

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده شیمی

گروه شیمی فیزیک

عنوان

مدل سازی ANN برای پیش بینی دانسیته ی سیالات تجمعی با استفاده از

توصیف کننده های مولکولی

دانشجو:

نازلی آزمونفر

استاد راهنما:

دکتر زهرا کلانتر

استاد مشاور:

دکتر منصور عرب چم جنگلی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ماه ۱۳۹۲

تقدیم به آنان که همیشه در مسیر زندگی پشتیبان، یاور و مایه‌ی
آرامش من بودند.

تقدیم به:

روح پدر بزرگوارم

مادر مهربانم

همسر صبور و فداکارم

به پاس جبران ذره‌ای از زحماتشان

تشکر و قدردانی

با سپاس ایزد منان که زیر سایه‌ی توجهاتش توفیق اتمام این کار نصیب من گشت، بر خود لازم می‌دانم که از همه‌ی عزیزانی که مرا در این راه یاری رساندند تشکر کنم.

با سپاس فراوان از استاد راهنمای ارجمندم سرکار خانم دکتر کلانتر، که در طی این دوران از راهنمایی‌های علمی و اخلاقی ایشان بهره برده‌ام. با تشکر از استاد مشاور بزرگوارم جناب آقای دکتر عرب به‌خاطر راهنمایی‌های بی‌دریغ ایشان و همچنین از اساتید گروه شیمی آقای دکتر گودرزی، آقای دکتر نیکوفرد خانم دکتر موسوی برای مساعدت‌های فراوانشان در طول انجام این تحقیق سپاسگذاری می‌نمایم.

از تمامی دوستان خوبم خانم‌ها زهرا ابراهیمی، زهرا باقرزاده، رقیه شایق، اکرم مرادی، مهدیه قاضی‌پور، یاسمین شهرکی و تمام دوستانی که در این مدت محبت‌ها و همدلی‌هایشان همراه من بودند کمال تشکر دارم.

در پایان از همسر عزیزم، مادر مهربانم، خانواده همسر و خواهران و برادران عزیزم که در تمامی مراحل، مشوق من بوده‌اند کمال تشکر و قدردانی را دارم.

چکیده

در این تحقیق مطالعات ارتباط کمی ساختار-خاصیت (QSPR)، برای پیش‌بینی دانسیته‌ی سیالات تجمعی شامل الکل‌ها، آمین‌های نوع اول و دوم و کربوکسیلیک اسیدها به کار گرفته شده‌اند. برای تولید توصیف‌کننده‌های مهم از دو روش مختلف استفاده شده است. در روش اول، ابتدا تعداد زیادی توصیف‌کننده شامل ۱۸ دسته‌ی مختلف، توسط نرم افزار Dragon محاسبه شدند و سپس برای انتخاب توصیف‌کننده‌های مهم از روش رگرسیون مرحله‌ای (SR) و الگوریتم ژنتیک بر اساس آنالیز حداقل مربعات (GA-PLS) استفاده شد. تعداد ۸ توصیف‌کننده توسط روش SR و ۵ توصیف‌کننده توسط روش GA-PLS انتخاب شدند. در روش دوم، توصیف‌کننده‌های بهینه بر اساس روش سهم گروه انتخاب شدند. پس از آنالیز ساختار شیمیایی تمام ترکیبات مورد مطالعه در این کار، نهایتاً ۱۰ گروه عاملی (-CH_3 , -CH_2 , >CH- , >CH-OH , $\text{-}\overset{|}{\underset{|}{\text{C}}}\text{-OH}$, COOH , $\text{-CH}_2\text{-NH}_2$, >CH-NH_2 , >NH) ایجاد شدند. توصیف‌کننده‌های انتخاب شده توسط این سه روش و دو توصیف‌کننده‌ی تجربی دما و فشار به عنوان ورودی شبکه‌ی عصبی مصنوعی مورد استفاده قرار گرفتند. ۴ نوع شبکه‌ی عصبی مصنوعی با ترکیب الگوریتم‌های آموزشی لونبرگ-مارکوارت و بایزین با توابع انتقال لگاریتم سیگموئید و تانژانت سیگموئید استفاده گردید. پس از آموزش و بهینه سازی پارامترهای ANN مانند تعداد ورودی، تعداد نرون لایه‌ی پنهان و تعداد دوره‌های آموزشی عملکرد مدل توسط سری تست مورد بررسی قرار گرفت. نتایج حاصل نشان می‌دهد که شبکه‌ی عصبی با الگوریتم آموزشی تنظیم لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید و توصیف‌کننده‌های حاصل از روش سهم گروه می‌تواند رابطه‌ی بین توصیف‌کننده‌های ساختاری و دانسیته‌ی مولکول‌های مورد نظر را شبیه‌سازی کند. میانگین مربعات خطا (MSE) و میانگین درصد انحراف مطلق (AAD) برای سری تست توسط این روش به ترتیب برابر ۱/۲۴۳۲ و ۰/۰۸۶۷٪ می‌باشد.

