ترکیبات آلی و خصوصیات فیزیکی، شیمیایی و بیولوژیکی را فراهم سازند. در تحقیق حاضر، برای پیش‌بینی فعالیت بیولوژیکی، توصیف‌کننده‌های Mor20u و Mor06v وارد مدل برتر شدند. توصیف‌کننده‌ی Mor20u دارای اثر متوسط مثبت بوده که انتظار می‌رود افزایش مقادیر آن همراه با افزایش مقادیر فعالیت بیولوژیکی باشد. با نگاهی به جدول (۳-۶) مربوط به توصیف‌کننده‌ها مشاهده می‌شود که مقادیر این توصیف‌کننده به طور نامنظمی از بالا به پایین در حال افزایش بوده که افزایش مقادیر فعالیت بیولوژیکی را توجیه می‌کند. توصیف‌کننده‌ی Mor06v که با حجم‌های واندروالس اتمی وزن‌دار شده است و نشان‌دهنده‌ی نقش مهم حجم واندروالس ترکیبات در تغییرات فعالیت بیولوژیکی می‌باشد، دارای اثر متوسط منفی بوده و افزایش مقادیر آن منجر به کاهش مقادیر فعالیت بیولوژیکی می‌شود.

ب) توصیف‌کننده‌های 2D – autocorrelation

توصیف‌کننده‌های دوبعدی خودارتباطی از گراف مولکولی و از طریق محاسبه‌ی مجموع اوزان اتم-های انتهایی کل مسیرها با طول مسیر مورد نظر (lag)، به دست می‌آیند. نوعی از این توصیف‌کننده‌ها گروه Geary Autocorrelation است که ضریب گری نام دارد و بدین صورت محاسبه می‌شود:

$$C(d) = \frac{\frac{1}{2\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - w_j)^2}{\frac{1}{A-1} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad \text{معادله‌ی ۲-۴}$$

که w یک ویژگی اتم، $\bar{w}$ میانگین مقدار آن ویژگی روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است و δ_{ij} نیز - که به تابع کرونکر^۱ معروف است - در حالتی که $d_{ij} = d$ باشد یک است، و در غیر اینصورت صفر است و Δ هم مجموعه‌ی δ هاست. مقدار این توصیف‌کننده که از جنس فاصله است، از صفر تا بی‌نهایت متغیر است.

نوع دیگری از این توصیف‌کننده‌ها، توصیف‌کننده‌های گروه Moran Autocorrelation هستند که از طریق رابطه زیر محاسبه می‌شوند:

$$I(d) = \frac{\frac{1}{2} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - \bar{w})(w_j - \bar{w})}{\frac{1}{A} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad \text{معادله‌ی ۳-۴}$$

پارامترها در این رابطه مشابه رابطه‌ی (۲-۴) می‌باشد.

از این گروه توصیف‌کننده‌ها، GATS2e و MATS2e وارد مدل برتر شده‌اند که نقش خاصیت فیزیکولوژی الکترون‌گاتیویته‌ی ساندerson را بیان می‌کنند. هر دو توصیف‌کننده دارای اثر متوسط

^۱ Kronecker

منفی بوده، بدین معنی که افزایش مقادیر این توصیف‌کننده‌ها باعث کاهش فعالیت بیولوژیکی می‌شود.

پ) توصیف‌کننده‌های GETAWAY

این توصیف‌کننده‌ها با توجه به مختصات فضایی اتم‌ها در یک مولکول به راحتی قابل محاسبه هستند که بر اساس ماتریس قدرت نفوذ^۱ تعریف می‌شوند و بیان کننده‌ی ویژگی‌های هندسی و توپولوژیک مولکول‌ها هستند. ماتریس قدرت نفوذ یا ماتریس تاثیر مولکول^۲ (MIM) با رابطه‌ی (۴-۴) بیان می‌شود:

$$H = M.(M^T.M)^{-1}.M^T \quad \text{معادله‌ی ۴-۴}$$

که در آن M ماتریس مختصات اتمی، T به معنای ماتریس معکوس و H ماتریس قدرت نفوذ است. H یک ماتریس متقارن $A \times A$ بوده که A تعداد اتم‌هاست.

اولین سری از این توصیف‌کننده‌های مولکولی، H- GETAWAY نامیده می‌شوند که تنها توسط اطلاعات به دست آمده توسط ماتریس نفوذ مولکولی محاسبه می‌شوند در حالی که سری مربوط به توصیف‌کننده‌های R- GETAWAY، ترکیبی از اطلاعات ماتریس تأثیر- فاصله است که به صورت زیر تعریف می‌شود:

$$[R]_{ij} = \left[\frac{\sqrt{h_{ii} h_{jj}}}{r_{ij}} \right]_{ij} \quad \text{معادله‌ی ۵-۴}$$

که h_{ii} و h_{jj} لوریج (قدرت نفوذ) دو اتم i و j و r_{ij} فاصله آن دو اتم است.

از این گروه، توصیف‌کننده‌های R1m، R5p و HATS0m وارد مدل برتر شده‌اند. اندیس‌های m و p نشان می‌دهند که این توصیف‌کننده‌ها به ترتیب بر اساس جرم اتمی و قطبش‌پذیری اتمی محاسبه

^۱Leverage matrix

^۲Molecular Influence Matrix

شده‌اند. توصیف‌کننده‌های R5p و HATS0m دارای اثر متوسط منفی و توصیف‌کننده‌ی R1m دارای اثر متوسط مثبت می‌باشد. همانطور که قبلاً گفته شد اثر متوسط منفی و اثر متوسط مثبت به ترتیب نشان‌دهنده‌ی ارتباط عکس و مستقیم با فعالیت بیولوژیکی می‌باشد.

ت) توصیف‌کننده‌های Galvez topological – Charge indices

این توصیف‌کننده‌ها جهت ارزیابی و بررسی انتقال بار بین جفت اتم‌ها در یک مولکول و بنابراین انتقال بار در تمام مولکول پیشنهاد شدند. برای توضیح این توصیف‌کننده‌ها ابتدا ماتریس M را بصورت زیر تعریف می‌کنیم:

$$M = A \times D^{-2} \quad \text{معادله‌ی ۴-۶}$$

که در آن A ماتریس مجاورت^۱ و D^{-2} ماتریس عکس مربع فاصله^۲ می‌باشد. توجه کنید که اعضای قطری ماتریس فاصله یکسان باقی می‌مانند. M ماتریس گالوز^۳ می‌باشد و یک ماتریس مربعی نامتقارن $A \times A$ می‌باشد که در آن A تعداد اتم‌های مولکول می‌باشد. M باعث یک ماتریس ترم بار^۴ نامتقارن، CT می‌شود که به صورت زیر تعریف می‌شود:

$$CT = \partial_i \quad \text{if } i=j \quad \text{معادله‌ی ۴-۷}$$

$$CT = m_{ij} - m_{ji} \quad \text{if } i \neq j$$

$$CT_{ij} = D(M) + [M - M^T] \quad \text{معادله‌ی ۴-۸}$$

که در آن m_{ij} عناصر M و ∂_i درجه‌ی بردار اتم i می‌باشد. ورودی‌های قطری CT ظرفیت توپولوژیکی اتم‌ها و ورودی‌های قطری CT_{ij} مقدار بار منتقل شده از اتم j به اتم i می‌باشد.

برای هر فاصله‌ی طولی k می‌توان یک اندیس بار توپولوژیکی، G_k ، به صورت زیر تعریف نمود:

^۱Adjacency

^۲Reciprocal square distance

^۳Galvez matrix

^۴Charge term matrix

$$G_k = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A |CT_{ij}| \cdot \delta(k; d_{ij}) \quad \text{معادله ی ۹-۴}$$

که در آن $\delta(k; d_{ij})$ دلتای کرونکر^۱ می باشد.

$$\begin{aligned} \delta(k; d_{ij}) &= 1 \quad \text{if } d_{ij} = k \\ &= 0 \quad \text{if } d_{ij} \neq k \end{aligned} \quad \text{معادله ی ۱۰-۴}$$

که در آن d_{ij} عناصر ماتریس فاصله هستند. بنابراین G_k بار کلی انتقالی بین اتم ها در یک فاصله ی توپولوژیکی k می باشد. حد اثر تعداد ترم های G_k برابر است با قطر توپولوژیکی D .

دو توصیف کننده ی JGI4 (با اثر متوسط منفی) و JGI6 (با اثر متوسط مثبت) از این گروه توصیف کننده ها وارد مدل برتر شده اند.

ث) توصیف کننده های WHIM

این شاخص از مختصات کارتزین ساختار سه بعدی مولکول، با استفاده از کنفورمری با حداقل انرژی محاسبه می شود و شامل اطلاعاتی درباره ی اندازه، شکل، تقارن و توزیع اتمی ساختار سه بعدی مولکول می باشد. این توصیف کننده از رابطه (۶-۸) به دست می آید:

$$S_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_j) (q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad \text{معادله ی ۱۱-۴}$$

که S_{jk} کوواریانس وزن دار شده بین کئوردینه j ام و k ام، A تعداد اتم ها، w_i وزن i امین اتم، q_{ij} و q_{ik} به ترتیب j امین و k امین کئوردینه های اتم i ام، و $\bar{q}$ مقدار میانگین متناظر است. از این توصیف کننده ها، E1s با اثر متوسط منفی وارد مدل برتر شد که با حالت های الکترو توپولوژیکی اتمی وزن دار شده است.

^۱Kronecker

۴-۳- بررسی میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی

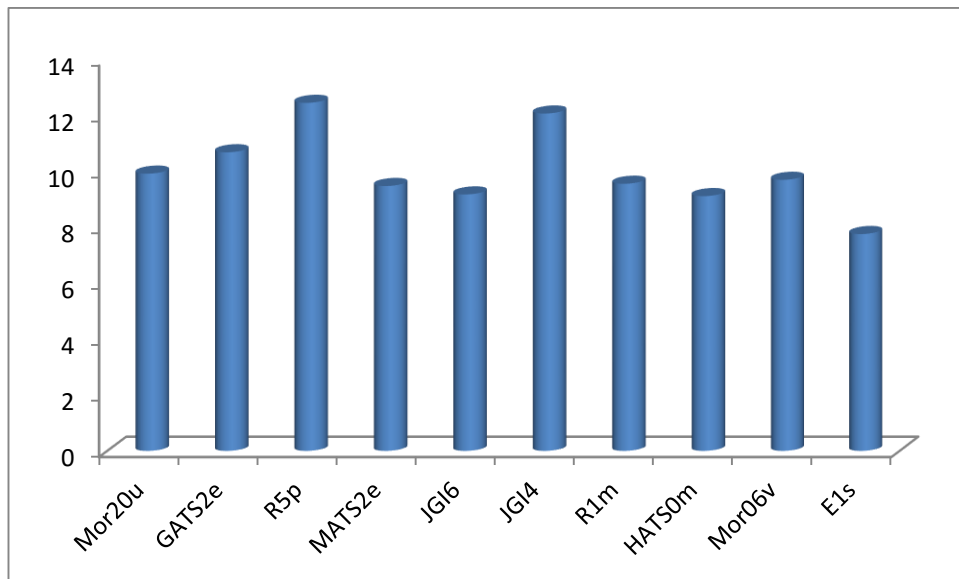
به منظور بررسی بیشتر اثر توصیف‌کننده‌های به کار گرفته شده در مدل شبکه‌ی عصبی مصنوعی، درصد مشارکت هر یک از توصیف‌کننده‌ها به صورت زیر محاسبه شد:

۱. توصیف‌کننده‌ی مورد نظر از سری توصیف‌کننده‌های ارائه شده به شبکه بهینه حذف شد.
۲. با استفاده از توصیف‌کننده‌های باقیمانده مقادیر پاسخ برای داده‌های سری ارزیابی محاسبه شد.
۳. با استفاده از مقادیر پاسخ محاسبه شده و مقادیر پاسخ تجربی، میانگین خطای مطلق در غیاب توصیف‌کننده‌ی مورد نظر محاسبه شد.
۴. مراحل ۱ تا ۳ برای هریک از توصیف‌کننده‌ها تکرار شد.
۵. در پایان درصد مشارکت هر یک از توصیف‌کننده‌ها به کمک رابطه‌ی (۴-۷) بدست آمد:

$$c_i = \frac{mae_i}{\sum_{i=1}^N mae_i} \times 100 \quad \text{معادله‌ی ۴-۱۲}$$

در رابطه‌ی بالا، c_i درصد مشارکت توصیف‌کننده‌ی i ام، N تعداد توصیف‌کننده‌های بکار رفته در مدل شبکه‌ی عصبی و mae_i میانگین خطای مطلق حاصل از سری ارزیابی در غیاب توصیف‌کننده‌ی i ام می‌باشد.

میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی ترکیبات CDK2 در شکل (۴-۱۱) نشان داده شده‌است. همانطور که مشاهده می‌شود، توصیف‌کننده‌های R5p و JGI4 بیشترین اثر مشارکت را در مدل ارائه شده دارند.



شکل ۴-۹- نمودار درصد مشارکت توصیف‌کننده‌های بکار گرفته شده در مدل شبکه عصبی برای ترکیبات CDK2

۴-۴- نتیجه‌گیری نهایی

هدف اصلی از کار حاضر، توسعه روشهای خطی و غیرخطی جهت مطالعه ارتباط کمی بین ساختار و فعالیت بیولوژیک مشتقات گوانین می‌باشد. نتایج نشان می‌دهند که مدل‌های خطی و غیرخطی توسعه یافته می‌توانند به عنوان روش‌های موفق برای مدل‌سازی و پیشگویی فعالیت بیولوژیک ترکیبات مورد مطالعه باشند و این مدل‌ها می‌توانند به محققان در پیش‌بینی فعالیت‌های بیولوژیک کمک کنند.

فصل پنجم

مدل سازی QSPR شاخصهای بازداري خطي کروماتوگرافي مربوط به
مونومتيل آلکانهای C4-C30 بر روی فاز ثابت متيل سليکون OV-1 با استفاده
از روش رگرسيون خطي و شبکه عصبي مصنوعي

۵-۱- معرفی مونومتیل آلکانها

مونومتیل آلکانها در خیلی از پدیده‌های طبیعی و مواد آلی سنتزی حضور دارند. این ترکیبات جزو اجزای تشکیل دهنده‌ی مهم مواد خام نفتی و محصولات تغییر یافته شان می‌باشند [۳۳]. این مواد همچنین در کولونیهای سیانوباکتریای موجود در روغن خام و رسوبات بعنوان یک منبع احتمالی از آلکانهای با شاخه‌های متیل یافت می‌شوند [۳۴]. حشرات انواع گوناگونی از آلکانهای با شاخه‌ی متیل مستقر در موقعیتهای خاصی (۲، ۳، ۷، ۹، ۱۱، ۱۳ و ۱۵) از زنجیر کربنی را تولید می‌کنند [۳۵]. مونومتیل آلکانها همچنین بعنوان ترکیبات تولیدی درون - بافتی موجود در بازدم نیز پیشنهاد شده اند [۳۶-۳۹] که در این پروژه، مدل‌سازی شاخص بازداری مونومتیل آلکان‌های سنتز شده در آزمایشگاه که احتمال دارد در بازدم وجود داشته باشند انجام شده است [۴۰].

آنالیز ترکیبات در بازدم می‌تواند در فهم فرآیندهای زیست - شیمی در افراد بیمار و سالم بسیار مفید باشد [۴۱ و ۴۲]. به غیر از اکسیژن و دی اکسید کربن، بازدم شامل تعدادی از ترکیبات مانند نیتریک اکسید (NO) [۴۳-۴۴]، مونوکسید کربن (CO) [۴۵ و ۴۶]، استون [۴۷]، ایزوپرن [۴۸-۵۰]، متانول [۵۱-۵۳]، اتان [۵۴-۵۷]، پنتان [۵۸ و ۵۹]، مرکاپتانها [۶۰]، دی‌متیل‌سولفید [۶۱ و ۶۲]، سولفید کربونیل [۶۳] و دیگر ترکیبات می‌باشد.

تکنیکهای تجزیه‌ای متفاوتی برای آنالیز نمونه‌های بازدم به کار رفته است که از آن جمله کروماتوگرافی گازی با آشکارساز اسپکترومتری جرمی (GC-MSS) می‌باشد. آنالیز بازدم بعنوان یک ابزار تشخیصی نیاز به تشخیص کمی در حد غلظتهای ppt^۱ تا ppb^۲ دارد. روشهای اساسی استفاده شده برای آنالیز مونومتیل آلکانها شامل کروماتوگرافی گازی مؤئینه با متیل سیلیکون به عنوان فاز ثابت در ستونهای با برنامه ریزی دمایی می‌باشد [۶۴].

^۱Part per thousand^۲Part per billion

به طور کلی بازداری در کروماتوگرافی گازی یک فرآیند بسیار پیچیده‌ای است که شامل برهمکنش چندین نیروی بین مولکولی می‌باشد. تمام این نیروها منجر به توزیع گونه بین فازهای گاز و مایع می‌شود. مدل سازی و پیش‌بینی خصوصیات فیزیکوشیمیایی ترکیبات آلی بسیار مهم می‌باشد. با مطالعات ارتباط کمی ساختار - خصوصیت (QSPR)، پیش‌بینی خصوصیات ترکیبات سنتز نشده یا پیش‌بینی خصوصیات اندازه‌گیری نشده‌ی ترکیبات معلوم امکان‌پذیر می‌باشد.

۵-۲- نرم افزارهای استفاده شده

لازم به ذکر است که نرم افزارهای استفاده شده مشابه نرم افزارهای بخش (۳-۲) می‌باشد.

۵-۳- انتخاب سری داده‌ها

تمام مونومتیل آلکانها تا C_{30} از مخلوط نرمال آلکانهای C_8-C_{29} توسط واکنش ورود متیلن تهیه شدند [۶۵] که این فرآیند با تهیه‌ی متیل آلکانهای C_4-C_8 کامل شد. مخلوط مونومتیل آلکان حاصله توسط کروماتوگرافی گازی موئینه در گستره‌ی دمایی $30^\circ C - 310^\circ C$ جداسازی شده و توسط آشکارساز جرمی تشخیص داده شدند [۶۴].

۵-۴- ویژگیهای سیستم اندازه‌گیری اندیس بازداری سری داده‌ها

اندازه‌گیرهای GC-MS جهت اندازه‌گیری اندیس بازداری مونومتیل آلکانها، روی دستگاه کروماتوگراف گازی ۶۸۹۰ سری N با یک آشکارساز انتخاب جرم ۵۹۷۳ اجرا شد. مقدار ۱ میکرولیتر از نمونه در مد تزریق شکاف با نسبت شکاف ۱:۱۰۰ در دمای $320^\circ C$ تزریق شد. مخلوط مونومتیل آلکانها با استفاده از ستون موئینه $100m \times 0.32mm I.D.$ که با یک فیلم با ضخامت ۰/۲۵ میکرون از متیل سیلیکون OV-1 به عنوان فاز ثابت پوشیده شده بود جداسازی شدند. دمای ستون در ابتدا ۳۰ درجه سلسیوس قرار داده شد، سپس دما تا ۳۱۰ درجه سلسیوس با سرعت $5^\circ C \min^{-1}$ افزایش یافت. دمای ستون در $310^\circ C$ به مدت ۵ دقیقه ثابت نگه داشته شد. هلیوم به عنوان گاز حامل با

سرعت خطی 26 cm.s^{-1} استفاده شد [۶۴]. در تحقیق حاضر، نتایج حاصل از این اندازه‌گیری به عنوان سری داده‌ها جهت مدل‌سازی QSPR به کار گرفته شد.

مشکلات موجود در اندازه‌گیری مونومتیل آلکانها با روشهایی مانند کروماتوگرافی گازی از جمله وجود عدم قطعیت در تهیه‌ی نمونه، بالا بودن هزینه‌ی استانداردها و بسیاری مشکلات دیگر از این قبیل، همچنین اهمیت این ترکیبات در موارد پزشکی، باعث شد تا در این پروژه با استفاده از روشهای کمومتریکس، مقادیر اندیس بازداري با استفاده از مقادیر اندازه‌گیری شده این پارامتر برای مونومتیل-آلکانهای $C_4 - C_{30}$ مدل‌سازی شود تا بتوان با مدل طراحی شده، اندیس بازداري ترکیبات مشابه را پیش‌بینی نمود. بدین منظور ابتدا مقادیر توصیف‌کننده‌ها محاسبه و سپس با استفاده از روشهای آنالیز رگرسیون خطی (MLR)، ارتباط خطی بین توصیف‌کننده‌ها و اندیس بازداري برقرار گردید. سپس با استفاده از شبکه عصبی مصنوعی (ANN) یک مدل غیرخطی برای پیش‌بینی اندیس بازداري ارائه شد.

۵-۵- محاسبه‌ی توصیف‌کننده‌ها

مقادیر اندیس بازداري ترکیبات به یکسری از خصوصیات هندسی، الکترونیکی و ساختاری‌شان ارتباط دارد. مقادیر این خصوصیات می‌تواند به صورت کمی توسط یکسری مقادیر عددی تحت عنوان توصیف‌کننده‌های مولکولی بیان شود. این توصیف‌کننده‌ها باید برای یافتن بهترین مدل QSPR مربوط به اندیس بازداري استفاده شوند. ساختارهای دو بعدی مولکولها با استفاده از نرم افزار HyperChem 8 رسم شدند [۲۴].

ساختارهای مولکولی با استفاده از الگوریتم گرادیان مزدوج بهینه شد که این کار تا جایی طول کشید که گرادیان RMS به $0.001 \text{ kcal.mol}^{-1}$ برسد. ساختارهای هندسی حاصله به نرم افزار Dragon منتقل و توصیف‌کننده‌های مرتبط تولید شد.

۵-۶- انتخاب توصیف کننده‌های مناسب

برای ساختن مدلی که بیانگر ارتباط بین ساختار ترکیبات مورد بررسی با اندیس بازداری آنها باشد از روش رگرسیون خطی چندگانه استفاده شد. توصیف کننده‌های محاسبه شده توسط نرم افزار Dragon، به عنوان متغیرهای مستقل و اندیس‌های بازداری به عنوان متغیر وابسته وارد نرم افزار SPSS گردید. از آنجایی که تعداد زیاد توصیف کننده‌ها باعث پیچیدگی محاسبات شده و همچنین احتمال وجود برهم کنش بین توصیف کننده‌ها افزایش می‌یابد، پس قبل از مدل سازی، تعدادی از توصیف کننده‌ها به خاطر داشتن تعداد زیاد مقادیر صفر یا مقادیر ثابت دیگر حذف شدند، چرا که حاوی اطلاعات کافی در مورد ساختار ترکیبات نبودند. همچنین توصیف کننده‌هایی که با یکدیگر همبستگی بیش از ۰/۹ داشتند، توصیف کننده‌ای که همبستگی کمتری با متغیر وابسته داشت حذف شد. در نهایت جهت به دست آوردن توصیف کننده‌هایی که بیشترین ارتباط را با اندیس‌های بازداری داشتند، روش رگرسیون مرحله‌ای بر روی تمام توصیف کننده‌های باقی مانده از مراحل قبل انجام گردید و چندین مدل مختلف بدست آمد. با در نظر گرفتن پارامترهای آماری از قبیل R^2 ، SE^۱، F و مجموع مربعات باقی مانده‌ها (SS_{Res})^۲ (رابطه‌ی ۲-۲۱)، مدل شماره‌ی ۷ انتخاب گردید. آماره‌های ۷ مدل اول حاصل از روش رگرسیون مرحله‌ای در جدول (۵-۱) آورده شده است.

^۱Standard Error^۲Sum of Squares of Residuals

جدول ۵-۱- هفت مدل اول حاصل از رگرسیون مرحله‌ای به همراه پارامترهای آماری مربوطه

شماره مدل	توصیف کننده	R^2	F	SE	SS_{Res}
۱	MW	۱/۰۰	۷۰۳۳۶۶/۴۴	۱۰/۸۹	۲۳۰۰۱/۳
۲	MW, Mor06u	۱/۰۰	۸۱۸۳۹۸/۴۳	۷/۱۴	۹۸۳۴/۷
۳	MW, Mor06u, GGI4	۱/۰۰	۷۴۷۸۰۲/۳۵	۶/۱	۷۱۳۸/۵
۴	MW, Mor06u, GGI4, PW5	۱/۰۰	۹۷۸۷۴۴/۹۱	۴/۶۲	۴۰۶۹/۴
۵	MW, Mor06u, GGI4, PW5, MATS4v	۱/۰۰	۱۰۰۶۱۳۱/۱	۴/۰۷	۳۱۴۹/۸
۶	MW, Mor06u, GGI4, PW5, MATS4v, GGI3	۱/۰۰	۱۲۴۵۳۶۶/۶	۳/۳۴	۲۱۰۹/۹
۷	MW, Mor06u, GGI4, PW5, MATS4v, GGI3, GGI2	۱/۰۰	۱۳۸۴۹۳۲/۴	۲/۹۳	۱۶۱۷/۶

این ۷ توصیف کننده به همراه طبقه آنها در جدول (۵-۲) ارائه شده است.

جدول ۵-۲ - توصیف کننده‌های انتخاب شده توسط رگرسیون خطی مرحله‌ای

No	Symbol	Class	Meaning
۱	MW	Constitutional	Molecular Weight
۲	Mor06u	3D –MORSE-signal	3D –MoRSE-signal 06/unweighted
۳	GGI4	Galvez topol	Topological charge index of order 4
۴	PW5	Topological	Path/Walk 5-Randic shape index
۵	MATS4v	2D-autocorrelations	Moran autocorrelation-lag 4/weighted by atomic van der waals volumes.
۶	GGI3	Galvez topol	Topological charge index of order 3
۷	GGI2	Galvez topol	Topological charge index of order 2

همچنین ماتریس همبستگی بین این توصیف کننده‌ها در جدول (۵-۳) ارائه شده که این ماتریس

عدم همبستگی بین توصیف کننده‌ها را نشان می‌دهد.

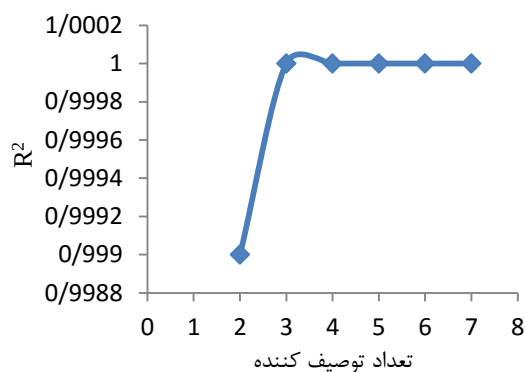
جدول ۵-۳- ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون خطی مرحله‌ای

	MW	Mor06u	GGI4	PW5	MATS4v	GGI3	GGI2
MW	۱						
Mor06u	۰/۲۷۸	۱					
GGI4	۰/۴۸۳	-۰/۲۸۲	۱				
PW5	۰/۵۹۶	۰/۰۸۸	۰/۸۳۲	۱			
MATS4v	-۰/۰۲۸	۰/۰۱۸	-۰/۰۶۱	-۰/۰۸۹	۱		
GGI3	۰/۴۰۳	-۰/۲۶۷	۰/۸۰۲	۰/۶۹۶	-۰/۱۲۶	۱	
GGI2	۰/۲۸۸	-۰/۲۰۹	-۰/۵۵۰	۰/۴۷۱	-۰/۶۵۴	۰/۶۹۹	۱

برای ساختن مدل خطی مناسب که بیانگر ارتباط ساختاری ترکیبات مورد مطالعه با اندیس بازداري آنها باشد، مدل‌های مختلفی (۷ مدل اول) که توسط رگرسیون خطی چندگانه مرحله‌ای به دست آمد، مورد بررسی قرار گرفتند. بدین منظور، داده‌ها به طور تصادفی به سه دسته‌ی آموزش، تست و ارزیابی با نسبت‌های به ترتیب ۶۰٪، ۲۰٪ و ۲۰٪ تقسیم شدند. سپس برای بررسی قدرت پیش‌بینی مدل‌های مختلف، ابتدا به کمک دو توصیف‌کننده‌ی اول و سری آموزش و با استفاده از روش Enter ضرائب توصیف‌کننده‌های وارد شده در مدل استخراج گردید و با استفاده از این ضرائب مقدار اندیس بازداري برای داده‌های سری ارزیابی پیش‌بینی گردید. این فرایند برای تمام هفت توصیف‌کننده تکرار شد. سپس ضریب تعیین و میانگین مربعات خطای سری ارزیابی تعیین گردید که این نتایج در جدول (۴-۵) آورده شده است. تاثیر تعداد توصیف‌کننده (شماره‌ی مدل) بر مقدار R^2 در شکل (۵-۱) نشان داده شده است.

جدول ۵-۴- مقادیر R^2 و MSE مربوط به سری ارزیابی

شماره مدل	MSE	R^2
۲	۳۶/۰۷	۰/۹۹۹
۳	۱۸/۳۹	۱
۴	۲۱/۹۷	۱
۵	۱۸/۴۵	۱
۶	۲۳۹/۳	۱
۷	۶۷/۵۸	۱

شکل ۵-۱- تغییرات R^2 بر حسب تعداد توصیف کننده برای سری ارزیابی

با توجه به مقادیر MSE و مجذور ضریب همبستگی و با توجه به اینکه هر چه تعداد توصیف کننده‌های وارد شده در مدل کمتر باشند، مدل از قدرت پیش‌بینی بالاتری برخوردار است، مدل شماره‌ی ۳ به عنوان مدل برتر انتخاب گردید که معادله‌ی آن به صورت زیر می‌باشد:

$$RI = -24.053 + 7.068 \times MW + 4.169 \times Mor06u - 92.26 \times GGI4 \quad \text{معادله‌ی ۵-۱}$$

جدول (۵-۵) نام توصیف کننده‌ها و اثر متوسط آنها را برای مدل منتخب نشان می‌دهد.

جدول ۵-۵- نام و معنای توصیف کننده به همراه اثر متوسط

توصیف کننده	ضریب	اثر متوسط
MW	۷/۰۶۸	۱/۰۲۱۲
Mor06u	۴/۱۶۹	-۰/۰۰۰۱۷۸
GGI4	-۹۲/۲۶	-۰/۰۰۹۲

همچنین جدول (۵-۶) مقادیر عددی توصیف کننده‌ها را برای ترکیبات سری داده‌ها نشان می‌دهد.

جدول ۵-۶- مقادیر عددی توصیف‌کننده‌ها و مقادیر تجربی اندیس بازداري برای ترکیبات مورد بررسی

شماره	نام ترکیب	MW	Mor06u	GGI4	PW5	MATS4v	GGI3	RI
۱	2-Methylpropane	۵۸	-۰/۸۷۹	۰/۰	۰/۰	۰/۴	۰/۰	۳۵۵/۷۷
۲	2-Methylbutane	۷۲	-۱/۲۵۶	۰/۰	۰/۰	-۰/۰۵۶	۰/۰	۴۶۶/۲۳
۳	2-Methylpentane	۸۶	-۱/۳۸	۰/۰	۰/۰	-۰/۲۲۸	۰/۲۵	۵۶۱/۳۱
۴	3-Methylpentane	۸۶	-۱/۲۴۲	۰/۰	۰/۰	۰/۰۶۱	۰/۱۲۵	۵۷۸/۰۵
۵	2-Methylhexane	۱۰۰	-۱/۲۵۴	۰/۱۶	۰/۰۴۴	-۰/۰۹۹	۰/۲۵	۶۶۲/۴۸
۶	3-Methylhexane	۱۰۰	-۱/۳۱۳	۰/۰۸	۰/۰۲۲	۰/۰۵۵	۰/۲۵	۶۷۲/۱۹
۷	2-Methylheptane	۱۱۴	-۱/۰۶۳	۰/۱۶	۰/۰۴۳	-۰/۰۰۸	۰/۲۵	۷۶۴/۳۲
۸	4-Methylheptane	۱۱۴	-۱/۳۵۷	۰/۰۸	۰/۰۳۲	-۰/۰۸۳	۰/۳۷۵	۷۶۵/۸۸
۹	3-Methylheptane	۱۱۴	-۱/۱۲۳	۰/۱۶	۰/۰۴۷	-۰/۰۰۶	۰/۲۵	۷۷۲/۱۷
۱۰	4-Methyloctane	۱۲۸	-۱/۱۳۱	۰/۱۶	۰/۰۵۲	۰/۰۴۹	۰/۳۷۵	۸۶۴/۰۶
۱۱	2-Methyloctane	۱۲۸	-۰/۷۳۳	۰/۱۶	۰/۰۴۵	-۰/۰۷۱	۰/۲۵	۸۶۵/۰۰
۱۲	3-Methyloctane	۱۲۸	-۰/۸۷۳	۰/۱۶	۰/۰۵۴	-۰/۰۰۵	۰/۲۵	۸۷۱/۸۹
۱۳	5-Methylnonane	۱۴۲	-۰/۸۸۱	۰/۲۴	۰/۰۶۶	-۰/۰۰۵	۰/۳۷۵	۹۶۱/۰۹
۱۴	4-Methylnonane	۱۴۲	-۰/۹۰۹	۰/۱۶	۰/۰۴۸	۰/۰۴۵	۰/۳۷۵	۹۶۲/۸۳
۱۵	2-Methylnonane	۱۴۲	-۰/۳۴۹	۰/۱۶	۰/۰۴۶	-۰/۰۶۳	۰/۲۵	۹۶۵/۳۹
۱۶	3-Methylnonane	۱۴۲	-۰/۵۴۴	۰/۱۶	۰/۰۴۵	-۰/۰۰۴	۰/۳۷۵	۹۷۲/۰۶
۱۷	5-Methyldecane	۱۵۶	-۰/۶۷۲	۰/۲۴	۰/۰۶۲	-۰/۰۰۴	۰/۳۷۵	۱۰۵۸/۹۴
۱۸	4-Methyldecane	۱۵۶	-۰/۵۶۹	۰/۱۶	۰/۰۴۸	۰/۰۴۱	۰/۲۵	۱۰۶۲/۰۴
۱۹	2-Methyldecane	۱۵۶	-۰/۱۲۲	۰/۱۶	۰/۰۴۸	-۰/۰۵۶	۰/۳۷۵	۱۰۶۵/۶۲
۲۰	3-Methyldecane	۱۵۶	-۰/۱۱۷	۰/۱۶	۰/۰۶۶	-۰/۰۰۳	۰/۳۷۵	۱۰۷۲/۰۶
۲۱	6-Methylundecane	۱۷۰	-۰/۴۸۲	۰/۲۴	۰/۰۵۸	-۰/۰۰۳	۰/۳۷۵	۱۱۵۶/۱۶
۲۲	5-Methylundecane	۱۷۰	-۰/۳۵۱	۰/۲۴	۰/۰۶۱	-۰/۰۰۳	۰/۲۵	۱۱۵۷/۳۶
۲۳	4-Methylundecane	۱۷۰	-۰/۳۶۲	۰/۲۴	۰/۰۶۱	۰/۰۳۸	۰/۲۵	۱۰۶۱/۲۱
۲۴	2-Methylundecane	۱۷۰	-۰/۶۰۱	۰/۱۶	۰/۰۴۹	-۰/۰۰۵	۰/۳۷۵	۱۱۶۵/۴۸
۲۵	3-Methylundecane	۱۷۰	۰/۳۲۵	۰/۱۶	۰/۰۴۷	-۰/۰۰۳	۰/۳۷۵	۱۱۷۲/۱۵
۲۶	6-Methyldodecane	۱۸۴	-۰/۲	۰/۲۴	۰/۰۵۷	-۰/۰۰۳	۰/۳۷۵	۱۲۵۴/۱۵
۲۷	5-Methyldodecane	۱۸۴	۰/۰	۰/۲۴	۰/۰۶	-۰/۰۰۳	۰/۲۵	۱۲۵۶/۱۸
۲۸	4-Methyldodecane	۱۸۴	۰/۱۹۱	۰/۱۶	۰/۰۵	۰/۰۳۶	۰/۲۵	۱۲۶۰/۷۵
۲۹	2-Methyldodecane	۱۸۴	۱/۱۵۱	۰/۱۶	۰/۰۵	-۰/۰۴۶	۰/۳۷۵	۱۲۶۵/۳۶
۳۰	3-Methyldodecane	۱۸۴	۰/۸۲۶	۰/۱۶	۰/۰۴۸	-۰/۰۰۳	۰/۳۷۵	۱۲۷۲/۱۲
۳۱	7-Methyltridecane	۱۹۸	-۰/۰۴	۰/۲۴	۰/۰۵۷	-۰/۰۰۳	۰/۳۷۵	۱۳۵۱/۹۴
۳۲	6-Methyltridecane	۱۹۸	۰/۰۶۵	۰/۲۴	۰/۰۵۷	-۰/۰۰۳	۰/۳۷۵	۱۳۵۲/۶
۳۳	5-Methyltridecane	۱۹۸	۰/۴۲۳	۰/۲۴	۰/۰۵۷	-۰/۰۰۳	۰/۲۵	۱۳۵۵/۴۳

ادامه‌ی جدول ۵-۶

شماره	نام ترکیب	MW	Mor06u	GGI4	PW5	MATS4v	GGI3	RI
۳۴	4-Methyltridecane	۱۹۸	۰/۶۵۳	۰/۱۶	۰/۰۵۱	۰/۰۰۳	۰/۲۵	۱۳۶۰/۳۵
۳۵	2-Methyltridecane	۱۹۸	۱/۶۸۴	۰/۱۶	۰/۰۵۱	-۰/۰۴۲	۰/۳۷۵	۱۳۶۵/۳۵
۳۶	3-Methyltridecane	۱۹۸	۱/۳۷۵	۰/۱۶	۰/۰۴۹	-۰/۰۰۲	۰/۳۷۵	۱۳۷۲/۳۳
۳۷	7-Methyltetradecane	۲۱۲	-۰/۶۷۱	۰/۲۴	۰/۰۵۶	-۰/۰۰۲	۰/۳۷۵	۱۴۵۰/۱۳
۳۸	6-Methyltetradecane	۲۱۲	-۰/۵۵۲	۰/۲۴	۰/۰۵۷	-۰/۰۰۲	۰/۳۷۵	۱۴۵۱/۶۳
۳۹	5-Methyltetradecane	۲۱۲	-۰/۲۲۵	۰/۲۴	۰/۰۵۸	-۰/۰۰۲	۰/۲۵	۱۴۵۴/۷۱
۴۰	4-Methyltetradecane	۲۱۲	۰/۰۶۹	۰/۱۶	۰/۰۵۱	۰/۰۳۱	۰/۲۵	۱۴۵۴/۷۱
۴۱	2-Methyltetradecane	۲۱۲	۱/۰۸۴	۰/۱۶	۰/۰۵۱	-۰/۰۳۹	۰/۳۷۵	۱۴۵۶/۳۷
۴۲	3-Methyltetradecane	۲۱۲	۰/۷۶	۰/۱۶	۰/۰۵۶	-۰/۰۰۲	۰/۳۷۵	۱۴۷۲/۵۱
۴۳	8-Methylpentadecane	۲۲۷	-۱/۰۰۹	۰/۲۴	۰/۰۵۶	-۰/۰۰۲	۰/۳۷۵	۱۵۴۸/۱۹
۴۴	7-Methylpentadecane	۲۲۷	-۱/۱۹۳	۰/۲۴	۰/۰۵۷	-۰/۰۰۲	۰/۳۷۵	۱۵۴۸/۸۵
۴۵	6-Methylpentadecane	۲۲۷	-۰/۷۶۳	۰/۲۴	۰/۰۵۷	-۰/۰۰۲	۰/۳۷۵	۱۵۵۰/۶۶
۴۶	5-Methylpentadecane	۲۲۷	-۰/۶۴۷	۰/۲۴	۰/۰۵۲	-۰/۰۰۲	۰/۲۵	۱۵۵۴/۲۴
۴۷	4-Methylpentadecane	۲۲۷	۰/۲۱۹	۰/۱۶	۰/۰۵۲	۰/۰۲۹	۰/۲۵	۱۵۵۹/۹۷
۴۸	2-Methylpentadecane	۲۲۷	۰/۷۲۵	۰/۱۶	۰/۰۵۲	-۰/۰۳۶	۰/۳۷۵	۱۵۶۵/۲۴
۴۹	3-Methylpentadecane	۲۲۷	۰/۴۶۹	۰/۱۶	۰/۰۵۱	-۰/۰۰۲	۰/۳۷۵	۱۵۷۲/۶۷
۵۰	8-Methylhexadecane	۲۴۱	-۱/۴۰۵	۰/۲۴	۰/۰۵۶	-۰/۰۰۲	۰/۳۷۵	۱۶۴۶/۹۶
۵۱	7-Methylhexadecane	۲۴۱	-۱/۵۵۳	۰/۲۴	۰/۰۵۷	-۰/۰۰۲	۰/۳۷۵	۱۶۴۷/۶۳
۵۲	6-Methylhexadecane	۲۴۱	-۱/۲۷۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۲	۰/۳۷۵	۱۶۵۰/۰۷
۵۳	5-Methylhexadecane	۲۴۱	-۱/۲۸۴	۰/۲۴	۰/۰۵۳	-۰/۰۰۲	۰/۲۵	۱۶۵۳/۹۷
۵۴	4-Methylhexadecane	۲۴۱	-۰/۶۶۴	۰/۱۶	۰/۰۵۳	۰/۰۲۸	۰/۲۵	۱۶۵۹/۹۱
۵۵	2-Methylhexadecane	۲۴۱	۱/۹۳	۰/۱۶	۰/۰۵۳	-۰/۰۳۴	۰/۳۷۵	۱۶۶۵/۳۵
۵۶	3-Methylhexadecane	۲۴۱	۰/۴۸۱	۰/۱۶	۰/۰۵۲	-۰/۰۰۱	۰/۳۷۵	۱۶۷۲/۹۹
۵۷	9-Methylheptadecane	۲۵۵	-۲/۱۰۱	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۷۴۵/۴
۵۸	8-Methylheptadecane	۲۵۵	-۲/۲۸۹	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۷۴۵/۵۵
۵۹	7-Methylheptadecane	۲۵۵	-۱/۶۰۴	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۷۴۶/۹۳
۶۰	6-Methylheptadecane	۲۵۵	-۰/۴۰۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۱۷۴۹/۷۱
۶۱	5-Methylheptadecane	۲۵۵	-۰/۲۴۴	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۲۵	۱۷۵۳/۵۶