کلمات کلیدی: QSPR، شبکه عصبی مصنوعی، رگرسیون مرحله‌ای، الگوریتم ژنتیک، سهم گروه.

نتایج حاصل از این پایان نامه در دو پوستر با عناوین:

- **“Density Prediction of Associating Fluids Using a New Artificial Neural Network-Group Contribution Method.”**
- **“QSPR study of Associating Fluids on Density Using CDFS Strategy .”**

در شانزدهمین کنگره شیمی ایران، دانشگاه یزد ارائه گردید.

فهرست مطالب

فصل اول

- ۱-۱- مقدمه..... ۲
- ۲-۱- معرفی سیالات تجمعی..... ۲
- ۳-۱- دانسیته..... ۴
- ۴-۱- روش‌های تجربی اندازه‌گیری دانسیته‌ی سیالات..... ۴
- ۴-۱-۱- اندازه‌گیری دانسیته با استفاده از دانسیتومتر رزونانسی..... ۴
- ۴-۱-۲- اندازه‌گیری دانسیته با استفاده از اندازه‌گیری سرعت صوت در دما و فشارهای مختلف..... ۶
- ۵-۱- روش‌های پیش‌بینی دانسیته‌ی سیالات تجمعی..... ۷
- ۵-۱-۱- پیش‌بینی دانسیته‌ی سیالات تجمعی با استفاده از معادله‌ی حالت..... ۷

فصل دوم

- ۱-۲- مقدمه..... ۱۲
- ۲-۲- کموتریکس..... ۱۲
- ۳-۲- رابطه‌ی کمی ساختار- خاصیت یا ساختار- فعالیت..... ۱۳
- ۳-۲-۱- فراهم کردن سری داده‌ها..... ۱۴
- ۳-۲-۲- رسم ساختار و بهینه‌سازی..... ۱۴
- ۳-۲-۳- محاسبه‌ی توصیف‌کننده‌ها..... ۱۵
- ۳-۲-۴- حذف توصیف‌کننده‌های نامناسب..... ۱۵
- ۳-۲-۵- دسته‌بندی داده‌ها..... ۱۷
- ۳-۲-۶- انتخاب بهترین توصیف‌کننده‌ها..... ۱۸
- ۳-۲-۶-۱- انتخاب بهترین توصیف‌کننده‌ها توسط رگرسیون مرحله‌ای (SR)..... ۱۸
- ۳-۲-۶-۲- انتخاب بهترین توصیف‌کننده‌ها توسط الگوریتم ژنتیک..... ۱۹
- ۳-۲-۶-۲- الف- مقدمه..... ۱۹
- ۳-۲-۶-۲- ب- الگوریتم ژنتیک..... ۲۰
- ۳-۲-۶-۲- ج- ساختار الگوریتم ژنتیکی..... ۲۳
- ۳-۲-۶-۳- انتخاب توصیف‌کننده‌ها بر مبنای روش سهم گروه (GCM)..... ۲۳

- ۲-۳-۷- ساختن مدل ۲۵
- ۲-۳-۷-۱- شبکه‌ی عصبی مصنوعی ۲۶
- ۲-۳-۷-۱-الف- نرون زیستی ۲۶
- ۲-۳-۷-۱-ب- نرون محاسباتی ۲۷
- ۲-۳-۷-۱-پ- تابع انتقال ۲۸
- ۲-۳-۷-۱-ت- ساختار لایه‌ای شبکه‌های عصبی ۲۹
- ۲-۳-۷-۱-ث- انواع شبکه‌ها از نظر ارتباطات بین نرونی ۳۰
- ۲-۳-۷-۱-ج- آموزش شبکه عصبی ۳۱
- ۲-۳-۷-۱-چ- تابع هدف ۳۲
- ۲-۳-۷-۱-ه- معرفی روش‌های مختلف بهینه‌سازی عملکرد شبکه عصبی ۲
- ۲-۳-۸- ارزیابی قدرت پیش‌بینی مدل ۳۵