ادامه‌ی جدول ۵-۶

شماره	نام ترکیب	MW	Mor06u	GGI4	PW5	MATS4v	GGI3	RI
۶۲	4-Methylheptadecane	۲۵۵	-۰/۳۵۰	۰/۱۶	۰/۰۵۳	-۰/۰۳۱	۰/۲۵	۱۷۵۲/۹۴
۶۳	2-Methylheptadecane	۲۵۵	-۰/۴۹	۰/۱۶	۰/۰۵۳	-۰/۰۳۱	۰/۳۷۵	۱۷۶۵/۲۹
۶۴	3-Methylheptadecane	۲۵۵	-۰/۸۸۴	۰/۱۶	۰/۰۵۲	-۰/۰۰۱	۰/۳۷۵	۱۷۷۳/۲۱
۶۵	9-Methyloctadecane	۲۶۹	-۱/۸۰۱	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۸۴۴/۰۳
۶۶	8-Methyloctadecane	۲۶۹	-۱/۶۲۲	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۸۴۴/۵۶
۶۷	7-Methyloctadecane	۲۶۹	-۱/۵۰۴	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۸۴۶/۵۱
۶۸	6-Methyloctadecane	۲۶۹	-۱/۲۷۵	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۱۸۴۹/۳۴
۶۹	5-Methyloctadecane	۲۶۹	-۰/۵۴۱	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۱۸۵۳/۶۱
۷۰	4-Methyloctadecane	۲۶۹	-۰/۴۴	۰/۱۶	۰/۰۵۴	۰/۰۲۵	۰/۲۵	۱۸۵۹/۹۷
۷۱	2-Methyloctadecane	۲۶۹	۱/۵۷۲	۰/۱۶	۰/۰۵۴	-۰/۰۳	۰/۳۷۵	۱۸۶۵/۲۸
۷۲	3-Methyloctadecane	۲۶۹	۰/۶۳۶	۰/۱۶	۰/۰۵۳	-۰/۰۰۱	۰/۳۷۵	۱۸۷۳/۴۴
۷۳	10-Methylnonadecane	۲۸۳	-۱/۸۶۲	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۹۴۲/۶۱
۷۴	9-Methylnonadecane	۲۸۳	-۱/۸۶۲	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۹۴۲/۶۸
۷۵	8-Methylnonadecane	۲۸۳	-۲/۱۹۷	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۱۹۴۳/۷۴
۷۶	7-Methylnonadecane	۲۸۳	-۲/۰۴۶	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۱۹۴۵/۷۹
۷۷	6-Methylnonadecane	۲۸۳	-۰/۷۶۳	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۱۹۴۸/۹۹
۷۸	5-Methylnonadecane	۲۸۳	-۱/۰۲۲	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۱۹۵۳/۴۵
۷۹	4-Methylnonadecane	۲۸۳	-۰/۹۹۶	۰/۱۶	۰/۰۵۴	۰/۰۲۴	۰/۲۵	۱۹۵۹/۹۴
۸۰	2-Methylnonadecane	۲۸۳	۰/۰۲۴	۰/۱۶	۰/۰۵۴	-۰/۰۲۸	۰/۳۷۵	۱۹۶۵/۲۳
۸۱	3-Methylnonadecane	۲۸۳	۰/۱۵	۰/۱۶	۰/۰۶۱	-۰/۰۰۱	۰/۳۷۵	۱۹۷۳/۸۴
۸۲	10-Methyleicosane	۲۹۷	-۱/۰۳۶	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۲۰۴۱/۶۵
۸۳	9-Methyleicosane	۲۹۷	-۰/۹۸۱	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۲۰۴۲/۱۷
۸۴	8-Methyleicosane	۲۹۷	-۱/۲	۰/۲۴	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۲۰۴۳/۲۸
۸۵	7-Methyleicosane	۲۹۷	-۱/۲۴۴	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۰۴۵/۴۵
۸۶	6-Methyleicosane	۲۹۷	-۱/۰۵۶	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۰۴۸/۷۹
۸۷	5-Methyleicosane	۲۹۷	-۰/۴۴۷	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۲۰۵۳/۳۹
۸۸	4-Methyleicosane	۲۹۷	-۰/۲۱۹	۰/۱۶	۰/۰۵۶	۰/۰۲۳	۰/۲۵	۲۰۶۰/۱۶
۸۹	2-Methyleicosane	۲۹۷	۰/۹۳۷	۰/۱۶	۰/۰۵۸	-۰/۰۲۶	۰/۳۷۵	۲۰۶۵/۳۴
۹۰	3-Methyleicosane	۲۹۷	۰/۶۵۹	۰/۱۶	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۰۷۴/۱۵
۹۱	11-Methylheneicosane	۳۱۱	-۰/۷۳۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۱۴۰/۳۷
۹۲	10-Methylheneicosane	۳۱۱	-۱/۹۰۹	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۱۴۰/۴۸
۹۳	9-Methylheneicosane	۳۱۱	-۰/۴۷۸	۰/۲۴	۰/۰۵۶	-۰/۰۰۱	۰/۳۷۵	۲۱۴۱/۵۵
۹۴	8-Methylheneicosane	۳۱۱	-۰/۶۴	۰/۲۴	۰/۰۵۶	-۰/۰۰۱	۰/۳۷۵	۲۱۴۲/۵۷
۹۵	7-Methylheneicosane	۳۱۱	۰/۸۸۷	۰/۲۴	۰/۰۵۶	-۰/۰۰۱	۰/۳۷۵	۲۱۴۴/۹۷

ادامه‌ی جدول ۵-۶

شماره	نام ترکیب	MW	Mor06u	GGI4	PW5	MATS4v	GGI3	RI
۹۶	6-Methylheneicosane	۳۱۱	۰/۹۰۴	۰/۲۴	۰/۰۵۶	-۰/۰۰۱	۰/۳۷۵	۲۱۴۸/۳۶
۹۷	5-Methylheneicosane	۳۱۱	-۰/۱۶۶	۰/۲۴	۰/۰۵۶	-۰/۰۰۱	۰/۲۵	۲۱۵۳/۲۴
۹۸	4-Methylheneicosane	۳۱۱	۲/۲۶۶	۰/۱۶	۰/۰۵۸	۰/۰۲۲	۰/۲۵	۲۱۶۲/۴۵
۹۹	2-Methylheneicosane	۳۱۱	۱/۵۰۵	۰/۱۶	۰/۰۵۸	-۰/۰۲۵	۰/۳۷۵	۲۱۶۵/۲۳
۱۰۰	3-Methylheneicosane	۳۱۱	۰/۱۱۲	۰/۱۶	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۲۱۰/۷۶
۱۰۱	11-Methyldocosane	۳۲۵	-۲/۷۳۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۲۳۹/۲۶
۱۰۲	10-Methyldocosane	۳۲۵	-۲/۱۳۸	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۲۳۹/۶۵
۱۰۳	9-Methyldocosane	۳۲۵	-۲/۵۹۲	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۲۴۰/۷۱
۱۰۴	8-Methyldocosane	۳۲۵	-۱/۸۸۷	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۲۴۲/۲۷
۱۰۵	7-Methyldocosane	۳۲۵	-۰/۰۹۶	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۲۴۴/۶۶
۱۰۶	6-Methyldocosane	۳۲۵	-۰/۶۸۹	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۲۴۸/۱۵
۱۰۷	5-Methyldocosane	۳۲۵	-۰/۶۷۴	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۲۲۵۰/۱۷
۱۰۸	4-Methyldocosane	۳۲۵	-۰/۰۱	۰/۱۶	۰/۰۵۵	۰/۰۲۱	۰/۲۵	۲۲۶۱/۳۲
۱۰۹	2-Methyldocosane	۳۲۵	-۰/۳۴۵	۰/۱۶	۰/۰۵۶	-۰/۰۲۴	۰/۳۷۵	۲۲۷۰/۹۸
۱۱۰	3-Methyldocosane	۳۲۵	۱/۶۳۱	۰/۱۶	۰/۰۵۵	-۰/۰۰۱	۰/۳۷۵	۲۲۷۴/۳۴
۱۱۱	12-Methyltricosane	۳۳۹	-۱/۸۷۵	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۳۲۰/۱۲
۱۱۲	11-Methyltricosane	۳۳۹	-۲/۸۱۷	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۳۳۸/۱۵
۱۱۳	10-Methyltricosane	۳۳۹	-۲/۱۵۶	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۳۳۸/۶۹
۱۱۴	9-Methyltricosane	۳۳۹	-۰/۷۹۴	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۳۴۰/۶۷
۱۱۵	8-Methyltricosane	۳۳۹	-۱/۰۴۴	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۳۴۱/۶۹
۱۱۶	7-Methyltricosane	۳۳۹	۰/۴۷۳	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۳۴۴/۲۵
۱۱۷	6-Methyltricosane	۳۳۹	۲/۰۵۸	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۳۷۵	۲۳۴۷/۹۲
۱۱۸	5-Methyltricosane	۳۳۹	۲/۴۴۸	۰/۲۴	۰/۰۵۶	-۰/۰۰۱	۰/۲۵	۲۳۵۲/۸۸
۱۱۹	4-Methyltricosane	۳۳۹	۳/۷۵۹	۰/۱۶	۰/۰۵۶	-۰/۰۰۱	۰/۲۵	۲۳۴۳/۵۴
۱۲۰	2-Methyltricosane	۳۳۹	۲/۹۳۸	۰/۱۶	۰/۰۵۵	-۰/۰۲۳	۰/۳۷۵	۲۳۸۹/۸۷
۱۲۱	3-Methyltricosane	۳۳۹	۱/۴۲۸	۰/۱۶	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۴۳۱/۶۵
۱۲۲	12-Methyltetracosane	۳۵۳	-۱/۵۲۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۴۳۷/۳۵
۱۲۳	11-Methyltetracosane	۳۵۳	-۱/۴۶۳	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۴۳۷/۶۱
۱۲۴	10-Methyltetracosane	۳۵۳	-۱/۲۳۷	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۴۳۸/۲۵
۱۲۵	9-Methyltetracosane	۳۵۳	-۰/۹۸۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۴۳۹/۷۴
۱۲۶	8-Methyltetracosane	۳۵۳	-۰/۱۰۳	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۴۴۱/۵۲
۱۲۷	7-Methyltetracosane	۳۵۳	۰/۱۴۱	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۴۳۰/۴۴
۱۲۸	6-Methyltetracosane	۳۵۳	۰/۱۴۳	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۴۴۷/۹۵
۱۲۹	5-Methyltetracosane	۳۵۳	۱/۵۹۹	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۲۴۵۲/۹۷

ادامه‌ی جدول ۵-۶

شماره	نام ترکیب	MW	Mor06u	GGI4	PW5	MATS4v	GGI3	RI
۱۳۰	4-Methyltetracosane	۳۵۳	۲/۴۲۸	۰/۱۶	۰/۰۵۶	۰/۰۱۹	۰/۲۵	۱۲۴۶۰
۱۳۱	2-Methyltetracosane	۳۵۳	۳/۵۵۵	۰/۱۶	۰/۰۵۶	-۰/۰۲۲	۰/۳۷۵	۲۴۷۰/۴۵
۱۳۲	3-Methyltetracosane	۳۵۳	۱/۲۰۸	۰/۱۶	۰/۰۵۵	-۰/۰۰۱	۰/۳۷۵	۲۴۷۴/۸۶
۱۳۳	13-Methylpentacosane	۳۶۷	-۱/۷۴۴	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۵۳۶/۴۷
۱۳۴	12-Methylpentacosane	۳۶۷	-۱/۶۱۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۵۳۶/۵۴
۱۳۵	11-Methylpentacosane	۳۶۷	-۱/۰۵۵	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۵۳۶/۹۸
۱۳۶	10-Methylpentacosane	۳۶۷	-۰/۱۷۳	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۵۳۷/۷۴
۱۳۷	9-Methylpentacosane	۳۶۷	۰/۰۸۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۵۳۹/۳۶
۱۳۸	8-Methylpentacosane	۳۶۷	۱/۵۷	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۵۴۱/۳۲
۱۳۹	7-Methylpentacosane	۳۶۷	۰/۶۶۲	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۵۴۳/۹۸
۱۴۰	6-Methylpentacosane	۳۶۷	۱/۸۹	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۵۴۷/۸۵
۱۴۱	5-Methylpentacosane	۳۶۷	۰/۸۵	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۲۵۵۳/۱۵
۱۴۲	4-Methylpentacosane	۳۶۷	۱/۸۸	۰/۱۶	۰/۰۵۶	۰/۰۱۹	۰/۲۵	۲۵۶۱/۱۳
۱۴۳	2-Methylpentacosane	۳۶۷	۱/۶۱۸	۰/۱۶	۰/۰۵۶	-۰/۰۲۱	۰/۳۷۵	۲۵۶۵/۲۹
۱۴۴	3-Methylpentacosane	۳۶۷	۱/۸۱۶	۰/۱۶	۰/۰۵۵	-۰/۰۰۱	۰/۳۷۵	۲۵۷۵/۴۵
۱۴۵	13-Methylhexacosane	۳۸۱	-۱/۶۹۱	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۶۵۵/۴۴
۱۴۶	12-Methylhexacosane	۳۸۱	-۲/۴۳۹	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۶۳۵/۸۷
۱۴۷	11-Methylhexacosane	۳۸۱	-۲/۱۱۴	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۶۳۶/۳۱
۱۴۸	10-Methylhexacosane	۳۸۱	-۱/۳۱۶	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۶۳۷/۳۵
۱۴۹	9-Methylhexacosane	۳۸۱	-۰/۸۷۸	۰/۲۴	۰/۰۵۸	-۰/۰۰۱	۰/۳۷۵	۲۶۴۰/۷۶
۱۵۰	8-Methylhexacosane	۳۸۱	-۰/۲۵۶	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۶۴۲/۵۴
۱۵۱	7-Methylhexacosane	۳۸۱	۰/۸۴۴	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۶۴۳/۸۴
۱۵۲	6-Methylhexacosane	۳۸۱	۰/۵۶۷	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۶۴۷/۹۱
۱۵۳	5-Methylhexacosane	۳۸۱	۰/۶۶۵	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۲۶۵۱/۴۸
۱۵۴	4-Methylhexacosane	۳۸۱	۱/۶۳۳	۰/۱۶	۰/۰۵۶	۰/۰۱۸	۰/۲۵	۲۶۶۰/۷۱
۱۵۵	2-Methylhexacosane	۳۸۱	۰/۹۱۲	۰/۱۶	۰/۰۵۶	-۰/۰۰۲	۰/۳۷۵	۲۶۷۲/۲۷
۱۵۶	3-Methylhexacosane	۳۸۱	۰/۶۶۵	۰/۱۶	۰/۰۵۶	-۰/۰۰۱	۰/۳۷۵	۲۶۷۵/۷۲
۱۵۷	14-Methylheptacosane	۳۹۵	-۲/۹۷۱	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۳۴/۹۳
۱۵۸	13-Methylheptacosane	۳۹۵	-۲/۶۲۴	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۳۵/۰
۱۵۹	12-Methylheptacosane	۳۹۵	-۳/۱۲۸	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۳۵/۴۵
۱۶۰	11-Methylheptacosane	۳۹۵	-۱/۵۴۶	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۳۶/۱۶
۱۶۱	10-Methylheptacosane	۳۹۵	-۰/۵۸۵	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۳۷/۲۱
۱۶۲	9-Methylheptacosane	۳۹۵	-۰/۷۷۲	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۳۹/۱۴
۱۶۳	8-Methylheptacosane	۳۹۵	-۰/۱۳۱	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۴۱/۲۹

ادامه‌ی جدول ۵-۶

شماره	نام ترکیب	MW	Mor06u	GGI4	PW5	MATS4v	GGI3	RI
۱۶۴	7-Methylheptacosane	۳۹۵	۰/۱۹۳	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۴۳/۸۷
۱۶۵	6-Methylheptacosane	۳۹۵	۲/۱۸۷	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۷۴۷/۸۲
۱۶۶	5-Methylheptacosane	۳۹۵	۱/۲۹۸	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۲۷۵۳/۲۲
۱۶۷	4-Methylheptacosane	۳۹۵	۳/۳۲۸	۰/۱۶	۰/۰۵۶	۰/۰۱۷	۰/۲۵	۲۷۶۰/۸۶
۱۶۸	2-Methylheptacosane	۳۹۵	۳/۰۱۷	۰/۱۶	۰/۰۵۷	-۰/۰۱۹	۰/۳۷۵	۲۷۶۵/۲۶
۱۶۹	3-Methylheptacosane	۳۹۵	۳/۱۷۸	۰/۱۶	۰/۰۵۶	-۰/۰۰۱	۰/۳۷۵	۲۷۷۶/۰۹
۱۷۰	14-Methyloctacosane	۴۰۹	-۰/۹۴	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۳۴/۴۲
۱۷۱	13-Methyloctacosane	۴۰۹	-۱/۵۵	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۳۴/۵۷
۱۷۲	12-Methyloctacosane	۴۰۹	-۱/۳۲	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۳۵/۱۴
۱۷۳	11-Methyloctacosane	۴۰۹	-۱/۱۸	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۳۵/۸۸
۱۷۴	10-Methyloctacosane	۴۰۹	-۰/۵۵	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۳۷/۱۷
۱۷۵	9-Methyloctacosane	۴۰۹	-۰/۳۳	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۳۹/۰۷
۱۷۶	8-Methyloctacosane	۴۰۹	-۰/۱۰	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۴۱/۱۹
۱۷۷	7-Methyloctacosane	۴۰۹	۰/۴۸۸	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۸۴۳/۹۶
۱۷۸	6-Methyloctacosane	۴۰۹	۱/۲۲۸	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۳۷۵	۲۸۴۸/۰۴
۱۷۹	5-Methyloctacosane	۴۰۹	۲/۱۸۶	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۲۵	۲۸۵۳/۴۰
۱۸۰	4-Methyloctacosane	۴۰۹	۲/۸۹۲	۰/۱۶	۰/۰۵۷	۰/۰۱۷	۰/۲۵	۲۸۶۳/۱۸
۱۸۱	2-Methyloctacosane	۴۰۹	۳/۱۳۹	۰/۱۶	۰/۰۵۷	-۰/۰۱۹	۰/۳۷۵	۲۸۶۵/۷۰
۱۸۲	3-Methyloctacosane	۴۰۹	۴/۷۶۴	۰/۱۶	۰/۰۵۶	-۰/۰۰۱	۰/۳۷۵	۲۸۷۶/۳۸
۱۸۳	15-Methylnonacosane	۴۲۳	۰/۱۵۷	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۳۳/۷۷
۱۸۴	14-Methylnonacosane	۴۲۳	-۰/۲۸	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۳۳/۸۲
۱۸۵	13-Methylnonacosane	۴۲۳	۰/۰۶۳	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۳۴/۲۶
۱۸۶	12-Methylnonacosane	۴۲۳	-۰/۲۶	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۳۴/۸۵
۱۸۷	11-Methylnonacosane	۴۲۳	۰/۲۰۳	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۳۵/۵۴
۱۸۸	10-Methylnonacosane	۴۲۳	۰/۲۷۲	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۳۷
۱۸۹	9-Methylnonacosane	۴۲۳	۰/۶۹	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۳۸/۹۰
۱۹۰	8-Methylnonacosane	۴۲۳	۱/۱۸۶	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۴۱/۱۱
۱۹۱	7-Methylnonacosane	۴۲۳	۱/۷۰۴	۰/۲۴	۰/۰۵۹	-۰/۰۰۱	۰/۳۷۵	۲۹۴۳/۹۳
۱۹۲	6-Methylnonacosane	۴۲۳	۲/۳۹۷	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۳۷۵	۲۹۴۸/۱۴
۱۹۳	5-Methylnonacosane	۴۲۳	۳/۱۲۹	۰/۲۴	۰/۰۶۱	-۰/۰۰۱	۰/۳۷۵	۲۹۵۳/۴۳
۱۹۴	4-Methylnonacosane	۴۲۳	۴/۵۱۱	۰/۱۶	۰/۰۵۷	-۰/۰۰۱	۰/۳۷۵	۲۹۶۱/۵۶
۱۹۵	2-Methylnonacosane	۴۲۳	۶/۴۳۸	۰/۱۶	۰/۰۵۷	۰/۰۱۶	۰/۲۵	۲۹۶۵/۷۲
۱۹۶	3-Methylnonacosane	۴۲۳	۴/۴۳۵	۰/۱۶	۰/۰۵۶	-۰/۰۱۸	۰/۲۵	۲۹۷۶/۴۳

۵-۶-۱- مدل سازی به روش شبکه عصبی

در این پروژه، شبکه‌ی عصبی مصنوعی برای مدل‌سازی غیرخطی جهت پیش‌بینی اندیس بازدارى داده‌ها مورد استفاده قرار گرفت. در این تحقیق، برنامه‌نویسی و اجرای یک شبکه‌ی عصبی پیش‌خور سه لایه با الگوی آموزش پس انتشار در محیط نرم‌افزار MATLAB انجام شد. در این روش تابع کارایی شبکه، میانگین مجموع مربعات خطا (MSE) می‌باشد. شبکه‌ی طراحی شده از سه لایه تشکیل شده است که لایه‌ی اول شامل توصیف‌کننده‌های مدل ۷ بوده که به عنوان ورودی به شبکه داده شد. تعداد گره لایه‌ی دوم که در حین آموزش شبکه بهینه شد و در نهایت تعداد گره لایه سوم یا لایه‌ی خروجی به طور پیش فرض شامل یک گره بود. مقدار تجربی اندیس بازدارى نیز به عنوان هدف وارد شبکه شد. به منظور افزایش تعمیم‌پذیری شبکه، سری داده‌ها به طور تصادفی به سه دسته‌ی آموزش (۶۰٪)، ارزیابی (۲۰٪) و تست (۲۰٪) تقسیم شدند. شبکه‌ی پیش‌خور طراحی شده مقادیر ابتدایی وزن‌ها و بایاس‌ها را به صورت خودکار تعیین می‌کند. مطابق تابع اجرای مربوطه مقدار خطای شبکه از تفاضل مقدار خروجی محاسبه شده و مقدار تجربی اندیس بازدارى محاسبه می‌گردد. مقادیر وزن‌ها طی آموزش شبکه آنقدر تغییر می‌کند تا مقدار خطای شبکه به کمترین مقدار ممکن برسد. برای بدست آوردن بهترین مدل و کمترین خطا، باید ابتدا مقادیر بهینه‌ی پارامترهای موثر بر شبکه تعیین گردد که این پارامترها عبارتند از:

- تعداد ورودی شبکه
- تعداد گره لایه‌ی پنهان
- نوع تابع انتقال
- نوع تابع آموزش
- تعداد دوره‌های آموزش
- پارامتر μ (مقدار اولیه برای μ)

۵-۶-۱-۱- انتخاب اولیه‌ی مقدار وزن‌ها و بایاس‌ها

قبل از آموزش یک شبکه‌ی پیش‌خور، مقادیر ابتدایی وزن‌ها و بایاس‌ها باید تعیین گردند. تابع newff به‌صورت خودکار به وزن‌ها مقدار می‌دهد ولی ممکن است با دستور init آنرا به‌صورت غیر خودکار نیز انجام داد.

۵-۶-۱-۲- بهینه سازی پارامترهای شبکه

برای بهینه‌سازی پارامترهای شبکه، شبکه‌هایی با تعداد ورودی‌های از ۲ تا ۷ توصیف‌کننده ایجاد شدند. هر شبکه با الگوریتم آموزشی لوبز-مارکوارت و تنظیم بایزین با تعداد گره متفاوت از ۲ تا ۱۰ گره در لایه‌ی پنهان آموزش داده شد. این در حالی بود که برای بدست آوردن بهترین تابع انتقال در لایه‌ی پنهان مدل‌های شبکه‌ی عصبی طراحی شده، از توابع لگاریتم سیگموئید و تانژانت هایپربولیک به‌عنوان تابع فعالیت استفاده گردید و تعداد دوره‌های آموزش نیز از ۲ تا ۲۰ تغییر داده شد. در روند بهینه‌سازی شبکه، به حداقل رساندن مقدار میانگین مربعات خطا (MSE) به‌عنوان معیار انتخاب شد. با توجه به اینکه خروجی‌های شبکه داده‌های کمی می‌باشند، لذا تابع انتقال خطی برای لایه‌ی خروجی برگزیده شد.

۵-۶-۱-۳- انتخاب تعداد متغیرهای ورودی شبکه، الگوریتم آموزشی، نوع تابع انتقال،

تعداد گره‌های لایه‌ی پنهان و تعداد دوره‌های آموزشی

در این پروژه بهینه‌سازی و انتخاب بهترین تعداد متغیرهای ورودی شبکه، الگوریتم آموزشی، نوع تابع تبدیل، تعداد نرون‌های لایه‌ی پنهان و تعداد دوره‌های آموزش به‌طور همزمان انجام گرفت (در این مرحله، مقدار پارامتر μ ، ۰/۰۰۱ در نظر گرفته شد). این کار موجب می‌شود بتوانیم نقش تمام پارامترها را به‌طور همزمان و یکجا بررسی نماییم. سپس در مقادیر بهینه‌ی این پارامترها، مقدار پارامتر μ از ۰/۰۰۰۵ تا ۰/۰۱ با گام‌های ۰/۰۰۰۵ تغییر داده شد تا با توجه به مقدار خطای شبکه، مقدار بهینه انتخاب گردد. بدین منظور به شبکه‌ای کاملاً تکرار پذیر نیازمندیم که بتواند به‌طور کاملاً

خودکار پارامترهای شبکه را تغییر داده و مقادیر MSE را که معیار انتخاب بهترین شبکه در این پروژه می باشد، برای هر شبکه‌ی خاص محاسبه نماید.

با توجه به نتایج حاصله برای تعداد متغیرهای ورودی شبکه، تعداد مختلف گره‌ها، دوره‌های آموزش و همچنین توابع متفاوت آموزش و تبدیل، بهترین شبکه‌های بدست آمده بر اساس کمترین مقدار MSE در جدول (۷-۵) خلاصه شده‌اند.

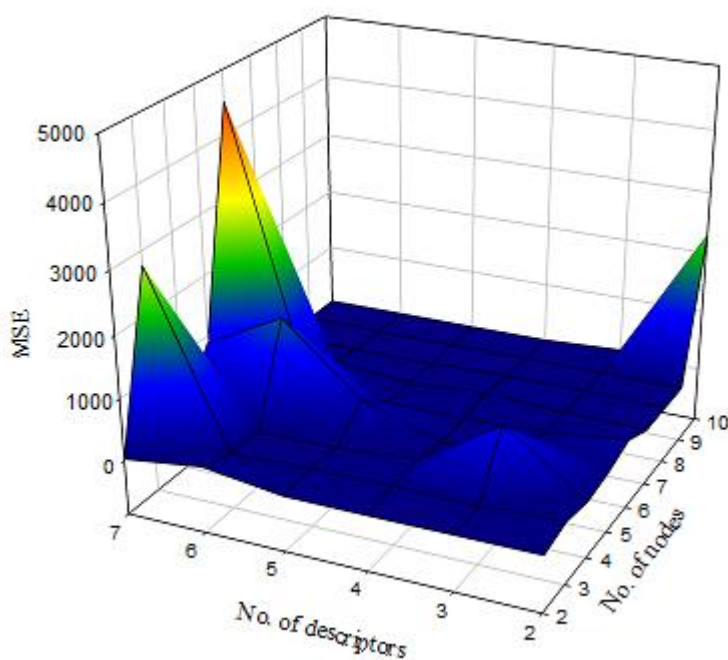
جدول ۷-۵- پارامترهای شبکه‌های بهینه بدست آمده بر اساس مقادیر میانگین مربعات خطا (MSE)

MSE	تعداد دوره‌های آموزش	تعداد گره‌ها	تابع تبدیل	تابع آموزش	تعداد توصیف‌کننده ها
۲/۲۶	۱۲	۴	لگاریتم سیگموئید	لونبرگ-مارکوارت	۶
۱۲/۲۹	۵	۱۰	لگاریتم سیگموئید	بایزین	۷
۳/۴۸	۱۷	۱۰	تانژانت هایپربولیک	لونبرگ-مارکوارت	۷
۱۰/۳۳	۲۰	۹	تانژانت هایپربولیک	بایزین	۷

با توجه به نتایج بدست آمده، الگوریتم آموزشی لونبرگ - مارکوارت نسبت به تنظیم بایزین دارای میانگین مربعات خطای کمتری است. همچنین تابع لگاریتم سیگموئیدی دارای خطای کمتری نسبت به تانژانت سیگموئیدی است. بنابراین تابع انتقال لگاریتم سیگموئیدی به عنوان تابع انتقال لایه پنهان استفاده شد. طبق نتایج به دست آمده، تعداد ۶ توصیف‌کننده، به عنوان توصیف‌کننده‌هایی که بیشترین ارتباط را با اندیس بازداري دارند، انتخاب شدند. نتایج محاسبات در جداول (۸-۵)، (۹-۵)، (۱۰-۵) و (۱۱-۵) و شکل‌های سه بعدی این جداول نیز در شکل‌های (۲-۵)، (۳-۵)، (۴-۵) و (۵-۵) ارائه شده است.

جدول ۵-۸- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزش لومبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

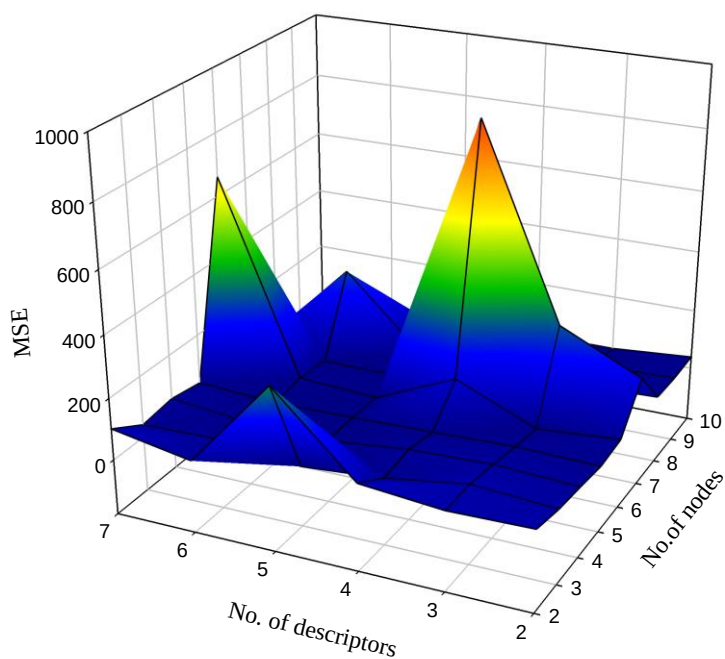
		تعداد ورودی‌های شبکه					
		۲	۳	۴	۵	۶	۷
تعداد گره‌های لایه پنهان	۲	۳۲/۷۹	۲۵/۹۳	۵۹۲/۳۶	۱۸/۹۰	۳۷۰/۹۹	۴۴/۶۰
	۳	۲۳۷/۳۷	۸۹/۶۴	۲۱۳/۷۵	۱۱/۲۷	۵/۱۵	۲۷۷۸/۹۷
	۴	۴۴/۹۵	۱۸۸۱/۲۰	۱۳/۹۱	۱۳/۷۰	۲/۲۶	۴۵۸/۹۰
	۵	۵۲/۳۰	۲۵۴/۳۷	۱۶/۲۹	۲۷۷/۲۶	۲۸۱۳/۵۵	۱۷۷۰/۹۲
	۶	۴۴۹/۵۵	۱۶۵/۳۷	۵۳/۷۹	۲۱/۰۶	۹۶/۴۷	۷۲۰/۳۳
	۷	۴۲/۶۷	۷۹/۵۰	۲۱۵/۰۴	۲۱/۳۲	۹/۸۲	۷/۵۸
	۸	۴۴/۶۷	۱۱۶/۲۳	۳۹۲/۹۰	۸/۵۶	۶/۷۰	۸/۴۰
	۹	۴۹۷/۸۲	۳۹/۸۷	۲۴/۶۴	۵۱۷/۸۴	۳/۱۰	۵/۶۶
	۱۰	۴۴۶۲/۲۷	۱۶۱/۰۰	۱۱/۳۵	۱۷/۱۹	۱۶/۲۳	۷/۰۹



شکل ۵-۲- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش لومبرگ - مارکوارت و تابع تبدیل لگاریتم سیگموئیدی

جدول ۵-۹- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزش تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

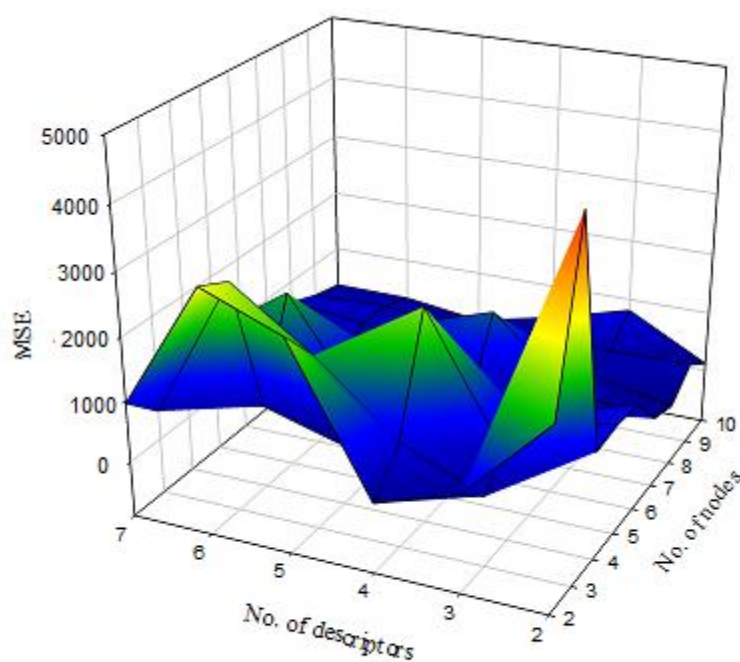
		تعداد ورودی های شبکه					
		۲	۳	۴	۵	۶	۷
تعداد گره‌های لایه پنهان	۲	۸۳/۱۷	۷۷/۸۹	۲۹۷/۴۳	۹۰۶/۸۰	۲۴/۱۹	۱۰۲/۰۳
	۳	۶۱/۰۴	۷۹/۱۲	۳۱/۴۶	۲۲/۶۱	۳۱/۴۶	۵۵/۰۶
	۴	۴۹/۸۰	۶۱/۷۰	۲۵/۸۲	۲۴/۹۱	۵۵/۴۲	۵۸/۳۷
	۵	۳۶/۱۷	۴۲/۰۳	۱۹/۱۲	۱۸/۷۲	۱۶/۴۷	۱۷۷۶/۴۰
	۶	۴۶/۴۰	۲۵/۲۱	۱۰۱/۹۱	۳۴۷/۷۳	۵۵/۴۸	۹۰۲/۵۹
	۷	۱۷۴/۷۶	۲۹۷/۴۳	۳۷/۸۶	۱۷/۱۸	۲۸/۷۹	۱۹/۲۶
	۸	۴۰/۸۰	۳۱/۴۶	۵۲/۵۱	۲۴/۶۳	۲۲/۰۲	۱۷/۰۳
	۹	۳۵/۶۹	۲۵/۸۲	۳۳/۴۷	۲۶/۸۴	۳۵/۵۸	۱۴/۷۸
	۱۰	۳۹/۴۷	۱۹/۱۲	۱۳۷/۴۶	۲۱/۵۰	۲۵/۱۵	۱۲/۲۹



شکل ۵-۳- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش بایزین و تابع تبدیل لگاریتم سیگموئید

جدول ۵-۱۰- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزش لوببرگ - مارکوارت و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

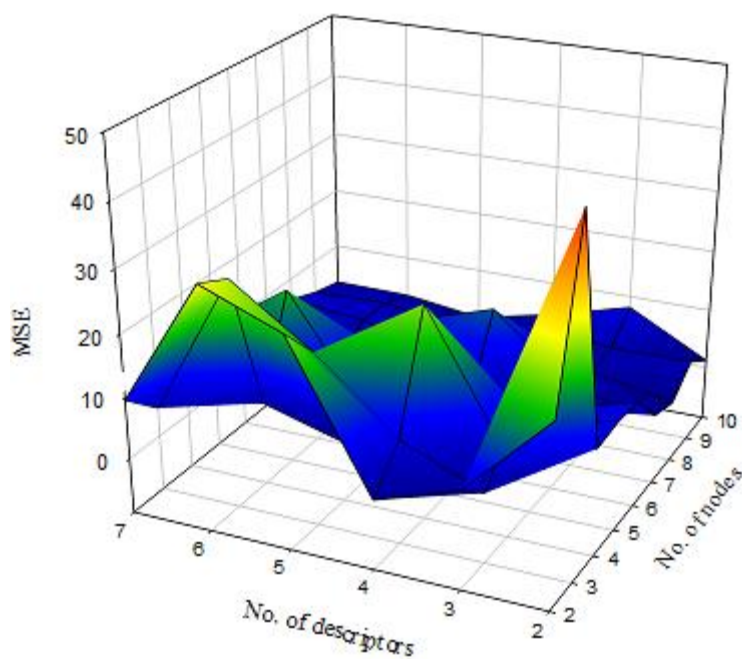
		تعداد ورودی های شبکه					
		۲	۳	۴	۵	۶	۷
تعداد گره‌های لایه پنهان	۲	۳۴/۳۳	۲۲۰/۷۵	۱۵/۳۰	۱۵/۵۹	۶۵/۴۷	۴۰/۶۴
	۳	۲۳۳/۳۰	۱۰۸/۰۶	۱۸/۳۸	۶۶۳/۱۵	۲۶۰/۱۸	۱۷/۹۱
	۴	۳۷/۸۷	۱۶/۱۱	۲۷/۶۴	۱۵/۴۸	۹/۹۶	۹/۳۰
	۵	۱۷۰/۵۵	۲۵/۱۰	۵۳/۷۲	۸۵۲/۴۲	۵/۶۷	۷/۶۹
	۶	۳۶۶/۰۴	۴۵/۰۳	۳۴/۱۹	۳۷۵/۲۷	۲۳/۸۱	۱۵/۶۵
	۷	۵۱/۷۲	۱۵/۵۶	۱۰۸۳/۹۴	۶۴/۰۹	۱۱/۵۲	۱۰/۶۱
	۸	۴۵/۰۸	۶۲/۶۰	۲۰۱/۸۰	۱۰/۱۵	۲۸۷/۵۲	۲۰۵/۴۳
	۹	۵۹/۰۱	۱۸۹۹/۸۳	۲۵۰/۲۶	۱۵/۴۴	۵۱/۰۳	۲۵/۸۷
	۱۰	۴۵۵۰/۴۲	۲۳۶/۱۳	۴۸/۸۶	۲۴/۲۴	۶۸/۷۱	۳/۴۸



شکل ۵-۴- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش لوببرگ - مارکوارت و تابع تبدیل تانژانت سیگموئید

جدول ۵-۱۱- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزش تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد ورودی‌های مختلف و تعداد گره‌های متفاوت در لایه پنهان

		تعداد ورودی های شبکه					
		۲	۳	۴	۵	۶	۷
تعداد گره های لایه پنهان	۲	۳۴/۹۷	۴۵/۰۵	۳۴/۸۴	۲۸/۱۹	۲۷/۹۵	۳۲/۸۹
	۳	۳۸/۶۸	۲۴/۵۷	۳۱/۰۴	۲۹/۹۹	۲۱/۹۱	۲۵/۹۷
	۴	۳۶/۹۶	۲۹/۰۵	۲۵/۱۵	۱۲/۱۶	۲۷/۶۱	۱۴/۵۷
	۵	۴۱/۵۹	۳۱/۰۰	۲۴/۰۰	۱۷/۶۰	۱۴/۶۲	۱۳/۰۳
	۶	۴۴/۵۱	۵۰/۰۰	۳۰/۵۱	۲۰/۵۳	۳۰/۰۱	۲۸/۰۳
	۷	۴۱/۷۵	۲۵/۲۱	۲۶/۵۵	۱۸/۱۷	۱۲/۷۸	۲۲/۸۲
	۸	۴۱/۲۱	۲۶/۰۷	۲۱/۲۴	۱۵/۰۳	۱۹/۷۸	۱۰/۲۱
	۹	۴۱/۵۲	۲۳/۵۱	۱۹/۵۸	۱۷/۵۱	۱۵/۲۰	۱۰/۳۳
	۱۰	۴۱/۲۷	۲۳/۷۶	۲۰/۴۱	۱۳/۴۲	۱۰/۶۸	۱۵/۴۳



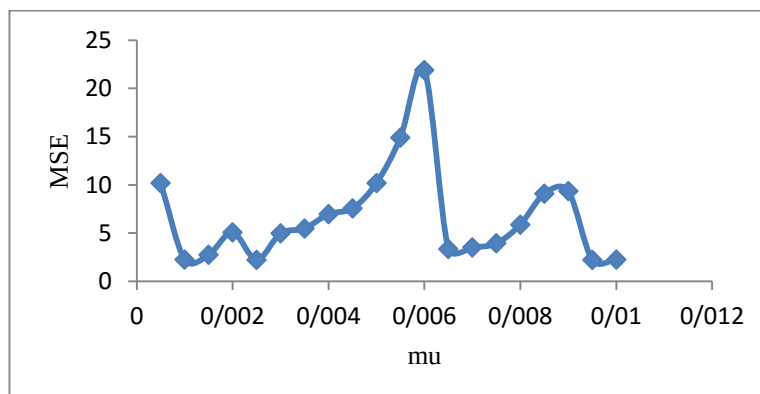
شکل ۵-۵- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزش تنظیم بایزین و تابع تبدیل تانژانت سیگموئید

۵-۶-۱-۴- بهینه سازی پارامتر μ

جهت بهینه کردن مقدار μ ، ساختار شبکه با ۶ متغیر ورودی، ۴ گره در لایه پنهان و الگوریتم آموزشی لونیبرگ - مارکوارت در نظر گرفته شد. سپس مقدار μ از ۰/۰۰۰۵ تا ۰/۰۱ با گام‌های ۰/۰۰۰۵ تغییر داده شد و آنگاه برای هر مورد مقدار میانگین مربع خطای سری ارزیابی محاسبه گردید. در نهایت مقدار میانگین مربع خطا بر حسب μ رسم گردید. نقطه‌ای که کمترین خطا را داشت به عنوان مقدار بهینه انتخاب شد که طبق جدول (۵-۱۲) و شکل (۵-۶) مقدار بهینه ۰/۰۰۲۵ به دست آمد.

جدول ۵-۱۲- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ

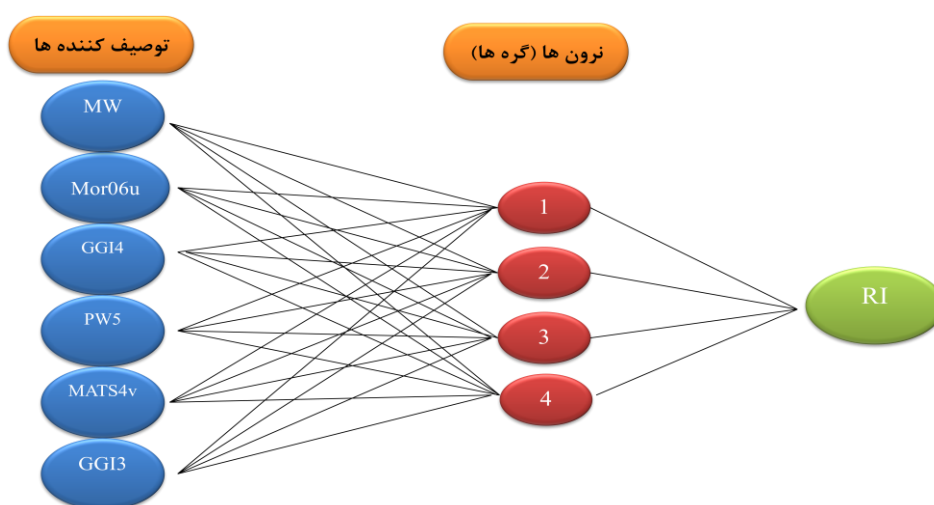
مقادیر μ	میانگین مربع خطا	مقادیر μ	میانگین مربع خطا
۰/۰۰۰۵	۱۰/۲۰	۰/۰۰۵۵	۱۴/۹۲
۰/۰۰۱	۲/۲۶	۰/۰۰۶	۲۱/۸۹
۰/۰۰۱۵	۲/۷۵	۰/۰۰۶۵	۳/۳۶
۰/۰۰۲	۵/۰۷	۰/۰۰۷	۳/۵۲
۰/۰۰۲۵	۲/۲۳	۰/۰۰۷۵	۳/۹۴
۰/۰۰۳	۴/۹۸	۰/۰۰۸	۵/۸۹
۰/۰۰۳۵	۵/۴۸	۰/۰۰۸۵	۹/۰۸
۰/۰۰۴	۶/۹۷	۰/۰۰۹	۹/۳۵
۰/۰۰۴۵	۷/۵۷	۰/۰۰۹۵	۲/۲۴
۰/۰۰۵	۱۰/۲۰	۰/۰۱	۲/۲۷



شکل ۵-۶- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی بر حسب پارامتر μ

۵-۶-۳- معماری شبکه‌ی عصبی بهینه شده

با توجه به نتایج حاصل در مرحله‌ی بهینه‌سازی، طرح شماتیک یک شبکه‌ی عصبی پیش‌خور سه لایه با الگوی آموزش پس انتشار و الگوریتم آموزشی لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد متغیر ورودی ۶ و تعداد نرون لایه‌ی پنهان ۴ در شکل (۵-۷) نشان داده شده است.



شکل ۵-۷- تصویر شماتیک ساختار شبکه عصبی بدست آمده پس از بهینه سازی

فصل ششم

بحث و نتیجه گیری

۶-۱- ارزیابی مدل‌های خطی و غیرخطی

۶-۱-۱- ارزیابی مدل‌های خطی و غیرخطی با استفاده از سری‌های ارزیابی و تست به منظور مطالعه‌ی قدرت پیش‌بینی مدل‌های منتخب خطی و غیرخطی به دست آمده، سری‌های ارزیابی و تست مورد استفاده قرار گرفتند. در این راستا در مدل خطی با استفاده از ضرائب مدل خطی مربوط به سری آموزش و در مدل غیرخطی با استفاده از شبکه عصبی بهینه شده، مقادیر اندیس بازداري مربوط به سری‌های ارزیابی و تست (هر سری شامل ۳۹ ترکیب) که در ساخت مدل استفاده نشده بودند، پیش‌بینی شد (جداول (۶-۱) و (۶-۲)). با نگاهی به مقادیر مجذور ضریب همبستگی در شکل‌های (۶-۱) و (۶-۲) مربوط به سری ارزیابی و شکل‌های (۶-۳) و (۶-۴) مربوط به سری تست، توانایی بالای مدل خطی و غیرخطی در پیش‌بینی اندیس بازداري این ترکیبات تصدیق می‌شود.