فصل سوم

- ۳-۱- مطالعه‌ی ارتباط کمی ساختار-خاصیت دانسیته‌ی سیالات تجمعی در محدوده‌ی وسیعی از دما و فشار ۴۲
- ۳-۲- مراحل مدل‌سازی ۴۲
- ۳-۲-۱- جمع‌آوری سری داده‌ها ۴۲
- ۳-۲-۲- رسم و بهینه‌سازی ساختار مولکول‌ها ۴۴
- ۳-۲-۳- محاسبه‌ی توصیف‌کننده‌ها ۴۵
- ۳-۲-۳-۱- محاسبه‌ی توصیف‌کننده‌ها با استفاده از نرم‌افزار Dragon ۴۵
- ۳-۲-۳-۲- محاسبه‌ی توصیف‌کننده‌ها به روش سهم گروه ۴۵
- ۳-۲-۴- حذف توصیف‌کننده‌های نامناسب ۴۹
- ۳-۲-۵- دسته‌بندی داده‌ها ۵۰
- ۳-۲-۶- انتخاب بهترین توصیف‌کننده‌ها ۵۰
- ۳-۲-۶-۱- انتخاب بهترین توصیف‌کننده‌ها به روش رگرسیون مرحله‌ای (SR) ۵۰
- ۳-۲-۶-۲- انتخاب توصیف‌کننده‌های معتبر به روش الگوریتم ژنتیک (GA) ۵۶
- ۳-۲-۷- مدل‌سازی غیرخطی ۵۷
- ۳-۲-۷-۱- مدل‌سازی غیرخطی با استفاده از شبکه عصبی مصنوعی ۵۷

مرحله‌ای (SR-ANN).....	۵۸
۲-۷-۱-الف- بهینه سازی شبکه‌ی عصبی مصنوعی با استفاده از توصیف‌کننده‌های حاصل از رگرسیون	
۲-۷-۱-ب- بهینه‌سازی مدل شبکه‌ی عصبی مصنوعی با توصیف‌کننده‌های الگوریتم ژنتیک (GA-ANN).....	۶۵
۲-۷-۱-ج- مدل‌سازی شبکه‌ی عصبی مصنوعی با استفاده از توصیف‌کننده‌های حاصل از روش سهم گروه	
.....(GCM-ANN)	۷۲
۲-۸-۱- ارزیابی قدرت پیش بینی مدل‌های غیرخطی.....	۸۰
۲-۸-۱-۱- ارزیابی مدل‌های غیر خطی بهینه شده با استفاده از پارامترهای آماری.....	۸۰
۲-۸-۲-۲- با استفاده از نمودار برگشتی.....	۸۱
۲-۸-۳-۳- با استفاده از نمودار خطای باقی‌مانده.....	۸۱
۲-۸-۴-۴- ارزیابی با استفاده از آزمون y -تصادفی.....	۸۶
۲-۸-۵-۵- ارزیابی با استفاده از اعتبارسنجی بیرونی.....	۸۷
۲-۹-۹- بررسی ارتباط توصیف‌کننده‌های وارد شده در مدل با دانسیته‌ی ترکیبات.....	۸۹
۲-۹-۱-۱- توصیف‌کننده‌ی $X0Av$	۸۹
۲-۹-۲-۲- توصیف‌کننده‌های گروه 2D-Autocorrelation.....	۹۰
۲-۹-۳-۳- توصیف‌کننده‌های Constitutional.....	۹۱
۲-۹-۴-۴- توصیف‌کننده‌های تابع توزیع شعاعی (RDF).....	۹۲
۲-۹-۵-۵- توصیف‌کننده‌های GETAWAY.....	۹۳
۲-۱۰-۱۰- بررسی میزان مشارکت توصیف‌کننده‌های منتخب شبکه عصبی.....	۹۵
۳-۳- نتیجه‌گیری.....	۹۷
۴-۴- آینده‌نگری:.....	۹۹
پیوست.....	۱۰۰
فهرست منابع:.....	۱۰۷

فهرست جداول

جدول (۱-۲) - مراحل کلی یک مدل سازی QSPR یا QSAR.....	۱۴
جدول (۲-۲) - انواع توصیف کننده های محاسبه شده توسط نرم افزار Dragon.....	۱۶
جدول (۱-۳) - محدوده ی دما و فشار و تعداد نقطه داده ی استفاده شده در مدل سازی برای سیالات تجمعی مختلف به همراه مرجع مورد استفاده برای هر یک از آن ها.....	۴۳
جدول (۲-۳) - گروه های عاملی موجود در مولکول های مورد بررسی به عنوان توصیف کننده های روش سهم گروه.....	۴۶
جدول (۳-۳) - تعداد گروه های عاملی موجود در مولکول های مختلف سری داده ها به عنوان توصیف کننده های روش سهم گروه.....	۴۷
جدول (۴-۳) - مقادیر همبستگی توصیف کننده های روش GCM با دانسیته.....	۴۹
جدول (۵-۳) - مقایسه ی آماره های سری ارزیابی مدل های به دست آمده از روش رگرسیون مرحله ای.....	۵۲
جدول (۶-۳) - توصیف کننده های انتخاب شده توسط روش رگرسیون مرحله ای.....	۵۳
جدول (۷-۳) - ماتریس همبستگی توصیف کننده های انتخاب شده توسط رگرسیون مرحله ای.....	۵۳
جدول (۸-۳) - مقادیر همبستگی توصیف کننده های انتخاب شده توسط رگرسیون مرحله ای با دانسیته.....	۵۳
جدول (۹-۳) - مقادیر توصیف کننده های انتخاب شده به روش SR.....	۵۴
جدول (۱۰-۳) - تعداد تکرار توصیف کننده های حاصل از GA به ازای ۱۲ بار اجرا.....	۵۶
جدول (۱۱-۳) - مقادیر همبستگی توصیف کننده های انتخابی توسط روش GA با دانسیته.....	۵۷
جدول (۱۲-۳) - توابع و پارامترهای بهینه ی شبکه های طراحی شده با توصیف کننده های حاصل از SR.....	۶۳
جدول (۱۳-۳) - توابع و مقادیر پارامترهای بهینه ی شبکه ی (SR-ANN).....	۶۵
جدول (۱۴-۳) - توابع و پارامترهای بهینه ی شبکه های GA-ANN.....	۷۰
جدول (۱۵-۳) - توابع و مقادیر پارامترهای بهینه شده ی GA-ANN.....	۷۲
جدول (۱۶-۳) - مقادیر MSE سری ارزیابی برای شبکه ی GCM-ANN با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۷۴
جدول (۱۷-۳) - مقادیر MSE سری ارزیابی برای شبکه ی GCM-ANN با الگوریتم آموزشی تنظیم لونبرگ مارکوارت و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۷۵
جدول (۱۸-۳) - مقادیر MSE سری ارزیابی برای شبکه ی GCM-ANN با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۷۶

جدول (۳-۱۹) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی تنظیم لونیبرگ -	
مارکوارت و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۷۷
جدول (۳-۲۰) - توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN.....	۷۸
جدول (۳-۲۱) - توابع و مقادیر پارامترهای بهینه شده GCM-ANN.....	۷۹
جدول (۳-۲۲) - پارامترهای آماری مربوط به هر یک از شبکه‌های بهینه شده برای سری‌های ارزیابی و تست.....	۸۰
جدول (۳-۲۳) - مقادیر R^2 به دست آمده برای سری ارزیابی و سری تست از آزمون y تصادفی در مدل SR-ANN،	
GCM-ANN و GA-ANN.....	۸۶
جدول (۳-۲۴) - مقادیر پارامترهای آماری برای سری تست بیرونی.....	۸۷

فهرست اشکال

- شکل (۱-۲) - ساختار یک نرون طبیعی ۲۷
- شکل (۲-۲) - شمای نرون مصنوعی نام ۲۹
- شکل (۳-۲) - ساختار یک شبکه‌ی سه لایه ۳۰
- شکل (۱-۳) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف کننده‌های حاصل از SR ۵۹
- شکل (۲-۳) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت و با استفاده از توصیف کننده‌های حاصل از SR ۶۰
- شکل (۳-۳) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف کننده‌های حاصل از SR ۶۱
- شکل (۴-۳) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت و با استفاده از توصیف کننده‌های حاصل از SR ۶۲
- شکل (۵-۳) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب تعداد دور آموزشی ۶۴
- شکل (۶-۳) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ ۶۴
- شکل (۷-۳) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف کننده‌های حاصل از GA ۶۶
- شکل (۸-۳) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت و با استفاده از توصیف کننده‌های حاصل از GA ۶۷

- شکل (۳-۹) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف کننده‌های حاصل از GA ۶۸
- شکل (۳-۱۰) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت با استفاده از توصیف کننده‌های حاصل از GA ۶۹
- شکل (۳-۱۱) - نمودار میانگین مربعات خطا برای سری ارزیابی بر حسب تعداد دورهای آموزش متفاوت با گام ۲ ۷۰
- شکل (۳-۱۲) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ ۷۱
- شکل (۳-۱۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۳) ۷۴
- شکل (۳-۱۴) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۴) ۷۵
- شکل (۳-۱۵) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۵) ۷۶
- شکل (۳-۱۶) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۶) ۷۷
- شکل (۳-۱۷) - نمودار میانگین مربعات خطا برای سری ارزیابی بر حسب تعداد دور آموزشی متفاوت با گام ۲ ۷۸
- شکل (۳-۱۸) - نمودار میانگین مربعات خطا بر حسب μ حول نقطه بهینه برای سری ارزیابی ۷۹
- شکل (۳-۱۹) - نمودار مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری ارزیابی (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN ۸۲
- شکل (۳-۲۰) - نمودار مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN ۸۳
- شکل (۳-۲۱) - نمودار خطای باقی‌مانده بر حسب مقدار تجربی دانسیته برای سری ارزیابی (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN ۸۴
- شکل (۳-۲۲) - نمودار خطای باقی‌مانده بر حسب مقدار تجربی دانسیته برای سری تست (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN ۸۵
- شکل (۳-۲۳) - نمودار مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست بیرونی (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN ۸۸
- شکل (۳-۲۴) - درصد مشارکت توصیف کننده‌ها در مدل SR-ANN ۹۵
- شکل (۳-۲۵) - درصد مشارکت توصیف کننده‌ها در مدل GA-ANN ۹۶
- شکل (۳-۲۶) - درصد مشارکت توصیف کننده‌ها در مدل GCM-ANN ۹۶