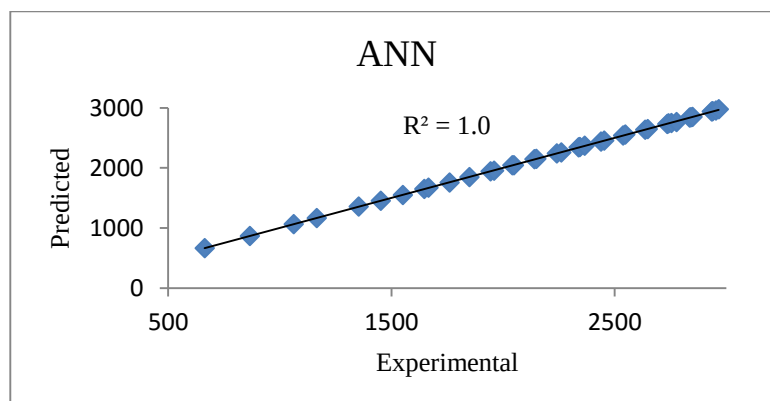
جدول ۶-۱- مقادیر تجربی و پیش بینی شده RI برای ترکیبات سری ارزیابی با استفاده از روش های

خطی و غیرخطی

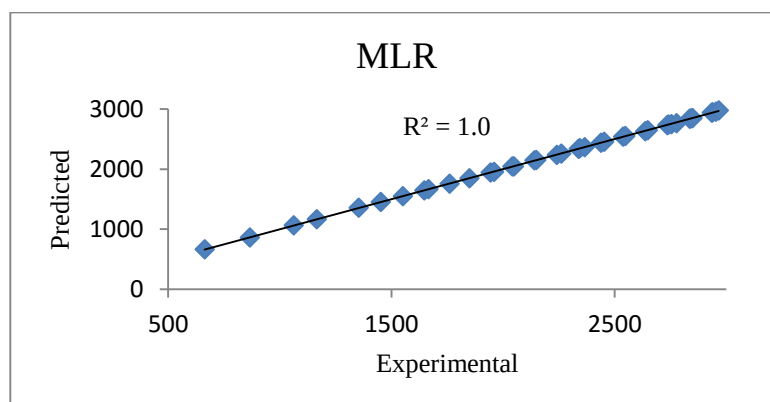
شماره‌ی ترکیب	مقدار واقعی	RI			
		مقدار پیش بینی شده		درصد خطای نسبی	
		MLR	ANN	MLR	ANN
۵	۶۶۲/۴۸	۶۶۴/۴۰	۶۶۰/۱۰	۰/۲۹	-۰/۳۶
۱۱	۸۶۵/۰۰	۸۶۴/۹۰	۸۶۴/۴۰	-۰/۰۱	-۰/۰۷
۱۸	۱۰۶۲/۰۴	۱۰۶۳/۹۰	۱۰۵۹/۲۰	۰/۱۸	-۰/۲۷
۲۴	۱۱۶۵/۴۸	۱۱۶۷/۹۰	۱۱۶۶/۶۰	۰/۲۱	۰/۱۰
۳۲	۱۳۵۲/۶۰	۱۳۵۶/۷۰	۱۳۵۳/۰۰	۰/۳۰	۰/۰۳
۳۸	۱۴۵۱/۶۳	۱۴۵۳/۲۰	۱۴۵۱/۷۰	۰/۱۲	۰/۰
۴۵	۱۵۵۰/۶۶	۱۵۵۱/۵۰	۱۵۵۰/۲۰	۰/۰۵	-۰/۰۳
۵۰	۱۶۴۶/۹۶	۱۶۴۸/۰۰	۱۶۴۷/۱۰	۰/۰۶	۰/۰۱
۵۵	۱۶۶۵/۳۵	۱۶۶۹/۳۰	۱۶۶۵/۶۰	۰/۲۴	۰/۰۲
۶۲	۱۷۵۹/۹۴	۱۷۵۷/۱۰	۱۷۵۹/۲۰	-۰/۱۶	-۰/۰۴
۶۸	۱۸۴۹/۳۴	۱۸۴۶/۹۰	۱۸۴۷/۴۰	-۰/۱۳	-۰/۱۰
۷۴	۱۹۴۳/۰۱	۱۹۴۳/۶۰	۱۹۴۳/۳۰	۰/۰۳	۰/۰۱
۷۹	۱۹۵۹/۹۴	۱۹۵۴/۶۰	۱۹۶۰/۳۰	-۰/۲۷	۰/۰۲
۸۲	۲۰۴۱/۶۵	۲۰۴۶/۲۰	۲۰۴۲/۱۰	۰/۲۲	۰/۰۲
۸۶	۲۰۴۸/۷۹	۲۰۴۶/۱۰	۲۰۴۷/۵۰	-۰/۱۳	-۰/۰۶
۹۱	۲۱۴۰/۳۷	۲۱۴۶/۶۰	۲۱۴۰/۷۰	۰/۳۰	۰/۰۲
۹۶	۲۱۴۸/۳۶	۲۱۵۳/۴۰	۲۱۴۷/۳۰	۰/۲۳	-۰/۰۵
۱۰۲	۲۲۳۹/۶۵	۲۲۳۹/۹۰	۲۲۴۱/۲۰	۰/۰۱	۰/۰۷
۱۰۸	۲۲۶۰/۰۳	۲۲۵۶/۲۰	۲۲۶۰/۶۰	-۰/۱۷	۰/۰۳
۱۱۲	۲۳۳۸/۱۵	۲۳۴۰/۴۰	۲۳۴۰/۰۰	۰/۱۰	۰/۰۸
۱۱۶	۲۳۴۴/۲۵	۲۳۵۰/۰۰	۲۳۴۲/۵۰	۰/۲۵	-۰/۰۷
۱۲۰	۲۳۶۵/۰۴	۲۳۶۷/۷۰	۲۳۶۷/۲۰	۰/۱۱	۰/۰۹
۱۲۳	۲۴۳۷/۶۱	۲۴۴۱/۱۰	۲۴۳۹/۰۰	۰/۱۴	۰/۰۶
۱۲۹	۲۴۵۲/۹۷	۲۴۵۳/۹۰	۲۴۵۳/۸۰	۰/۰۴	۰/۰۳
۱۳۴	۲۵۳۶/۵۴	۲۵۳۹/۶۰	۲۵۳۷/۰۰	۰/۱۲	۰/۰۲
۱۴۰	۲۵۴۷/۸۵	۲۵۵۴/۲۰	۲۵۴۶/۳۰	۰/۲۵	-۰/۰۶
۱۴۵	۲۶۳۵/۴۴	۲۶۳۴/۳۰	۲۶۳۲/۷۰	-۰/۰۴	-۰/۱۰
۱۵۲	۲۶۴۷/۹۱	۲۶۴۷/۹۰	۲۶۴۳/۶۰	۰/۰	-۰/۱۶
۱۵۷	۲۷۳۴/۹۳	۲۷۳۲/۳۰	۲۷۳۴/۳۰	۰/۱۰	-۰/۰۲
۱۶۳	۲۷۴۱/۰۷	۲۷۴۴/۱۰	۲۷۴۰/۷۰	۰/۱۱	-۰/۰۱
۱۶۶	۲۷۵۳/۲۲	۲۷۵۰/۱۰	۲۷۵۲/۶۰	-۰/۱۱	-۰/۰۲

ادامه‌ی جدول ۱-۶

شماره‌ی ترکیب	RI				
	مقدار واقعی	مقدار پیش‌بینی شده		درصد خطای نسبی	
		MLR	ANN	MLR	ANN
۱۶۹	۲۷۷۶/۰۹	۲۷۶۵/۳۰	۲۷۷۶/۹۰	-۰/۳۹	۰/۰۳
۱۷۳	۲۸۳۵/۸۸	۲۸۳۸/۹۰	۲۸۳۵/۶۰	۰/۱۱	-۰/۰۱
۱۷۶	۲۸۴۱/۱۹	۲۸۴۳/۴۰	۲۸۳۸/۲۰	۰/۰۸	-۰/۱۱
۱۷۸	۲۸۴۸/۰۴	۲۸۴۹/۰۰	۲۸۴۶/۲۰	۰/۰۳	-۰/۰۶
۱۸۴	۲۹۳۳/۸۲	۲۹۴۱/۸۰	۲۹۳۴/۳۰	۰/۲۷	۰/۰۲
۱۸۹	۲۹۳۸/۹۰	۲۹۴۵/۹۰	۲۹۳۶/۹۰	۰/۲۴	-۰/۰۷
۱۹۳	۲۹۵۳/۴۳	۲۹۵۶/۱۰	۲۹۵۲/۷۰	۰/۰۹	-۰/۰۲
۱۹۵	۲۹۶۵/۷۲	۲۹۷۷/۲۰	۲۹۶۶/۸۰	۰/۳۹	۰/۰۴



شکل ۱-۶- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل ANN



شکل ۲-۶- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل MLR

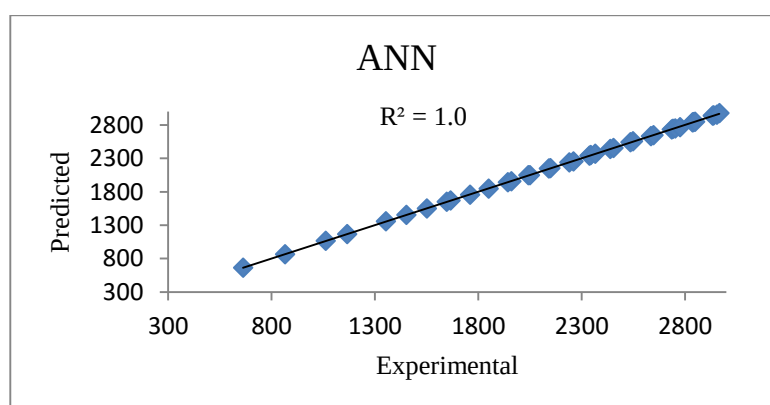
جدول ۶-۲- مقادیر تجربی و پیش بینی شده‌ی اندیس بازداري برای ترکیبات سری تست با استفاده از روش‌های خطی و

غیرخطی

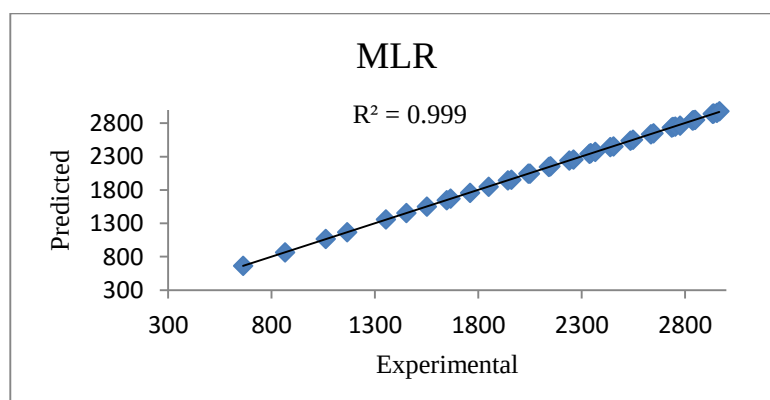
شماره‌ی ترکیب	مقدار واقعی	RI			
		مقدار پیش بینی شده		درصد خطای نسبی	
		MLR	ANN	MLR	ANN
۲	۴۶۶/۲۳	۴۸۰/۸۰	۴۵۸/۹۰	۳/۱۳	-۱/۵۷
۷	۷۶۴/۳۲	۷۶۴/۴۰	۷۶۱/۶۰	۰/۰۱	-۰/۳۶
۹	۷۷۲/۱۷	۷۶۴/۱۰	۷۷۴/۷۰	-۱/۰۵	۰/۳۳
۱۵	۹۶۵/۳۹	۹۶۵/۶۰	۹۶۵/۷۰	۰/۰۲	۰/۰۳
۲۱	۱۱۵۶/۱۶	۱۱۵۶/۰۰	۱۱۵۶/۶۰	-۰/۰۱	۰/۰۴
۲۸	۱۲۶۰/۷۵	۱۲۶۵/۴۰	۱۲۵۷/۶۰	۰/۳۷	-۰/۲۵
۳۴	۱۳۶۰/۳۵	۱۳۶۶/۵۰	۱۳۵۶/۴۰	۰/۴۵	-۰/۲۹
۳۶	۱۳۷۲/۳۳	۱۳۶۹/۵۰	۱۳۷۲/۳۰	-۰/۲۱	۰/۰
۴۲	۱۴۷۲/۵۱	۱۴۶۶/۱۰	۱۴۷۲/۸۰	-۰/۴۴	۰/۰۲
۴۷	۱۵۵۹/۹۷	۱۵۶۱/۲	۱۵۵۸/۱۰	۰/۰۸	-۰/۱۲
۵۱	۱۶۴۷/۶۳	۱۶۴۷/۴	۱۶۴۸/۷۰	-۰/۰۱	۰/۰۶
۵۸	۱۷۴۵/۵۵	۱۷۴۳/۵۰	۱۷۴۶/۶۰	-۰/۱۲	۰/۰۶
۶۱	۱۷۵۳/۶۵	۱۷۵۲/۰۰	۱۷۵۲/۱۰	-۰/۰۹	-۰/۰۹
۶۵	۱۸۴۴/۰۳	۱۸۴۴/۷۰	۱۸۴۵/۱۰	۰/۰۴	۰/۰۶
۷۱	۱۸۶۵/۲۸	۱۸۶۶/۱۰	۱۸۶۶/۳۰	۰/۰۴	۰/۰۵
۷۷	۱۹۴۸/۹۹	۱۹۴۴/۰۰	۱۹۴۵/۹۰	-۰/۲۶	-۰/۱۶
۸۰	۱۹۶۵/۲۳	۱۹۶۳/۰۰	۱۹۶۶/۰۰	-۰/۱۱	۰/۰۴
۸۵	۲۰۴۵/۴۵	۲۰۴۵/۴۰	۲۰۴۴/۷۰	۰/۰	-۰/۰۴
۸۸	۲۰۶۰/۱۶	۲۰۵۷/۰۰	۲۰۵۹/۵۰	-۰/۱۵	-۰/۰۳
۹۳	۲۱۴۱/۲۰	۲۱۴۷/۷۰	۲۱۴۳/۹۰	۰/۳۰	۰/۱۳
۹۹	۲۱۶۵/۲۳	۲۱۶۳/۳۰	۲۱۶۶/۴۰	-۰/۰۹	۰/۰۵
۱۰۶	۲۲۴۸/۱۵	۲۲۴۶/۰۰	۲۲۴۵/۸۰	-۰/۱۰	-۰/۱۰
۱۱۰	۲۲۷۴/۳۴	۲۲۶۳/۰۰	۲۲۷۴/۱۰	-۰/۵۰	-۰/۰۱
۱۱۵	۲۳۴۱/۶۹	۲۳۴۳/۷۰	۲۳۴۰/۹۰	۰/۰۹	-۰/۰۳
۱۱۸	۲۳۵۲/۸۸	۲۳۵۸/۲۰	۲۳۵۳/۹۰	۰/۲۳	۰/۰۴
۱۲۲	۲۴۳۷/۳۵	۲۴۴۰/۸۰	۲۴۳۸/۹۰	۰/۱۴	۰/۰۶
۱۲۶	۲۴۴۱/۵۲	۲۴۴۶/۸۰	۲۴۴۰/۸۰	۰/۲۲	-۰/۰۳
۱۳۲	۲۴۷۴/۸۶	۲۴۵۹/۶۰	۲۴۷۵/۴۰	-۰/۶۲	۰/۰۲
۱۳۷	۲۵۳۹/۳۶	۲۵۴۶/۷۰	۲۵۳۹/۸۰	۰/۲۹	۰/۰۲
۱۴۲	۲۵۶۰/۶۰	۲۵۶۱/۶۰	۲۵۶۰/۹۰	۰/۰۴	۰/۰۱
۱۴۸	۲۶۳۷/۳۵	۲۶۴۰/۰۰	۲۶۳۵/۶۰	۰/۱۰	-۰/۰۷

ادامه‌ی جدول ۶-۲

شماره‌ی ترکیب	RI				
	مقدار واقعی	مقدار پیش‌بینی شده		درصد خطای نسبی	
		MLR	ANN	MLR	ANN
۱۵۴	۲۶۶۱/۷۰	۲۶۵۹/۷۰	۲۶۶۰/۳۰	-۰/۰۸	-۰/۰۵
۱۶۰	۲۷۳۶/۱۶	۲۸۶۵/۳۰	۲۷۳۷/۷۰	۴/۷۲	۰/۰۶
۱۶۵	۲۷۴۷/۸۲	۲۷۴۵/۵۰	۲۷۴۹/۵۰	-۰/۰۸	۰/۰۶
۱۷۱	۲۸۳۴/۵۷	۲۸۳۹/۹۰	۲۸۳۴/۷۰	۰/۱۹	۰/۰
۱۷۵	۲۸۳۹/۰۷	۲۸۴۱/۵۰	۲۸۳۷/۷۰	۰/۰۹	-۰/۰۵
۱۸۱	۲۸۶۵/۷۰	۲۸۶۳/۳۰	۲۸۶۵/۸۰	-۰/۰۸	۰/۰
۱۸۶	۲۹۳۴/۸۵	۲۹۴۱/۹۰	۲۹۳۴/۴۰	۰/۲۴	-۰/۰۲
۱۹۱	۲۹۴۳/۹۳	۲۹۵۰/۱۰	۲۹۳۹/۵۰	۰/۲۱	-۰/۱۵



شکل ۶-۳- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل ANN



شکل ۶-۴- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل MLR

۶-۱-۲- ارزیابی مدل‌های خطی و غیرخطی با استفاده از روش رد مرحله‌ای تک‌تک

به منظور بررسی بیشتر مدل‌های خطی و غیرخطی، با استفاده از روش رد مرحله‌ای تک‌تک مقدار اندیس بازداری ترکیبات محاسبه شد که نتایج حاصله در جدول (۳-۶) ارائه شده است. این نتایج تعمیم‌پذیری بالای مدل‌های طراحی شده توسط هر دو روش خطی و غیرخطی را بیان می‌کند. در روش رگرسیون خطی، یک ترکیب از سایر ترکیبات جدا شده و با بقیه‌ی ترکیبات با روش اجباری مدل ساخته و مقدار اندیس بازداری برای ترکیب کنار گذاشته شده محاسبه می‌شود. در روش غیرخطی یک ترکیب به عنوان سری تست و بقیه‌ی ترکیبات به عنوان سری آموزش در نظر گرفته شده و شبکه‌ی بهینه مقدار اندیس بازداری برای سری تست را پیش‌بینی می‌کند. با رسم منحنی مقدار تجربی اندیس بازداری بر حسب مقدار پیش‌بینی شده‌ی آن و محاسبه‌ی ضریب همبستگی خطی می‌توان این دو مدل را مقایسه کرد.