فصل اول

مقدمه

۱-۱- مقدمه

سیالات تجمعی به علت داشتن پیوند هیدروژنی و ایجاد کمپلکس‌های تجمعی (دایمر، تریمر، ...) رفتارهای غیرایده‌آلی زیادی از خود نشان می‌دهند، به گونه‌ای که این موضوع باعث می‌شود رفتار فازی و حجمی این سیالات بسیار پیچیده باشد [۱]. بنابراین، بررسی خواص ترمودینامیکی سیالات تجمعی از جمله دانسیته، به عنوان یک مقوله‌ی جداگانه، از اهمیت بالایی در صنایع شیمیایی و پتروشیمیایی و در طراحی‌های صنعتی، خطوط لوله و پمپاژها برخوردار است.

۱-۲- معرفی سیالات تجمعی

توسعه‌ی تئوری دقیق برای سیالات تجمعی، یک زمینه‌ی چالش برانگیز در تحقیقات ترمودینامیکی است. یکی از مهم‌ترین آثار شیمیایی موجود در ترمودینامیک این سیالات، پیوند هیدروژنی می‌باشد [۱]. اگرچه پیوند هیدروژنی از اغلب نیروهای بین مولکولی دیگر قوی‌تر است، اما این پیوند را می‌توان به عنوان یک برهم‌کنش واندروالسی قوی در نظر گرفت که قدرت آن بین پیوندهای لاندن^۱ و کووالانسی قرار می‌گیرد.

برای نمایش پیوند هیدروژنی معمولاً از نماد $A-H \cdots B$ استفاده می‌گردد که A و B عموماً اتم‌هایی هستند که الکترونگاتیوتر از هیدروژن می‌باشند مثل، F, Cl, O, S, N. گروه A-H به عنوان پذیرنده‌ی الکترون یا دهنده‌ی پیوند هیدروژنی و B به عنوان دهنده‌ی الکترون یا پذیرنده‌ی پیوند هیدروژنی عمل می‌کنند. هنگام تشکیل پیوند هیدروژنی اتم الکترونگاتیو A با جذب الکترون از ابر الکترونی اتم هیدروژن باعث القای بار جزئی مثبت بر روی اتم هیدروژن می‌گردد. در نتیجه اتم هیدروژن جفت الکترون‌های آزاد اتم دیگر را که پذیرنده‌ی پیوند هیدروژنی است جذب می‌کند. به همین دلیل معمولاً گفته می‌شود که پیوند هیدروژنی در نتیجه‌ی نیروهای جاذبه بین بارهای مخالف

1- London

الکتریکی جزئی ایجاد می گردد[۲]. نتیجه‌ی مهم این رویداد این است که یک ماده‌ی خالص دارای پیوند هیدروژنی به صورت مخلوطی از گونه‌های تجمعی است. در این گونه سیستم‌ها، رفتار غیرایده‌الی، هم ناشی از برهمکنش‌های واندروالسی و هم ناشی از تجمع شیمیایی می‌باشد. در نتیجه رفتار فازی و حجمی این مواد به علت برهمکنش‌های قوی بین مولکول‌ها بسیار پیچیده است [۳-۴]. سیالات تجمعی می‌توانند به عنوان مخلوطی از مولکول‌های منفرد و متراکم در نظر گرفته شوند. برهمکنش‌های واندروالسی همچنین روی خواص مخلوط اثر می‌گذارند. آب نمونه‌ای از سیالات تجمعی است که رفتار غیرعادی بسیاری در خواص فیزیکی و شیمیایی آن شناخته شده است. به طوری که بیش از ۶۰ رفتار غیر عادی برای آن ثبت شده است که مربوط به ساختار سه‌بعدی و پیچیده‌ی ناشی از پیوند هیدروژنی می‌باشد [۵]. الکل‌ها، آمین‌ها و کربوکسیلیک اسیدها مثال‌های دیگری از این نوع ترکیبات می‌باشند.