جدول ۶-۳- مقادیر تجربی، پیش بینی شده و خطای حاصله به روش رد مرحله‌ای تک تک با مدل‌های ANN و MLR

No.	Monomethylalkane	RI	نتایج رد مرحله‌ای تک تک توسط		خطای مطلق	
			ANN	MLR	ANN	MLR
۱	2-Methylpropane	۳۵۴/۷۷	۳۷۸/۱۳	۳۳۶/۱۴	۲۳/۳۶	-۱۸/۶۳
۲	2-Methylbutane	۴۶۶/۲۳	۴۷۲/۷۵	۴۸۰/۴۲	۶/۵۲	۱۴/۱۹
۳	2-Methylpentane	۵۶۱/۳۱	۵۵۲/۰۵	۵۴۸/۸۵	-۹/۲۶	-۱۲/۴۶
۴	3-Methylpentane	۵۷۸/۰۵	۵۳۷/۷۲	۵۷۹/۹۷	-۴۰/۳۳	۱/۹۲
۵	2-Methylhexane	۶۶۲/۴۸	۶۹۱/۴۳	۶۶۳/۶۳	۲۸/۹۵	۱/۱۵
۶	3-Methylhexane	۶۷۲/۱۹	۶۷۴/۶۱	۶۶۷/۴۴	۲/۴۲	-۴/۷۵
۷	2-Methylheptane	۷۶۴/۳۲	۷۷۰/۲۷	۷۶۲/۳۱	۵/۹۵	-۲/۰۱
۸	4-Methylheptane	۷۶۵/۸۸	۷۸۶/۵۲	۷۶۳/۴۶	۲۰/۶۴	-۲/۴۲
۹	3-Methylheptane	۷۷۲/۱۷	۸۱۲/۰۲	۷۷۵/۲۸	۳۹/۸۵	۳/۱۱
۱۰	4-Methyloctane	۸۶۴/۰۶	۹۰۱/۹۹	۸۶۶/۹۷	۳۷/۹۳	۲/۹۱
۱۱	2-Methyloctane	۸۶۵/۰۰	۸۶۷/۹۹	۸۶۴/۴۸	۲/۹۹	-۰/۵۲
۱۲	3-Methyloctane	۸۷۱/۸۹	۸۸۱/۳۸	۸۷۰/۷۹	۹/۴۹	-۱/۱۰
۱۳	5-Methylnonane	۹۶۱/۰۹	۹۷۰/۰۰	۹۶۴/۰۰	۸/۹۱	۲/۹۱
۱۴	4-Methylnonane	۹۶۲/۸۳	۹۷۱/۲۲	۹۶۲/۱۵	۸/۳۹	-۰/۶۸
۱۵	2-Methylnonane	۹۶۵/۳۹	۹۵۸/۵۸	۹۶۵/۵۸	-۶/۸۱	۰/۱۹
۱۶	3-Methylnonane	۹۷۲/۰۶	۹۷۳/۷۶	۹۷۱/۰۰	۱/۷۰	-۱/۰۵
۱۷	5-Methyldecane	۱۰۵۸/۹۴	۱۰۷۳	۱۰۵۹/۲۰	۱۴/۰۶	۰/۲۶
۱۸	4-Methyldecane	۱۰۶۲/۰۴	۱۰۶۰/۳	۱۰۶۱/۸۴	-۱/۷۴	-۰/۲۰
۱۹	2-Methyldecane	۱۰۶۵/۶۲	۱۰۵۵/۸	۱۰۶۷/۹۰	-۹/۸۲	۲/۲۸
۲۰	3-Methyldecane	۱۰۷۲/۰۶	۱۰۶۸/۶	۱۰۷۱/۴۹	-۳/۴۶	-۰/۵۷
۲۱	6-Methylundecane	۱۱۵۶/۱۶	۱۱۴۹/۵	۱۱۵۴/۴۶	-۶/۶۶	-۱/۷۰
۲۲	5-Methylundecane	۱۱۵۷/۳۶	۱۱۶۱/۷	۱۱۵۷/۸۲	۴/۳۴	۰/۴۶
۲۳	4-Methylundecane	۱۱۶۱/۲۱	۱۱۶۱/۴	۱۱۵۷/۶۶	۰/۱۹	-۳/۵۵
۲۴	2-Methylundecane	۱۱۶۵/۴۸	۱۱۵۲	۱۱۶۹/۱۳	-۱۳/۴۸	۳/۶۵
۲۵	3-Methylundecane	۱۱۷۲/۱۵	۱۱۶۵/۷	۱۱۷۲/۰۶	-۶/۴۵	-۰/۰۹
۲۶	6-Methyldodecane	۱۲۵۴/۱۵	۱۲۴۱/۲	۱۲۵۳/۰۹	-۱۲/۹۵	-۱/۰۶
۲۷	5-Methyldodecane	۱۲۵۶/۱۸	۱۲۵۳/۳	۱۲۵۶/۵۶	-۲/۸۸	۰/۳۸
۲۸	4-Methyldodecane	۱۲۶۰/۷۵	۱۲۵۳/۷	۱۲۶۳/۵۲	-۷/۰۵	۲/۷۷
۲۹	2-Methyldodecane	۱۲۶۵/۳۶	۱۲۵۰/۴	۱۲۷۰/۴۵	-۱۴/۹۶	۵/۰۹

ادامه ی جدول ۳-۶

No.	Monomethylalkane	مقادیر تجربی RI	نتایج رد مرحله ای تک تک توسط		خطای مطلق	
			ANN	MLR	ANN	MLR
۳۰	3-Methyldodecane	۱۲۷۲/۱۲	۱۲۶۵/۱	۱۲۷۲/۸۸	-۷/۰۲	۰/۷۶
۳۱	7-Methyltridecane	۱۳۵۱/۹۴	۱۳۳۹/۱	۱۳۵۲/۶۸	-۱۲/۸۴	۰/۷۴
۳۲	6-Methyltridecane	۱۳۵۲/۶	۱۳۳۹/۱	۱۳۵۲/۷۰	-۱۳/۵۰	۰/۱۰
۳۳	5-Methyltridecane	۱۳۵۵/۴۳	۱۳۴۹/۷	۱۳۵۶/۴۶	-۵/۷۳	۱/۰۳
۳۴	4-Methyltridecane	۱۳۶۰/۳۵	۱۳۵۰/۳	۱۳۶۳/۵۰	-۱۰/۰۵	۳/۱۵
۳۵	2-Methyltridecane	۱۳۶۵/۳۵	۱۳۵۱/۵	۱۳۷۱/۸۰	-۱۳/۸۵	۶/۴۵
۳۶	3-Methyltridecane	۱۳۷۲/۳۳	۱۳۶۶/۴	۱۳۷۳/۸۰	-۵/۹۳	۱/۴۷
۳۷	7-Methyltetradecane	۱۴۵۰/۱۳	۱۴۳۳/۴	۱۴۴۹/۲۴	-۱۶/۷۳	-۰/۸۹
۳۸	6-Methyltetradecane	۱۴۵۱/۶۳	۱۴۳۷/۴	۱۴۵۰/۴۷	-۱۴/۲۳	-۱/۱۶
۳۹	5-Methyltetradecane	۱۴۵۴/۷۱	۱۴۴۹	۱۴۵۴/۱۶	-۵/۷۱	-۰/۵۵
۴۰	4-Methyltetradecane	۱۴۵۴/۷۱	۱۴۴۹/۴	۱۴۶۲/۵۵	-۵/۳۱	۷/۸۴
۴۱	2-Methyltetradecane	۱۴۵۶/۳۷	۱۴۴۸	۱۴۶۹/۷۳	-۸/۳۷	۱۳/۳۶
۴۲	3-Methyltetradecane	۱۴۷۲/۵۱	۱۴۶۷/۴	۱۴۷۲/۴۳	-۵/۱۱	-۰/۰۸
۴۳	8-Methylpentadecane	۱۵۴۸/۱۹	۱۵۳۳	۱۵۴۷/۶۵	-۱۵/۱۹	-۰/۵۴
۴۴	7-Methylpentadecane	۱۵۴۸/۸۵	۱۵۳۶/۴	۱۵۴۸/۳۱	-۱۲/۴۵	-۰/۵۴
۴۵	6-Methylpentadecane	۱۵۵۰/۶۶	۱۵۳۷/۲	۱۵۴۹/۱۲	-۱۳/۴۶	-۱/۵۳
۴۶	5-Methylpentadecane	۱۵۵۴/۲۴	۱۵۴۸/۱	۱۵۵۲/۳۸	-۶/۱۴	-۱/۸۶
۴۷	4-Methylpentadecane	۱۵۵۹/۹۷	۱۵۴۹/۶	۱۵۶۱/۸۳	-۱۰/۳۷	۱/۸۶
۴۸	2-Methylpentadecane	۱۵۶۵/۲۴	۱۵۴۹/۲	۱۵۶۹/۲۳	-۱۶/۰۴	۳/۹۹
۴۹	3-Methylpentadecane	۱۵۷۲/۶۷	۱۵۶۹/۳	۱۵۷۱/۷۳	-۳/۳۷	-۰/۹۴
۵۰	8-Methylhexadecane	۱۶۴۶/۹۶	۱۶۳۳/۱	۱۶۴۵/۹۴	-۱۳/۸۶	-۱/۰۲
۵۱	7-Methylhexadecane	۱۶۴۷/۶۳	۱۶۳۶/۳	۱۶۴۶/۶۶	-۱۱/۳۳	-۰/۹۷
۵۲	6-Methylhexadecane	۱۶۵۰/۰۷	۱۶۴۰/۴	۱۶۴۸/۲۱	-۹/۶۷	-۱/۸۶
۵۳	5-Methylhexadecane	۱۶۵۳/۹۷	۱۶۴۷/۱	۱۶۵۰/۱۸	-۶/۸۷	-۳/۸۰
۵۴	4-Methylhexadecane	۱۶۵۹/۹۱	۱۶۵۰/۳	۱۶۶۱/۰۲	-۹/۶۱	۱/۱۱
۵۵	2-Methylhexadecane	۱۶۶۵/۳۵	۱۶۵۶/۴	۱۶۷۱/۷۹	-۸/۹۵	۶/۴۴
۵۶	3-Methylhexadecane	۱۶۷۲/۹۹	۱۶۷۲/۹	۱۶۷۱/۶۹	-۰/۰۹	-۱/۳۰
۵۷	9-Methylheptadecane	۱۷۴۵/۴۰	۱۷۳۵/۷	۱۷۴۴/۶۰	-۹/۷۰	-۰/۸۰
۵۸	8-Methylheptadecane	۱۷۴۵/۵۵	۱۷۳۵/۱	۱۷۴۴/۲۲	-۱۰/۴۵	-۱/۳۳
۵۹	7-Methylheptadecane	۱۷۴۶/۹۳	۱۷۳۷/۱	۱۷۴۵/۵۶	-۹/۸۳	-۱/۳۷
۶۰	6-Methylheptadecane	۱۷۴۹/۷۱	۱۷۴۳/۸	۱۷۴۸/۹۲	-۵/۹۱	-۰/۷۹
۶۱	5-Methylheptadecane	۱۷۵۳/۶۵	۱۷۵۱	۱۷۵۱/۲۲	-۲/۶۵	-۲/۴۳

۶۲	4- Methylheptadecan	۱۷۵۹/۹۴	۱۷۵۷/۱	۱۷۵۹/۷۱	-۲/۸۴	-۰/۲۳
----	---------------------	---------	--------	---------	-------	-------

ادامه‌ی جدول ۳-۶

No.	Monomethylalkane	RI مقادیر تجربی	نتایج رد مرحله‌ای تک تک توسط		خطای مطلق	
			ANN	MLR	ANN	MLR
۶۳	2-Methylheptadecane	۱۷۶۵/۲۹	۱۷۴۷/۸	۱۷۶۶/۱۲	-۱۷/۴۹	۰/۸۳
۶۴	3-Methylheptadecane	۱۷۷۳/۲۱	۱۷۷۴/۲	۱۷۷۱/۳۴	۰/۹۹	-۱/۸۷
۶۵	9-Methyloctadecane	۱۸۴۴/۰۳	۱۸۳۸/۳	۱۸۴۴/۲۶	-۵/۷۳	۰/۲۳
۶۶	8-Methyloctadecane	۱۸۴۴/۵۶	۱۸۳۹/۹	۱۸۴۴/۶۱	-۴/۶۶	۰/۰۵
۶۷	7-Methyloctadecane	۱۸۴۶/۵۱	۱۸۳۹/۲	۱۸۴۴/۸۲	-۷/۳۱	-۱/۶۹
۶۸	6-Methyloctadecane	۱۸۴۹/۳۴	۱۸۴۳/۲	۱۸۴۶/۲۶	-۶/۱۴	-۳/۰۸
۶۹	5-Methyloctadecane	۱۸۵۳/۶۱	۱۸۵۵/۴	۱۸۵۰/۷۲	۱/۷۹	-۲/۸۹
۷۰	4-Methyloctadecane	۱۸۵۹/۹۷	۱۸۵۴/۳	۱۸۶۰/۵۰	-۵/۶۷	۰/۵۳
۷۱	2-Methyloctadecane	۱۸۶۵/۲۸	۱۸۶۰/۹	۱۸۷۰/۳۷	-۴/۳۸	۵/۰۹
۷۲	3-Methyloctadecane	۱۸۷۳/۴۴	۱۸۷۸	۱۸۷۰/۸۴	۴/۵۶	-۲/۶۰
۷۳	10-Methylnonadecane	۱۹۴۲/۶۱	۱۹۴۰	۱۹۴۳/۲۱	-۲/۶۱	۰/۶۰
۷۴	9-Methylnonadecane	۱۹۴۳/۰۱	۱۹۴۰	۱۹۴۳/۲۰	-۳/۰۱	۰/۱۹
۷۵	8-Methylnonadecane	۱۹۴۳/۷۴	۱۹۳۸/۸	۱۹۴۲/۵۳	-۴/۹۴	-۱/۲۱
۷۶	7-Methylnonadecane	۱۹۴۵/۷۹	۱۹۴۲/۵	۱۹۴۳/۸۳	-۳/۲۹	-۱/۹۶
۷۷	6-Methylnonadecane	۱۹۴۸/۹۹	۱۹۴۳/۴	۱۹۴۴/۳۵	-۵/۵۹	-۴/۶۴
۷۸	5-Methylnonadecane	۱۹۵۳/۴۵	۱۹۵۵/۷	۱۹۴۸/۸۳	۲/۲۵	-۴/۶۲
۷۹	4-Methylnonadecane	۱۹۵۹/۹۴	۱۹۵۲/۳	۱۹۵۸/۳۷	-۷/۶۴	-۱/۵۷
۸۰	2-Methylnonadecane	۱۹۶۵/۲۳	۱۹۶۰	۱۹۶۸/۳۹	-۵/۲۳	۳/۱۶
۸۱	3-Methylnonadecane	۱۹۷۳/۸۴	۱۹۷۶/۸	۱۹۶۸/۷۴	۲/۹۶	-۵/۱۰
۸۲	10-Methyleicosane	۲۰۴۱/۶۵	۲۰۴۴/۹	۲۰۴۳/۹۰	۳/۲۵	۲/۲۵
۸۳	9-Methyleicosane	۲۰۴۲/۱۷	۲۰۴۵	۲۰۴۴/۰۰	۲/۸۳	۱/۸۳
۸۴	8-Methyleicosane	۲۰۴۳/۲۸	۲۰۴۴/۲	۲۰۴۳/۵۶	۰/۹۲	۰/۲۸
۸۵	7-Methyleicosane	۲۰۴۵/۴۵	۲۰۴۷/۲	۲۰۴۴/۴۷	۱/۷۵	-۰/۹۸
۸۶	6-Methyleicosane	۲۰۴۸/۷۹	۲۰۵۰/۹	۲۰۴۵/۸۳	۲/۱۱	-۲/۹۶
۸۷	5-Methyleicosane	۲۰۵۳/۳۹	۲۰۵۹/۵	۲۰۴۹/۰۳	۶/۱۱	-۴/۳۶
۸۸	4-Methyleicosane	۲۰۶۰/۱۶	۲۰۵۶/۹	۲۰۵۸/۹۸	-۳/۲۶	-۱/۱۸
۸۹	2-Methyleicosane	۲۰۶۵/۳۴	۲۰۶۴	۲۰۶۸/۳۶	-۱/۳۴	۳/۰۲
۹۰	3-Methyleicosane	۲۰۷۴/۱۵	۲۰۸۲/۵	۲۰۶۹/۷۰	۸/۳۵	-۴/۴۵
۹۱	11-Methylheneicosane	۲۱۴۰/۳۷	۲۱۴۷/۶	۲۱۴۳/۵۶	۷/۲۳	۳/۱۹
۹۲	10-Methylheneicosane	۲۱۴۰/۴۸	۲۱۴۳/۱	۲۱۴۱/۲۵	۲/۶۲	۰/۷۷

۹۳	9-Methylheneicosane	۲۱۴۱/۲۰	۲۱۵۱/۶	۲۱۴۵/۰۷	۱۰/۴۰	۳/۸۷
۹۴	8-Methylheneicosane	۲۱۴۲/۵۷	۲۱۵۱	۲۱۴۴/۷۴	۸/۴۳	۲/۱۷

ادامه ی جدول ۳-۶

No.	Monomethylalkane	RI مقادیر تجربی	نتایج رد مرحله ای تک تک توسط خطای مطلق			
			ANN	MLR	ANN	MLR
۹۵	7-Methylheneicosane	۲۱۴۴/۹۷	۲۱۵۶/۷	۲۱۴۷/۷۵	۱۱/۷۳	۲/۷۸
۹۶	6-Methylheneicosane	۲۱۴۸/۳۶	۲۱۵۹/۸	۲۱۴۸/۷۵	۱۱/۴۴	۰/۳۹
۹۷	5-Methylheneicosane	۲۱۵۳/۲۴	۲۱۶۲	۲۱۴۸/۶۴	۸/۷۶	-۴/۶۰
۹۸	4-Methylheneicosane	۲۱۶۰/۰۵	۲۱۷۲/۸	۲۱۶۴/۱۳	۱۲/۷۵	۴/۰۸
۹۹	2-Methylheneicosane	۲۱۶۵/۲۳	۲۱۶۷/۸	۲۱۶۸/۶۰	۲/۵۷	-۶/۷۸
۱۰۰	3-Methylheneicosane	۲۱۷۴/۳۰	۲۱۸۰/۵	۲۱۶۷/۵۲	۶/۲۰	۰/۴۶
۱۰۱	11-Methyldocosane	۲۲۳۹/۲۶	۲۲۴۳/۳	۲۲۳۹/۷۲	۴/۰۴	۰/۴۶
۱۰۲	10-Methyldocosane	۲۲۳۹/۶۵	۲۲۴۵/۸	۲۲۴۰/۸۹	۶/۱۵	۱/۲۴
۱۰۳	9-Methyldocosane	۲۲۴۰/۷۱	۲۲۴۳/۸	۲۲۳۹/۹۷	۳/۰۹	-۰/۷۴
۱۰۴	8-Methyldocosane	۲۲۴۲/۲۷	۲۲۴۶/۸	۲۲۴۱/۳۴	۴/۵۳	-۰/۹۳
۱۰۵	7-Methyldocosane	۲۲۴۴/۶۶	۲۲۴۵/۹	۲۲۴۰/۸۹	۱/۲۴	-۳/۷۷
۱۰۶	6-Methyldocosane	۲۲۴۸/۱۵	۲۲۵۴/۶	۲۲۴۴/۶۷	۶/۴۵	-۳/۴۸
۱۰۷	5-Methyldocosane	۲۲۵۳/۰۴	۲۲۶۰/۷	۲۲۴۶/۷۰	۷/۶۶	-۶/۳۴
۱۰۸	4-Methyldocosane	۲۲۶۰/۰۳	۲۲۶۱/۳	۲۲۵۸/۵۱	۱/۲۷	-۱/۵۲
۱۰۹	2-Methyldocosane	۲۲۶۵/۰۶	۲۲۵۸/۷	۲۲۶۳/۹۸	-۶/۳۶	-۱/۰۸
۱۱۰	3-Methyldocosane	۲۲۷۴/۳۴	۲۲۸۷/۱	۲۲۶۹/۶۰	۱۲/۷۶	-۴/۷۴
۱۱۱	12-Methyltricosane	۲۳۳۸/۰۳	۲۳۴۶/۷	۲۳۴۰/۴۸	۸/۶۷	۲/۴۵
۱۱۲	11-Methyltricosane	۲۳۳۸/۱۵	۲۳۴۶/۹	۲۳۴۰/۵۹	۸/۷۵	۲/۴۴
۱۱۳	10-Methyltricosane	۲۳۳۸/۶۹	۲۳۴۵/۵	۲۳۳۹/۹۲	۶/۸۱	۱/۲۳
۱۱۴	9-Methyltricosane	۲۳۴۰/۶۷	۲۳۵۱/۲	۲۳۴۲/۵۷	۱۰/۵۳	۲/۵۶
۱۱۵	8-Methyltricosane	۲۳۴۱/۶۹	۲۳۵۰/۱	۲۳۴۲/۰۶	۸/۴۱	۰/۳۷
۱۱۶	7-Methyltricosane	۲۳۴۴/۲۵	۲۳۵۶/۴	۲۳۴۵/۰۳	۱۲/۱۵	۰/۷۸
۱۱۷	6-Methyltricosane	۲۳۴۷/۹۲	۲۳۶۵/۹	۲۳۴۹/۱۶	۱۷/۹۸	۱/۲۴
۱۱۸	5-Methyltricosane	۲۳۵۲/۸۸	۲۳۷۳/۵	۲۳۵۱/۸۶	۲۰/۶۲	-۱/۰۲
۱۱۹	4-Methyltricosane	۲۳۶۰/۰۷	۲۳۷۹/۸	۲۳۶۵/۳۲	۱۹/۷۳	۵/۲۵
۱۲۰	2-Methyltricosane	۲۳۶۵/۰۴	۲۳۷۷/۹	۲۳۷۰/۷۵	۱۲/۸۶	۵/۷۱
۱۲۱	3-Methyltricosane	۲۳۷۴/۷۰	۲۳۸۷/۷	۲۳۶۹/۱۷	۱۳/۰۰	-۵/۵۳
۱۲۲	12-Methyltetracosane	۲۴۳۷/۳۵	۲۴۴۶/۸	۲۴۴۰/۲۳	۹/۴۵	۲/۸۸
۱۲۳	11-Methyltetracosane	۲۴۳۷/۶۱	۲۴۴۷/۱	۲۴۴۰/۳۴	۹/۴۹	۲/۷۳
۱۲۴	10-Methyltetracosane	۲۴۳۸/۲۵	۲۴۴۸	۲۴۴۰/۷۸	۹/۷۵	۲/۵۳
۱۲۵	9-Methyltetracosane	۲۴۳۹/۷۴	۲۴۴۹/۱	۲۴۴۱/۲۶	۹/۳۶	۱/۵۲

۱۲۶	8-Methyltetracosane	۲۴۴۱/۵۲	۲۴۵۲/۸	۲۴۴۲/۹۷	۱۱/۲۸	۱/۴۵
-----	---------------------	---------	--------	---------	-------	------

ادامه ی جدول ۳-۶

No.	Monomethylalkane	مقادیر تجربی RI	نتایج رد مرحله ای تک تک توسط			
			خطای مطلق			
			ANN	MLR	ANN	MLR
۱۲۷	7-Methyltetracosane	۲۴۴۴/۲۰	۲۴۵۶/۶	۲۴۴۴/۴۵	۱۲/۴۰	۰/۲۵
۱۲۸	6-Methyltetracosane	۲۴۴۷/۹۵	۲۴۵۶/۶	۲۴۴۴/۴۱	۸/۵۶	-۳/۵۴
۱۲۹	5-Methyltetracosane	۲۴۵۲/۹۷	۲۴۶۸/۴	۲۴۴۹/۲۳	۱۵/۴۳	-۳/۷۴
۱۳۰	4-Methyltetracosane	۲۴۶۰/۱۴	۲۴۷۳/۵	۲۴۶۲/۵۸	۱۳/۳۶	۲/۴۴
۱۳۱	2-Methyltetracosane	۲۴۶۵/۰۷	۲۴۷۹/۳	۲۴۷۱/۱۲	۱۴/۲۳	۶/۰۵
۱۳۲	3-Methyltetracosane	۲۴۷۴/۸۶	۲۴۸۴	۲۴۶۷/۶۷	۹/۱۴	-۷/۱۹
۱۳۳	13-Methylpentacosane	۲۵۳۶/۴۷	۲۵۴۳	۲۵۳۸/۸۷	۶/۵۳	۲/۴۰
۱۳۴	12-Methylpentacosane	۲۵۳۶/۵۴	۲۵۴۳/۶	۲۵۳۹/۱۳	۷/۰۶	۲/۵۹
۱۳۵	11-Methylpentacosane	۲۵۳۶/۹۸	۲۵۴۶	۲۵۴۰/۲۱	۹/۰۲	۳/۲۳
۱۳۶	10-Methylpentacosane	۲۵۳۷/۷۴	۲۵۴۹/۸	۲۵۴۱/۹۳	۱۲/۰۶	۴/۱۹
۱۳۷	9-Methylpentacosane	۲۵۳۹/۳۶	۲۵۵۰/۹	۲۵۴۲/۴۲	۱۱/۵۴	۳/۰۶
۱۳۸	8-Methylpentacosane	۲۵۴۱/۳۲	۲۵۵۷/۳	۲۵۴۵/۳۸	۱۵/۹۸	۴/۰۶
۱۳۹	7-Methylpentacosane	۲۵۴۳/۹۸	۲۵۵۶	۲۵۴۴/۵۳	۱۲/۰۲	۰/۵۵
۱۴۰	6-Methylpentacosane	۲۵۴۷/۸۵	۲۵۶۱/۲	۲۵۴۶/۸۸	۱۳/۳۵	-۰/۹۷
۱۴۱	5-Methylpentacosane	۲۵۵۳/۱۵	۲۵۶۲/۳	۲۵۴۶/۸۳	۹/۱۵	-۶/۳۲
۱۴۲	4-Methylpentacosane	۲۵۶۰/۶۰	۲۵۶۶/۸	۲۵۶۰/۴۶	۶/۲۰	-۰/۱۴
۱۴۳	2-Methylpentacosane	۲۵۶۵/۲۹	۲۵۶۵/۶	۲۵۶۶/۲۱	۰۰/۳۱	۰/۹۲
۱۴۴	3-Methylpentacosane	۲۵۷۵/۴۵	۲۵۸۲/۶	۲۵۶۷/۸۶	۷/۱۵	-۷/۵۹
۱۴۵	13-Methylhexacosane	۲۶۳۵/۴۴	۲۶۳۴/۱	۲۶۳۶/۰۶	-۱/۳۴	۰/۶۲
۱۴۶	12-Methylhexacosane	۲۶۳۵/۸۷	۲۶۳۵/۳	۲۶۳۶/۵۵	-۰/۵۷	۰/۶۸
۱۴۷	11-Methylhexacosane	۲۶۳۶/۳۱	۲۶۳۶/۸	۲۶۳۷/۱۹	۰/۴۹	۰/۸۸
۱۴۸	10-Methylhexacosane	۲۶۳۷/۳۵	۲۶۴۰/۴	۲۶۳۸/۷۴	۳/۰۵	۱/۳۹
۱۴۹	9-Methylhexacosane	۲۶۳۹/۰۹	۲۶۴۲/۳	۲۶۳۹/۵۸	۳/۲۱	۰/۴۹
۱۵۰	8-Methylhexacosane	۲۶۴۱/۰۹	۲۶۴۷/۶	۲۶۴۱/۸۱	۶/۵۱	۰/۷۲
۱۵۱	7-Methylhexacosane	۲۶۴۳/۸۴	۲۶۵۲/۳	۲۶۴۳/۹۴	۸/۴۶	۰/۱۰
۱۵۲	6-Methylhexacosane	۲۶۴۷/۹۱	۲۶۵۱	۲۶۴۳/۳۴	۳/۰۹	-۴/۵۷
۱۵۳	5-Methylhexacosane	۲۶۵۳/۰۶	۲۶۵۶/۷	۲۶۴۵/۵۳	۳/۶۴	-۷/۵۳
۱۵۴	4-Methylhexacosane	۲۶۶۰/۷۱	۲۶۵۹/۸	۲۶۵۸/۹۶	-۰/۹۱	-۱/۷۵
۱۵۵	2-Methylhexacosane	۲۶۶۵/۳۰	۲۶۵۶/۷	۲۶۶۳/۸۷	-۸/۶۰	-۱/۴۳
۱۵۶	3-Methylhexacosane	۲۶۷۵/۷۲	۲۶۷۳/۶	۲۶۶۵/۳۷	-۲/۱۲	-۱۰/۳۵
۱۵۷	14-Methylheptacosane	۲۷۳۴/۹۳	۲۷۲۸/۵	۲۷۳۵/۶۰	-۶/۴۳	۰/۶۷

۱۵۸ | 13-Methylheptacosane ۲۷۳۵/۰۰ ۲۷۳۰/۳ ۲۷۳۶/۳۰ -۴/۷۰ ۱/۳۰

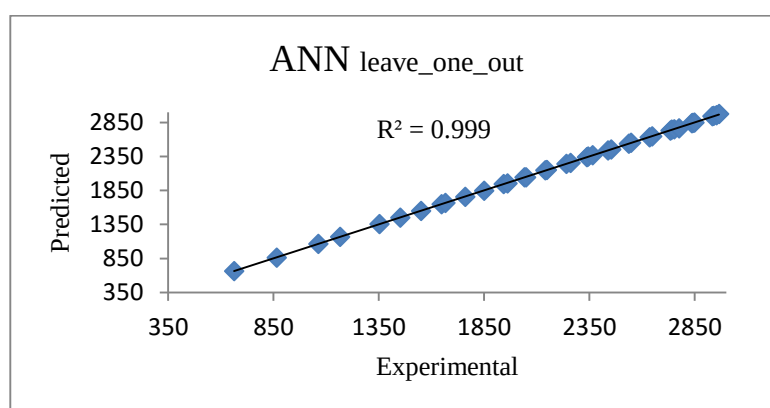
ادامه‌ی جدول ۳-۶

No.	Monomethylalkane	RI مقادیر تجربی	نتایج رد مرحله‌ای تک تک توسط		خطای مطلق	
			ANN	MLR	ANN	MLR
۱۵۹	12-Methylheptacosane	۲۷۳۵/۴۵	۲۷۲۲/۴	۲۷۳۵/۲۶	-۱۳/۰۵	-۰/۱۹
۱۶۰	11-Methylheptacosane	۲۷۳۶/۱۶	۲۷۳۵/۴	۲۷۳۸/۳۹	-۰/۷۶	۲/۲۳
۱۶۱	10-Methylheptacosane	۲۷۳۷/۲۱	۲۷۳۹/۷	۲۷۴۰/۲۶	۲/۴۹	۳/۰۵
۱۶۲	9-Methylheptacosane	۲۷۳۹/۱۴	۲۷۳۸/۸	۲۷۳۹/۸۷	-۰/۳۴	۰/۷۳
۱۶۳	8-Methylheptacosane	۲۷۴۱/۰۷	۲۷۴۱/۶	۲۷۴۱/۱۰	۰/۵۳	۰/۰۳
۱۶۴	7-Methylheptacosane	۲۷۴۳/۸۷	۲۷۴۳	۲۷۴۱/۷۱	-۰/۸۷	-۲/۱۶
۱۶۵	6-Methylheptacosane	۲۷۴۷/۸۲	۲۷۵۴	۲۷۴۶/۵۹	۶/۱۸	-۱/۲۳
۱۶۶	5-Methylheptacosane	۲۷۵۳/۲۲	۲۷۵۲/۷	۲۷۴۵/۷۹	-۰/۵۲	-۷/۴۳
۱۶۷	4-Methylheptacosane	۲۷۶۰/۸۶	۲۷۶۰/۸	۲۷۶۱/۴۵	-۰/۰۶	۰/۵۹
۱۶۸	2-Methylheptacosane	۲۷۶۵/۲۶	۲۷۶۳	۲۷۶۸/۲۹	-۲/۲۶	۳/۰۳
۱۶۹	3-Methylheptacosane	۲۷۷۶/۰۹	۲۷۷۷/۴	۲۷۶۹/۵۱	۱/۳۱	-۶/۵۸
۱۷۰	14-Methyloctacosane	۲۸۳۴/۴۲	۲۸۲۹/۴	۲۸۳۸/۶۶	-۵/۰۲	۴/۲۴
۱۷۱	13-Methyloctacosane	۲۸۳۴/۵۷	۲۸۲۶/۶	۲۸۳۷/۵۱	-۷/۹۷	۲/۹۴
۱۷۲	12-Methyloctacosane	۲۸۳۵/۱۴	۲۸۲۷/۶	۲۸۳۷/۹۰	-۷/۵۴	۲/۷۶
۱۷۳	11-Methyloctacosane	۲۸۳۵/۸۸	۲۸۲۸/۲	۲۸۳۸/۱۷	-۷/۶۸	۲/۲۹
۱۷۴	10-Methyloctacosane	۲۸۳۷/۱۴	۲۸۳۱/۱	۲۸۳۹/۳۸	-۶/۰۴	۲/۲۴
۱۷۵	9-Methyloctacosane	۲۸۳۹/۰۷	۲۸۳۲	۲۸۳۹/۷۸	-۷/۰۷	۰/۷۱
۱۷۶	8-Methyloctacosane	۲۸۴۱/۱۹	۲۸۳۳	۲۸۴۰/۲۰	-۸/۱۹	-۰/۹۹
۱۷۷	7-Methyloctacosane	۲۸۴۳/۹۶	۲۸۳۵/۵	۲۸۴۱/۳۳	-۸/۴۶	-۲/۶۳
۱۷۸	6-Methyloctacosane	۲۸۴۸/۰۵	۲۸۴۱	۲۸۴۳/۷۴	-۷/۰۵	-۴/۳۱
۱۷۹	5-Methyloctacosane	۲۸۵۳/۴۰	۲۸۴۷/۴	۲۸۴۶/۵۰	-۶/۰۰	-۶/۹۰
۱۸۰	4-Methyloctacosane	۲۸۶۱/۱۸	۲۸۵۱	۲۸۶۰/۶۰	-۱۰/۱۸	-۰/۵۸
۱۸۱	2-Methyloctacosane	۲۸۶۵/۷۰	۲۸۵۳/۵	۲۸۶۷/۵۵	-۱۲/۲۰	۱/۸۵
۱۸۲	3-Methyloctacosane	۲۸۷۶/۳۸	۲۸۷۴/۳	۲۸۷۱/۶۴	-۲/۰۸	-۴/۷۳
۱۸۳	15-Methylnonacosane	۲۹۳۳/۷۷	۲۹۲۳	۲۹۳۹/۹۰	-۱۰/۷۷	۶/۱۳
۱۸۴	14-Methylnonacosane	۲۹۳۳/۸۲	۲۹۲۱	۲۹۳۹/۰۲	-۱۲/۸۲	۵/۲۵
۱۸۵	13-Methylnonacosane	۲۹۳۴/۲۶	۲۹۲۲/۶	۲۹۳۹/۷۰	-۱۱/۶۶	۵/۴۴
۱۸۶	12-Methylnonacosane	۲۹۳۴/۸۵	۲۹۲۱/۱	۲۹۳۹/۰۴	-۱۳/۷۵	۴/۱۹
۱۸۷	11-Methylnonacosane	۲۹۳۵/۵۴	۲۹۲۳/۱	۲۹۳۹/۹۶	-۱۲/۴۴	۴/۴۲
۱۸۸	10-Methylnonacosane	۲۹۳۷/۰۰	۲۹۲۳/۴	۲۹۴۰/۰۷	-۱۳/۶۰	۳/۰۷
۱۸۹	9-Methylnonacosane	۲۹۳۸/۹۰	۲۹۲۵/۱	۲۹۴۰/۸۷	-۱۳/۸۰	۱/۹۷

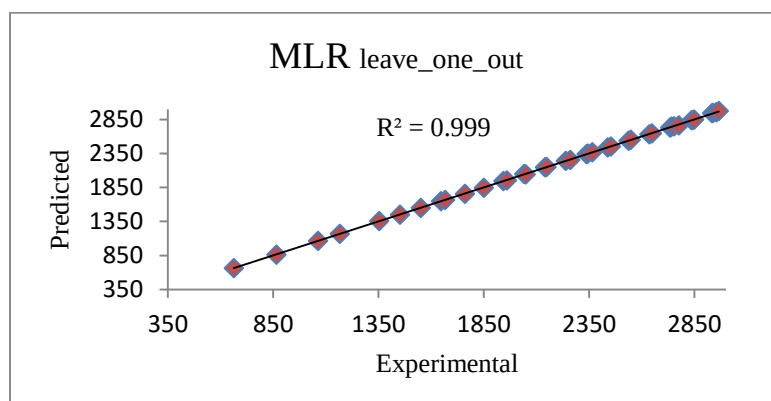
۱۹۰	8-Methylnonacosane	۲۹۴۱/۱۱	۲۹۲۷/۱	۲۹۴۱/۸۱	-۱۴/۰۱	۰/۷۰
۱۹۱	7-Methylnonacosane	۲۹۴۳/۹۳	۲۹۲۹/۱	۲۹۴۲/۷۷	-۱۴/۸۳	-۱/۱۶

ادامه‌ی جدول ۳-۶

No.	Monomethylalkane	مقادیر تجربی RI	نتایج رد مرحله‌ای تک تک توسط		خطای مطلق	
			ANN	MLR	ANN	MLR
۱۹۲	6-Methylnonacosane	۲۹۴۸/۱۴	۲۹۳۴/۲	۲۹۴۵/۰۵	-۱۳/۹۴	-۳/۰۹
۱۹۳	5-Methylnonacosane	۲۹۵۳/۴۳	۲۹۳۹/۴	۲۹۴۷/۳۱	-۱۴/۰۳	-۶/۱۲
۱۹۴	4-Methylnonacosane	۲۹۶۱/۵۶	۲۹۴۵/۹	۲۹۶۲/۹۶	-۱۵/۶۶	۱/۴۰
۱۹۵	2-Methylnonacosane	۲۹۶۵/۷۲	۲۹۶۳/۷	۲۹۷۳/۸۰	-۲/۰۲	۸/۰۸
۱۹۶	3-Methylnonacosane	۲۹۷۶/۴۳	۲۹۶۰/۸	۲۹۶۹/۹۲	-۱۵/۶۳	-۶/۵۱

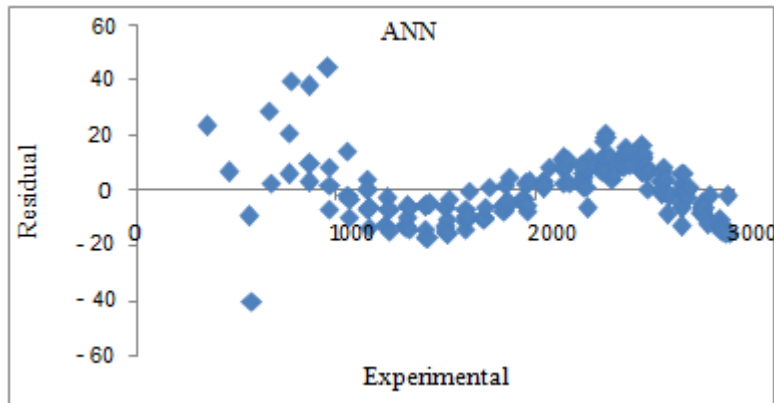


شکل ۵-۶- نمودار مقادیر پیش بینی شده اندیس بازداري با استفاده از شبکه‌ی عصبی مصنوعی بهینه شده برحسب مقادیر واقعی به روش رد مرحله ای تک تک

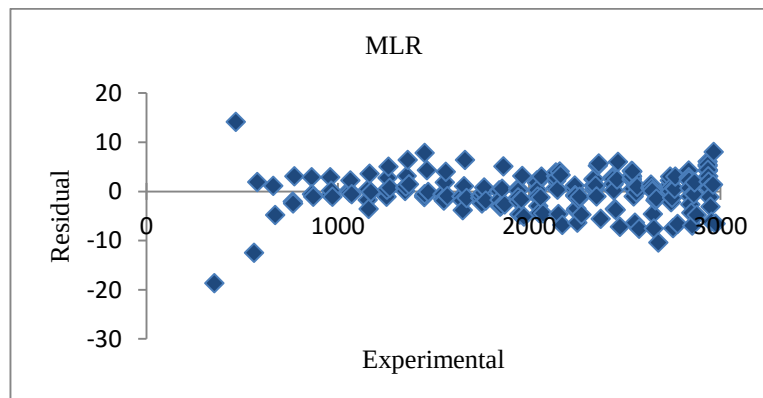


شکل ۶-۶- نمودار مقادیر پیش بینی شده اندیس بازداري به روش خطی برحسب مقادیر واقعی به روش رد مرحله ای تک تک

همچنین در شکل‌های (۶-۷) و (۶-۸) مقادیر باقیمانده‌ها بر حسب مقادیر واقعی اندیس بازداری ترکیبات مورد بحث، ترسیم شده است. تمرکز نقاط به سمت محورهای افقی نشان‌دهنده‌ی عدم وجود خطای معین می‌باشد که البته این تقارن در روش ANN بیشتر از روش MLR می‌باشد.



شکل ۶-۷- نمودار مقادیر خطای مطلق در ارزیابی مدل شبکه‌ی عصبی مصنوعی به روش رد مرحله‌ای تک‌تک در مقابل مقادیر تجربی



شکل ۶-۸- نمودار مقادیر خطای مطلق در ارزیابی مدل خطی به روش رد مرحله‌ای تک‌تک در مقابل مقادیر تجربی

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

جهت بررسی و مقایسه‌ی میزان کارایی دو روش خطی و غیرخطی در پیش‌بینی اندیس بازداری، چندین پارامتر آماری محاسبه گردید. آماره‌های این دو روش در جدول (۶-۴) آمده است.

جدول ۶-۴- مقایسه‌ی آماره‌های بدست آمده از دو مدل ANN و MLR

پارامتر	سری تست (N=39)		سری ارزیابی (N=39)	
	MLR	ANN	MLR	ANN
MAE	۳۱/۷۸	۱۹/۴۳	۲۹/۵۰	۱۷/۲۸
MSE	۲۸/۹۶	۴/۲۱	۱۸/۴۴	۲/۲۴
PRESS	۵۶۶/۴۴	۳۲۴	۳۹۰۲	۲۸۲/۲۴
SEP	۵/۳۸	۲/۰۵	۴/۲۹	۱/۵۰
R ²	۰/۹۹۹	۱/۰	۱/۰	۱/۰
REP (%)	۰/۰۳	۰/۰۲	۰/۰۷	۰/۰۲
RMAE	۰/۲۶	۰/۱۲	۰/۱۶	۰/۰۶

تمام روابط مربوط به آماره‌های محاسبه شده در این قسمت در بخش (۲-۱-۸) آورده شده است. همانطور که در جدول (۶-۴) مشاهده می‌شود، مقادیر خطاهای روش غیر خطی برای هر دو سری تست و ارزیابی نسبت به روش خطی بسیار کوچکتر می‌باشد.

۴-۱-۶- ارزیابی شبکه به روش Y -تصادفی

در اینجا نیز مانند قسمت (۳-۴) برای جلوگیری از همبستگی‌های تصادفی، تست Y -تصادفی انجام شد. نتایج حاصل از چندین بار اجرای این آزمون در جدول (۵-۶) نشان داده شده است که مجذور ضریب همبستگی پایین، نشان‌دهنده‌ی عدم وابستگی شانسی در مدل توسعه یافته توسط شبکه عصبی مصنوعی است.

جدول ۵-۶- مقادیر مجذور ضریب همبستگی سری تست بعد از چندین Y -تصادفی

تکرار	R^2 سری تست
۱	۰/۰۵۷
۲	۰/۱۵
۳	۰/۱۴
۴	۰/۰۲۵
۵	۰/۰۲۱
۶	۰/۰۳۳
۷	۰/۱۰۵
۸	۰/۰۳۳
۹	۰/۰۶۷
۱۰	۰/۰۰۳۵

۶-۲- بررسی ارتباط توصیف‌کننده‌های ساختاری وارد شده در مدل با اندیس

بازداری ترکیبات مورد بررسی

در این قسمت با توجه به توصیف‌کننده‌های وارد شده در مدل‌های MLR و ANN یک بررسی اجمالی بر روابط و اثرات توصیف‌کننده‌های موجود در مدل با شاخص بازداری ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده در مدل خطی و غیر خطی به ترتیب شامل سه و شش توصیف‌کننده است که هر کدام بیان‌گر خصوصیات متفاوتی از مولکولهای مورد بررسی است.

الف) توصیف‌کننده‌های زیر ساختاری^۱

توصیف‌کننده‌های زیر ساختاری ساده‌ترین و پراستفاده‌ترین توصیف‌کننده‌ها بوده که ساختار مولکولی یک ترکیب را بدون هیچ گونه اطلاعاتی در رابطه با هندسه‌ی مولکول‌ها بررسی می‌کند. معمول‌ترین توصیف‌کننده‌های زیرساختاری شامل تعداد اتم‌ها، تعداد پیوندها، تعداد مطلق و نسبی پیوندهای یگانه، دوگانه، سه‌گانه و آروماتیک، تعداد حلقه‌ها، تعداد حلقه‌های بنزن و ... می‌باشد. از میان این توصیف‌کننده‌ها MW^2 در مدل ظاهر شده است که به صورت مجموع وزنهای اتمی محاسبه می‌شود.

این توصیف‌کننده در ارتباط با اندازه‌ی مولکول بوده و حساس به نوع اتم می‌باشد و از رابطه‌ی زیر محاسبه می‌شود:

$$MW = \sum_{i=1}^A m_i \quad \text{معادله‌ی ۶-۱}$$

که در آن m جرم اتمی و A تعداد اتم موجود در مولکول می‌باشد. با توجه به این توصیف‌کننده می‌توان اینطور نتیجه‌گیری کرد که هر چه جرم مولکولی بیشتر شود به علت بالا رفتن سطح تماس

^۱Constitutional

^۲Molecular Weight

ترکیب با ستون، شاخص بازداري افزایش می‌یابد. مثبت بودن اثر متوسط این توصیف‌کننده نیز نشان-
دهنده‌ی ارتباط مستقیم این توصیف‌کننده و شاخص بازداري می‌باشد. چند مثال از اثر این توصیف-
کننده در جدول زیر آمده است.

جدول ۶-۶- اثر توصیف‌کننده‌ی MW بر اندیس بازداري

نام ترکیب	MW	RI
2-methylpropane	۵۸	۳۵۵/۷۷
2-methylbutane	۷۲	۴۶۶/۲۳
2-methylpentane	۸۶	۵۶۱/۳۱
2-methylhexane	۱۰۰	۵۷۸/۰۵

ب) توصیف‌کننده‌های 3D MORSE

همانطور که در بخش (۴-۲-الف) توضیح داده شد، توصیف‌کننده‌های 3D - Morse (نمایش سه
بعدی ساختار مولکول براساس تفرق الکترون) از طریق معادله تبدیلی که در پراش الکترون استفاده
می‌شود، محاسبه می‌گردند:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}} \quad \text{معادله‌ی ۶-۲}$$

I شدت الکترون پراکنده شده، A_i و A_j خاصیت اتمی اتم i و j ، s زاویه پراکندگی، r_{ij} فاصله بین
اتم‌های i و j ، N تعداد کل اتم‌ها را نشان می‌دهد. این روش باعث می‌شود که ساختار سه بعدی
مولکول به یک کد ثابت تبدیل شود.