همان‌طور که توضیح داده شد، سیستم‌های تجمعی از جمله سیستم‌های غیرایده‌ال هستند که جهت بررسی آن‌ها نیاز به معادلات خاصی است که اثرات تجمعی در آن وارد شده باشد. این کار می‌تواند به چند صورت انجام شود [۱، ۵ و ۷]:

۱- استفاده از معادلاتی که مبنای تئوری داشته و برای این سیستم‌ها طراحی شده است، به عنوان مثال استفاده از تئوری آماری سیالات تجمعی^۱ (SAFT)، تئوری اختلال، تئوری شیمیایی و غیره.

۲- استفاده از معادلات ساده‌ی سیستم‌های غیرتجمعی که عبارت تجمعی به آن اضافه شده باشد.

۳- استفاده از معادلات حالت با پارامترهای اصلاح شده برای سیستم‌های تجمعی.

۱-۳- دانسیته

نسبت جرم به حجم یک ماده در دما و فشار مشخص دانسیته نامیده می‌شود. واحد دانسیته در سیستم SI، kg m^{-3} است، اما واحدهای g cm^{-3} یا g ml^{-1} نیز ممکن است استفاده شوند. اگر شرایط اندازه‌گیری دانسیته، شرایط ترمودینامیکی نرمال ($T=273/15 \text{ K}$ و $p=101/325 \text{ kPa}$) یا استاندارد ($T=298/15 \text{ K}$ و $P=101/325 \text{ kPa}$) باشد، دانسیته‌ی اندازه‌گیری شده به ترتیب، دانسیته‌ی نرمال و دانسیته‌ی استاندارد نامیده می‌شود.

۱-۴- روش‌های تجربی اندازه‌گیری دانسیته‌ی سیالات

در سال‌های اخیر، برای دستیابی به داده‌های دانسیته‌ی سیالات تجمعی بر حسب دما و فشار از دو روش کلی زیر استفاده شده است:

۱- استفاده از دانسیتومتر ارتعاشی^۱

۲- استفاده از اندازه‌گیری سرعت صوت در دما و فشارهای مختلف

که در ادامه راجع به هریک از این روش‌ها توضیحاتی ارائه می‌شود.

۱-۴-۱- اندازه‌گیری دانسیته با استفاده از دانسیتومتر رزونانسی

این روش بر مبنای تعیین فرکانس رزونانس یک سیستم نوسان کننده قرار دارد. این سیستم نوسان کننده می‌تواند لوله‌ای باشد که از یک سیال پر شده است، یا استوانه یا سیمی باشد که درون سیال مورد آزمایش قرار گرفته است [۹].

دانسیتومتری که معمولاً برای این منظور استفاده می‌شود، دانسیتومتر دیجیتالی رزونانسی لوله‌ای

1- Vibrating Densitometer

آنتون پار^۱ مدل DMA 512P می‌باشد. بخش حساس و نوسان‌کننده‌ی آن لوله‌ی U-شکلی از جنس فولاد ضد زنگ است که حاوی سیال نمونه است. آهن‌ربا و سیم‌پیچی روی لوله تعبیه شده است که عامل ایجاد رزونانس ارتعاشی آن می‌باشد [۱۰].

به منظور اندازه‌گیری دانسیته در دما و فشارهای مختلف، توسط دانسیتومتر ارتعاشی لوله‌ای مراحل زیر انجام می‌شود:

۱- ابتدا دانسیتومتر باید کالیبره شود. برای کالیبراسیون معمولاً از آب، n -هگزان یا نیتروژن به

عنوان سیالات مرجع استفاده می‌شود. مقادیر دانسیته‌ی این سیالات به کمک معادلات حالت

پیشنهاد شده توسط نیسن^۲، یا وانگ و پراس^۳ و یا اسپیم^۴ و همکارانش به‌دست می‌آید.

۲- دانسیتومتر به‌وسیله‌ی پمپ خلأ گاز‌گیری می‌شود.

۳- نمونه با استفاده از سرنگ به درون لوله‌ی U-شکل تزریق می‌شود.

۴- نمونه در شرایط دمایی و فشاری مورد نظر قرار می‌گیرد تا دوره‌ی رزونانس تقریباً ثابت شود،

به‌طوری‌که سیستم به تعادل دمایی و مکانیکی برسد. در این مرحله، دما و فشار سیستم

به‌طور پیوسته کنترل می‌شود.

فشار مورد نیاز توسط تولیدکننده‌ی فشار تولید و کنترل می‌شود. فشار با استفاده از آب یا استون

به عنوان سیال هیدرولیک به نمونه منتقل می‌شود. البته نمونه و سیال دارای فشار بالا توسط

دیافراگم تفلونی از هم جدا می‌شوند که این دیافراگم برای انتقال فشار مناسب می‌باشد. برای

اندازه‌گیری فشار نمونه از فشارسنج‌های مختلفی استفاده می‌شود: مانند فشارسنج HBM^۵، فشارسنج

دیجیتالی روشکا^۶ ۷۶۱۵ و یا فشارسنج ویکا^۷.

1- Anton paar

2- Niesen

3- Wange and pruss

4- Spam

5- Hottinger Boldwin Messtechnik

6- Rushka

7- Wika

دمای نمونه معمولاً توسط حمام در گردش جولابو^۱ F32 کنترل و توسط سنسور پلاتینی مقاوم که متصل به دماسنج ARK F2500 و یا دماسنج آنتون پار CK100 می‌باشد، اندازه‌گیری می‌شود. سپس دانسیته‌ی سیال مورد آزمایش در دما و فشار مورد نظر به کمک معادله‌ی (۱-۱)، محاسبه می‌شود.

$$\rho = K(\tau^2 - \tau_0^2) + \rho_0 \quad (1-1)$$

در این رابطه K پارامتر مشخصه‌ی لوله‌ی U-شکل، ρ دانسیته‌ی سیال مورد آزمایش، ρ_0 دانسیته‌ی سیال مرجع، τ دوره‌ی رزونانس لوله‌ی U-شکل وقتی که با سیال مورد آزمایش پر شده است، و τ_0 دوره رزونانس لوله‌ی U-شکل وقتی که با سیال مرجع پر شده است، می‌باشند [۱۴-۱۱].

۱-۴-۲- اندازه‌گیری دانسیته با استفاده از اندازه‌گیری سرعت صوت در دما و

فشارهای مختلف

در این روش برای اندازه‌گیری دانسیته‌ی سیالات، ابتدا مقادیر سرعت صوت سیال مورد آزمایش در فشار اتمسفر و فشارهای مختلف، توسط دستگاه اندازه‌گیری سرعت صوت، اندازه‌گیری می‌شود. همچنین باید مقادیر تراکم‌پذیری هم‌دما κ_T و دانسیته‌ی سیال را در فشار اتمسفر به روش تجربی اندازه‌گیری نمود.

سپس برای محاسبه‌ی داده‌های دانسیته در فشارهای مختلف از روش اصلاح‌شده‌ی داویس^۲ و گردن^۳ استفاده می‌شود. این روش بر اساس روابطی است که تراکم‌پذیری آدیاباتیک، κ_S ، را به سرعت صوت، c ، و تراکم‌پذیری هم‌دما، κ_T ، را به تراکم‌پذیری آدیاباتیک، κ_S ، مربوط می‌کند (روابط (۲-۱) تا (۴-۱)).

$$\kappa_T = \frac{1}{\rho} \left(\frac{\partial \rho}{\partial P} \right)_T \quad (2-1)$$

1- Julabo
2- Davis
3- Gordon

$$\kappa_S = (\rho c^2)^{-1} \quad (3-1)$$

$$\kappa_T - \kappa_S = T \alpha_p^2 / (\rho C_p) \quad (4-1)$$

که در این رابطه α_p ضریب انبساط پذیری هم فشار و C_p ظرفیت گرمایی هم فشار می باشد. از رابطه ی (۳-۱) برای بیان تغییر دانسیته نسبت به فشار در دمای ثابت به شکل زیر استفاده می شود.

$$(\partial \rho / \partial P) = 1/c^2 + T \alpha_p^2 / C_p \quad (5-1)$$

با انتگرال گیری نسبت به فشار می توان مقادیر دانسیته را بر حسب فشار به دست آورد:

$$\rho(P, T) - \rho(P_0, T) = \int_{P_0}^P 1/c^2 dP + T \int_{P_0}^P (\alpha_p^2 / C_p) dP \quad (6-1)$$

در این رابطه P_0 فشار مرجع می باشد [۱۵ و ۱۶].

۵-۱- روش های پیش بینی دانسیته ی سیالات تجمعی

۵-۱-۱- پیش بینی دانسیته ی سیالات تجمعی با استفاده از معادله ی حالت

از آن جایی که انجام آزمایش های تجربی برای تعیین خواص ترمودینامیکی نیاز به تلاش و هزینه ی زیادی دارد از معادلات حالت برای پیش بینی خواص ترمودینامیکی مواد استفاده می شود.

در سال ۱۹۹۱ ایم^۱ و همکارانش معادله ی حالت ایم- سانگ- میسون^۲ (ISM) را برای پیش بینی

دانسیته ی نرمال الکل ها از متانول تا هپتانول پیشنهاد کردند [۱۷]:

$$P / \rho k_B T = 1 + \frac{\bar{\alpha} p}{1 - \lambda b p} - \frac{(\bar{\alpha} - B_2) \rho}{1 + \delta b p} \quad (7-1)$$

که در این رابطه، p فشار، ρ دانسیته، k ثابت بولتزمن، T دما و λ به عنوان یک پارامتر تنظیم پذیر در معادله ی حالت، می باشد. در این رابطه $\bar{\alpha}$ ضریب تصحیح نیروهای دافعه برای ضریب دوم ویریال، B_2 ، می باشد و b پارامتر وابسته به دما مشابه هم حجم واندروالسی است. همچنین δ

1- Iham

2- Iham-Song-Mosan

تصحیح کوچکی است، که مقدار آن توسط فاکتورهای تجربی به دست می آید، که برای کرات سخت $\delta = 0$ ، برای گازهای نجیب $\delta = 1$ ، و برای گازهای مولکولی $\delta \leq 0.1$ است.