این توصیف‌کننده‌ها قادرند ارتباط بین ساختار سه بعدی ترکیبات آلی و خصوصیات فیزیکی،
شیمیایی و بیولوژیکی را فراهم سازند. در تحقیق حاضر، برای پیش‌بینی اندیس بازداري، توصیف-
کننده‌ی Mor06u از این گروه توصیف‌کننده‌ها وارد مدل برتر شد که با توجه به اثر متوسط منفی آن
در مدل می‌توان گفت که با افزایش مقادیر این توصیف‌کننده‌ها RI کاهش یابد.

پ) توصیف کننده های Galvez topol . charge indices

توضیحات مربوط به این گروه از توصیف کننده ها در بخش (۴-۲-ت) آمده است. در این تحقیق دوتا از این توصیف کننده ها، GGI3 و GGI4، به کار گرفته شدند که توصیف کننده ی GGI4 در مدل برگزیده ی خطی انتخاب شد و دارای اثر متوسط منفی می باشد، بدین معنی که با افزایش آن مقادیر اندیس بازداري کاهش می یابد.

ت) توصیف کننده های Topological

این دسته از توصیف کننده ها بعنوان توصیف کننده های شکلی^۱ بیان می شوند. PW5^۲ از گروه توصیف کننده های دوبعدی توپولوژیکی می باشد که در تحقیق حاضر وارد مدل برتر در شبکه شده است.

اندیسهای path / walk برای هر اتم بصورت نسبت بین p_i^m (atomic path count) و awc_i^m (atomic walk count) تعریف می شود:

$$(p/w)_i^m = p_i^m / awc_i^m \quad \text{معادله ی ۳-۶}$$

همانطور که نسبت شمارش path / walk مستقل از اندازه ی مولکولی است، این دسته از توصیف کننده ها به عنوان توصیف کننده های شکلی بیان می شوند.

ث) توصیف کننده های 2D-autocorrelation

توضیحات مربوط به این توصیف کننده ها در بخش (۴-۲-ب) آمده است. از میان این توصیف کننده ها، MATS4v که با حجم و اندروالس وزن دار شده در مدل انتخاب شده و نشان می دهد که

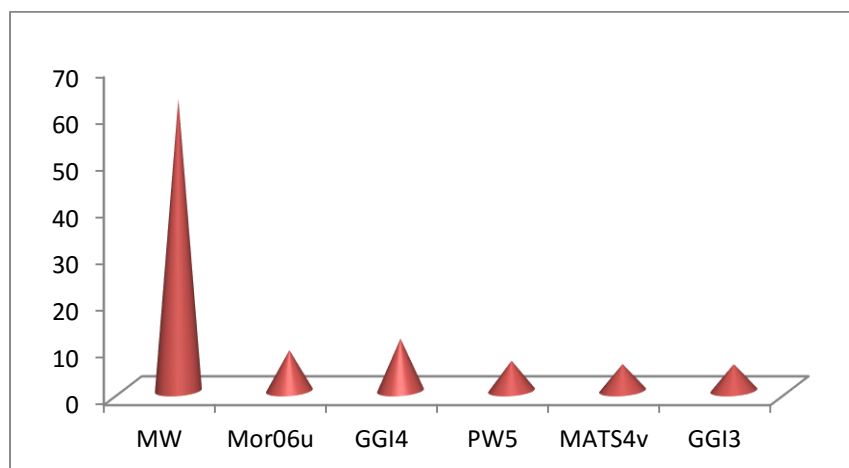
^۱Shape descriptors

^۲Path/Walk

حجم کلی استخلاف، فاکتوری تعیین کننده برای اندیس بازداری آن ترکیب می‌باشد و افزایش این توصیف‌کننده باعث کاهش RI می‌شود.

۳-۶- بررسی میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی

همانطور که در بخش (۳-۴) اشاره شد، هدف از این بخش بررسی میزان تاثیر هر توصیف‌کننده بر روی پارامتر مورد بررسی بوده که در این بخش پارامتر مورد نظر شاخص بازداری می‌باشد. میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی مربوط به مونومتیل آلکان‌ها در شکل (۶-۸) نشان داده شده‌است. همانطور که مشاهده می‌شود، توصیف‌کننده‌ی MW بیشترین اثر مشارکت را در مدل ارائه شده دارد.



شکل ۶-۹- نمودار درصد مشارکت توصیف‌کننده‌های بکار گرفته شده در مدل شبکه عصبی مربوط به مونومتیل آلکان‌ها

۴-۶- نتیجه‌گیری نهایی

هدف اصلی از کار حاضر، توسعه روش‌های خطی و غیرخطی جهت مطالعه‌ی ارتباط کمی بین ساختار و اندیس بازداري مونومتیل آلکان‌ها می‌باشد. نتایج نشان می‌دهند که مدل‌های خطی و غیرخطی توسعه یافته می‌توانند به عنوان روش‌های موفق برای مدل‌سازی و پیش‌بینی اندیس بازداري ترکیبات مورد مطالعه باشند و این مدل‌ها می‌توانند به محققان در پیش‌بینی اندیس‌های بازداري ترکیبات جدید کمک کنند.

آینده‌نگری

- پژوهشگران می‌توانند نتایج حاصل از این مطالعات را در طراحی دارو و پیش‌بینی اندیس‌های بازداري ترکیبات جدید به کار ببرند.
- می‌توان از روش‌هایی همچون ژنتیک الگوریتم و طرح‌ریزی متوالی برای انتخاب توصیف‌کننده‌های مهم و از روش‌هایی مانند SVM^۱ و LS-SVM^۲ به جای ANN استفاده کرد.
- روش‌های سه بعدی مانند CoMFA^۳ و CoMSIA^۴ برای بررسی برهمکنش‌های دارویی و پیش‌بینی‌های مربوطه می‌تواند جایگزین روش‌های معمول دو بعدی گردد.

^۱Successive Projection Algorithm^۲Supported Vector Machine^۳Copative Molecular Field Analysis^۴Copative Molecular Similarity Analysis

References

- [1]. M. Fernandez, A. Tundidor- Camba, J Caballero, (2005) "Modelling of cyclin-dependent kinase inhibition by 1H-pyrazolo[3,4-d] pyrimidine derivatives using artificial neural network ensembles " *J. Chem. Inf. Model.* **45**, 6, pp 1884–1895.
- [2]. S. Samanta, B. Debnath, A. Basu, Gayen, K. SSrikanth, T. Jha, (2006) "Exploring QSAR on 3-aminopyrazoles as anti tumor agents for their inhibitory activity of CDK2/cyclin A" *European Journal of Medical Chemistry*, **41**, pp 1190-1195.
- [3]. J. Li, and Liu, (2007) "structure activity relationship study of oxindole – based inhibitors of cyclin-dependent kinases on least squares support vector machines " *Anal. Chim. Acta.* **581**, 2, pp 333–342.
- [4]. A.R. Katritzky, K. Chen, U. Maran, D.A. Carlson, (2000) "QSPR Correlation and Predictions of GC Retention Indexes for Methyl-Branched Hydrocarbons Produced by Insects" *Anal. Chem.* **72**, 1, pp 101-109.
- [5]. Y. Du, Y. Liang, (2003) "Data mining for seeking accurate quantitative relationship between molecular structure and GC retention indices of alkanes by projection pursuit" *Comput. Biol.Chem.* **27**, 3, pp 339-353.
- [6]. Z. Ziang, Y. Liang, Q. Hu, (2005) "Relativity study of the topological index of methylalkane structures and chromatographic retention index" *Chinese journal of Chromatography*, **23**, 2, pp 117-122.
- [7]. K. Zarei, M. Atabati, (2008) "Prediction of GC Retention Indices for insect-produced methyl substituted alkanes using a wavelet neural network" *Anal. Chem.*, **55**, pp 732-739.
- [8]. منصور عرب چم جنگلی (۱۳۸۶) "پیش بینی فعالیت دارویی ضد ایدز (سیتوتوکسیتی) مشتقات ۵- فیل -۱- فیل آمینو ۱-هیدروکسی ایمیدازول بوسیله ی شبکه عصبی مصنوعی" *دانشگاه صنعتی شاهرود*, گزارش طرح پژوهشی.
- [9]. C. Hansch, T. Fujita, (1964) "p-sigma-pi Analysis: a method for the correlation of biological activity and chemical structure", *J. Am. Chem. Soc.* **86**, pp 1616-1626.
- [10]. T. Clark, (1985) "A Handbook of Computational Chemistry" John Wiley & Sons, New York. pp 145-151
- [11]. M. Karelson, (2000), "Molecular Descriptors in QSAR/QSPR", Wiley-VCH, New York.
- [12]. I.E. Frank, R. Todeschini, (1994). "The Data Analysis Handbook". Elsevier, Amsterdam (The Netherlands), pp 366.
- [13]. I.A. Basheer, M. Hajmeer, (2000). "Artificial neural networks: Fundamentals, computing, design, and application" *J. of Microbiological Methods* **43**, pp 3–31.

- [14]. M.S. Bazaraa, H.D. Sherali, C.M. Shetty, (2006). "Nonlinear Programming Theory and Algorithms", third ed. John-Wiley, New York. pp 853 .
- [15]. W.S. McCulloch, W.H. Pitts, (1943). "A logical calculus of the ideas immanent in neural nets". *Bull. Math Biophys.* 5, pp 115-133.
- [16]. D. O. Hebb, (1949), "The organization of behavior", Wiley, New York. 355 pages.
- [17]. J. Zupan, J. Gasteiger, (1999). "Neural Network in Chemistry and Drug Design". Wiley-VCH Publishers, Weinheim(Germany).
- [18]. J. Zupan, (1994), "Introduction to Artificial Neural Network(ANN) Methods: What They Are and How to Use Them", *Acta Chemica Slovenica*, 41,3 , pp 327-352.
- [19]. E. Rich, K. Knight, (1991). "Artificial Intelligence", 2nd ed.; McGraw-Hill: New York.
- [20]. K.A. Smith, (2005). "Applications of neural networks". *Computers and Operations Research.*, 32, 10, pp 2495–7.
- [21]. M. Lam, (2004). "Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis" *Decision Support Systems*, 37, pp 567–581.
- [22]. D. Rummelhurt, G. Hinton, R. Williams, (1986) "Learning Representation by Back Propagation Error", *Nature*, 323, pp 533-536.
- [23]. J.G. Taylor, C.L.T. Mannion, (1992), "Theory and application of neural networks", Springer-Verlag, New York.
- [24]. راهنمای نرم افزار مطلب ۲۰۰۷.
- [25]. ع. ا. حسینی پاک، م. شرف‌الدین، (۱۳۸۱)، "تحلیل داده‌های اکتشافی"، انتشارات دانشگاه تهران
- [26]. D.E. Rumelhart, G.E. Hinton, and R.J. Williams, (1986), "Learning internal representations by error propagation" D.E. Rumelhart and J. McClelland, editors, *Parallel Data Processing*, Vol.1, Chapter 8, The M.I.T. Press, Cambridge, MA, , pp 318-362.
- [27]. M.T. Hagan, H.B. Demuth, M.H. Beale, "Neural Network Design", chapter 11-12, ISBN 0-9717321-0-8 (available from John Stovall, john.stovall@colorado.edu, 303.492.3648).
- [28]. M.P. Freitas, (2007) "Multivariate QSAR: From Classical Descriptors to New Perspectives" *Curr. Comput. Aided Drug Des.* 3, 4, pp 235–239.
- [29]. C. Norbury, P. Nurse, (1992) "Animal-Cell Cycles and Their Control". *Annu.Rev. Biochem.*, 61, 441–470.
- [30]. D. O. Morgan, (1995) "Principles of CDK Regulation". *Nature (London)*, 374, 6518, pp 131–134.

- [31] M. Hall, G. Peters, **(1996)** "Genetic alterations of cyclins, cyclin-dependent kinases and Cdk inhibitors in human cancer". *Adv. Cancer Res.*, **68**, pp 67–108.
- [32]. H. Jans, Alzate-Morales, Julio Caballero, Ariela Vergara Jague, D. Fernando, J. Gonzalez Nilo, **(2009)** "Insights into the Structural Basis of N2 and O6 Substituted Guanine Derivatives as Cyclin-Dependent Kinase 2 (CDK2) Inhibitors" *Chem. Inf. Model.* **49**, pp 886–899.
- [33]. Y.V. Kissin, G.P. Feulmer, **(1986)** "Catagenesis and composition of petroleum: Origin of *n*-alkanes and isoalkanes in petroleum crudes" *J. Chromatogr. Sci.* **24**, pp 53-59.
- [34]. F. Kenig, J.S.S. Damste, A.C.K. Dalen, W.I.C. Rijpstra, A.Y. Huc, J.W. Deleeuw, **(1995)** "Branched aliphatic alkanes with quaternary substituted carbon atoms in modern and ancient geologic samples" *Geochim. Cosmochim. Acta.* **59**, pp 2999-3015.
- [35]. A.R. Katritzky, K. Chen, U. Maran, D.A. Carlson, **(2000)** "QSPR Correlation and Predictions of GC Retention Indexes for Methyl-Branched Hydrocarbons Produced by Insects" *Anal. Chem.* **72**, 1, pp 101-109.
- [36] M. Phillips, K. Gleeson, J.M. Hughes, J. Greenberg, R.N. Cataneo, L. Baker, W.P. McVay, **(1999)** "Volatile organic compounds in breath as markers of lung cancer: A cross-sectional study" *Lancet*, **353**, pp 1930-1933.
- [37]. M. Phillips, R.N. Cataneo, A.R. Cummin, A.J. Gagliardi, K. Gleeson, J. Greenberg, R.A. Maxfield, W.N. Rom, **(2003)** "Detection of lung cancer with volatile markers in the breath" *Chest*, **123**, pp 2115-2123.
- [38]. M. Phillips, R.N. Cataneo, R. Condos, G.A. Ring Erickson, J. Greenberg, V. La Bombardi, M.I. Munawar, O. Tietje, **(2007)** "Volatile biomarkers of pulmonary tuberculosis in the breath" *Tuberculosis(Edinb.)*, **87**, pp 44-52.
- [39]. W. Miekisch, J.K. Schubert, G.F. Noeldge-Schomburg, **(2004)** "Diagnostic potential of breath analysis--focus on volatile organic compounds" *Clin. Chim. Acta.* **347**, pp 25-39.
- [40]. M. Phillips, J.P. Boehmer, R.N. Cataneo, T. Cheema, H.J. Eisen, J.T. Fallon, P.E. Fisher, A. Gass, J. Greenberg, J. Kobashigawa, D. Mancini, B. Rayburn, M.J. Zucker, **(2004)** "Heart Allograft Rejection: Detection With Breath Alkanes in Low Levels (the HARDBALL Study)" *J. Heart Lung Transplant.* **23**, pp 701-708.
- [41]. T.S.Wang, P. Spanel, D. Smith, in: A. Amann, D. Smith (Eds.), **(2005)** "Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring", World Scientific, Singapore, pp. 479–490.
- [42]. A. Amann, P. Spanel, D. Smith, **(2007)** "Breath Analysis: the Approach towards Clinical Applications" *Mini Rev. Med. Chem.* **7**, pp 115–129.

- [43]. J. Lundberg, in: A. Amann, D. Smith (Eds.), **(2005)** "Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring" World Scientific, Singapore, pp 3-34.
- [44]. L. Gustafsson, in: A. Amann, D. Smith (Eds.), **(2005)** "Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring", World Scientific, Singapore, pp 11-18 (xi-xviii)
- [45]. S.W. Ryter, L.E. Otterbein, **(2004)** "Carbon monoxide in biology and medicine" *Bioessays*, **26**, 3, pp 270-280.
- [46]. D.J. Slebos, S.W. Ryter, A.M. Choi, **(2003)** "Heme oxygenase-1 and carbon monoxide in pulmonary medicine" *Respir. Res.* **4**, pp7.
- [47]. C. Turner, P. Spanel, D. Smith, **(2006)** "A longitudinal study of ammonia, acetone and propanol in the exhaled breath of 30 subjects using selected ion flow tube mass spectrometry, SIFT-MS" *Physiol. Meas.* **27**, pp 321-337.
- [48]. C. Turner, P. Spanel, D. Smith, **(2006)** "A longitudinal study of breath isoprene in healthy volunteers using selected ion flow tube mass spectrometry (SIFT-MS)" *Physiol. meas.* **27**, pp 13-22.
- [49]. T. Karl, P. Prazeller, D. Mayr, A. Jordan, J. Rieder, R. Fall, W. Lindinger, **(2001)** "Human breath isoprene and its relation to blood cholesterol levels: new measurements and modeling" *J. Appl. Physiol.* **91**, pp 762-770.
- [50]. A. Amann, S. Telser, L. Hofer, A. Schmid, H. Hinterhuber, in: A. Amann, D. Smith (Eds.), **(2005)** "Breath Analysis for Clinical Diagnosis and Therapeutic Monitoring" World Scientific, Singapore, pp. 305-316.
- [51]. W. Lindinger, J. Taucher, A. Jordan, A. Hansel, W. Vogel, **(1997)** "production of methanol after consumption of fruit" *Alcohol Clin. Exp. Res.* **21**, pp 939-943.
- [52]. C. Turner, P. Spanel, D. Smith, **(2006)** "A longitudinal study of breath isoprene in healthy volunteers using selected ion flow tube mass spectrometry (SIFT-MS)" *Physiol. Meas.* **27**, 7, pp 637-648.
- [53]. G. von Basum, D. Halmer, P. Hering, M. Murtz, S. Schiller, F. Muller, A. Popp, F. Kuhnemann, **(2004)** "Parts per billion sensitivity for ethane in air with an optical parametric oscillator cavity leak-out spectrometer" *Opt. Lett.* **29**, 8, pp 797-799.
- [54]. H. Dahnke, D. Kleine, P. Hering, M. Murtz, **(2001)** "Real-time monitoring of ethane in human breath using mid-infrared cavity leak-out spectroscopy" *J. Appl. Phys. B*, **72**, pp 971-975.

- [55]. P. Paredi, S.A. Kharitonov, D. Leak, S.Ward, D. Cramer, P.J. Barnes, **(2000)** “ Exhaled ethane, a marker of lipid peroxidation, is elevated in chronic obstructive pulmonary disease” *Am. J. Respir. Crit Care Med.* **162**, pp 369-373.
- [56]. P. Paredi, S.A. Kharitonov, P.J. Barnes, **(2000)**, “Elevation of Exhaled Ethane Concentration in Asthma” *Am. J. of Respir. Crit. Care Med.* **162**, 4, pp 1450-1454.
- [57]. W. Miekisch, J.K. Schubert, D.A. Vagts, K. Geiger, **(2001)** “Analysis of volatile Disease Markers in Blood ” *Clin. Chem.* **47**, pp 1053-1060.
- [58]. J. Scholpp, J.K. Schubert, W. Miekisch, K. Geiger, **(2002)**” Breath markers and soluble lipid peroxidation markers in critically ill patients ” *Clin. Chem. Lab. Med.* **40**, pp 587-594.
- [59]. S. Chen, L. Zieve, V. Mahadevan, **(1970)** “Mercaptans and dimethyl sulfide in the breath of patients with cirrhosis of the liver” *J. Lab. Clin. Med.* **75**, pp 628-635.
- [60]. N. Ochiai, M. Takino, S. Daishima, D.B. Cardin, **(2001)** “Analysis of volatile sulphur compounds in breath by gas chromatography-mass spectrometry using a three-stage cryogenic trapping preconcentration system. ” *J. chromatogr.B: Biomed. Sci. Appl.* **762**, 1, pp 67-75.
- [61]. S.M. Studer, J.B. Orens, I. Rosas, J.A. Krishnan, K.A. Cope, S. Yang, J.V. Conte, P.B. Becker, T.H. Risby, **(2001)** “Patterns and significance of exhaled-breath biomarkers in lung transplant recipients with acute allograft rejection” *J. Heart Lung Transplant*, **20**, 11, pp 1158-1166.
- [62]. C. Roller, A. Kosterev, F. Tittel, K. Uehara, C. Gmachl, D. Sivco, **(2003)** “Carbonyl sulfide detection with a thermoelectrically cooled midinfrared quantum cascade laser ” *Opt. Lett.* **28**, 21, pp 2052-2054.
- [63]. Y.V. Kissin, G.P. Feulmer, **(1986)** ”Gas Chromatographic Analysis of Alkyl-Substituted Paraffins ” *J. Chromatogr. Sci.* **24**, pp 53-59 (7).
- [64]. Zofia Krko sova, Robert Kubinec, Ladislav Sojak, Anton Amann **(2008)** “Temperature-programmed gas chromatography linear retention indices of all C4–C30 monomethylalkanes on methylsilicone OV-1 stationary phase Contribution towards a better understanding of volatile organic compounds in exhaled breath” *J. of Chromatogr. A*, **1179**, pp 59–68.
- [65]. M.C. Simmons, D.B. Richardson, I. Dvoretzky, in: R.P.W. Scott (Ed.), **(1960)** “*Gas Chromatography*”, Butterworth, London, pp 211.

Abstract

In the first section, quantitative structure-activity relationship (QSAR) models of 56 compounds of CDK2 inhibitors were constructed using linear (multiple linear regression; MLR) and nonlinear (artificial neural network; ANN) methods. The cyclin dependent kinases (CDKs) are compounds which play a fundamental role in cell cycle regulation and inhibit the overactive CDK2 which result in excessive reproduction of cells and finally in cancer. The prediction ability of the models was evaluated using the validation and test data sets, leave-one-out cross validation method and Y-randomization. The mean squared error (MSE) for test set by ANN and MLR models are 0.063 and 0.069, respectively.

In the second section, quantitative structure-property relationship (QSPR) models of retention indices of 196 monomethyl alkanes were constructed using two linear (multiple linear regression; MLR) and nonlinear (artificial neural network; ANN) methods. The prediction ability of the models was evaluated using the validation and test data sets, leave-one-out cross validation method and Y-randomization. The mean squared error (MSE) for test set by ANN and MLR models are 4.21 and 28.96, respectively. The results of both sections show better prediction ability for nonlinear models.

Keywords: Quantitative structure-activity relationship, Quantitative structure-property relationship, Multiple linear regression, Artificial neural network, Biological activity, Retention index.



Shahrood University of Technology

Faculty of Chemistry

**Quantitative structure – activity relationship study of some guanine
derivatives as CDK2 inhibitors in body**

Payam Kalhor

Supervisor:

Dr. N. Goodarzi

Advisor:

Dr. H. Nikoofard

Date: may-2011