ضابته و همکارانش در سال ۲۰۱۱ با استفاده از تئوری اختلال، مقادیر ترمودینامیکی $\bar{\alpha}$ ، b و B_2 را بر اساس برهم کنش های درون مولکولی به دست آوردند. آن ها برای این کار از پتانسیل لنارد-جونز به عنوان پتانسیل درون مولکولی استفاده کردند. سپس مقادیر این پارامترها را به رابطه ی (۷-۱) وارد کرده و مقادیر دانسیته ی الکلهای از متانول تا هپتانول را با توافق خوبی با مقادیر تجربی پیش بینی کردند [۱۸].

در سال ۲۰۰۵ دکتر پارسافر و همکارانش، معادله ی حالت MLIR^۱ را برای پیش بینی خواص ترمودینامیکی الکلهای نوع اول، نوع دوم و نوع سوم و ۱-کربوکسیلیک اسیدها ارائه دادند. آن ها برای ارائه ی این معادله، هر ترکیب آلی را به صورت یک مخلوط فرضی از گروه های تشکیل دهنده اش در نظر گرفتند، که در این مخلوط پتانسیل برهم کنش هر جفت به صورت متوسط پتانسیل جفت مؤثر^۲ در نظر گرفته شد؛ و برای این منظور از پتانسیل لنارد جونز استفاده کردند. سپس معادله ی حالت (۸-۱) را برای این سیالات ارائه دادند [۱۹].

$$\left(\frac{Z}{n} - 1\right)V^2 = A_m n^2 + B_m n^4 \rho^4 \quad (8-1)$$

که در آن $Z = \frac{P}{\rho RT}$ فاکتور تراکم پذیری، n تعداد گروه های کربنی، ρ دانسیته ی مولی و A_m و B_m پارامترهای معادله حالت MLIR به ازای یک گروه کربنی می باشد. سپس آن ها برای پیش بینی پارامترهای A_m و B_m این سیالات از روش سهم گروه استفاده کردند و با جای گذاری در معادله ی (۸-۱) توانستند دانسیته ی این ترکیبات را در دما و فشارهای مختلف پیش بینی کنند. میانگین درصد خطای پیش بینی دانسیته برای الکلهای و کربوکسیلیک اسیدها به ترتیب کمتر از ۰/۸۶ و ۰/۸۹ گزارش شده است [۱۹].

1- Modified Linear Isotherm Regularity Equation of State

2- Effective pair potential

کلانتر و همکارانش در سال ۲۰۱۳ نیز توسط این روش، دانسیته و دیگر خواص ترمودینامیکی آمین‌های نوع اول را با متوسط خطای کمتر از ۱/۵۴ پیش‌بینی کردند [۲۰].

موسوی در سال ۲۰۱۱ توانست از ترکیب روش سهم گروه با معادله حالت GMA^1 که توسط گوهرشادی و همکارانش ارائه شده بود [۲۱]، دانسیته‌ی الکلهای نوع اول، نوع دوم و نوع سوم و آمین‌های نوع اول و نوع دوم را پیش‌بینی کند [۲۲]. سپس او در سال ۲۰۱۲ این کار را به کربوکسیلیک اسیدها [۲۳] گسترش داد. او نیز در این کار هریک از این ترکیبات را به عنوان یک مخلوط فرضی از متیل، متیلن و یک گروه عاملی تشکیل‌دهنده‌ی آن‌ها در نظر گرفت و سپس معادله‌ی حالت GMA را به این مخلوط فرضی تعمیم داد. رابطه‌ی (۹-۱).

$$\left(\frac{2Z}{n} - 1\right)V_m^3 = A_m n^3 + B_m n^4 \quad (9-1)$$

که در این جا نیز Z فاکتور تراکم‌پذیری، $V_m = \frac{1}{\rho}$ حجم مولی، n تعداد گروه‌های کربنی، A_m و B_m پارامترهای معادله حالت می‌باشد. موسوی با نوآوری رابطه‌ی (۹-۱)، رابطه‌ی ساده شده‌ی (۱۰-۱) را به دست آورد که با قرار دادن پارامترهای A_m و B_m هر یک از این ترکیبات، دانسیته‌ی این ترکیبات را با میانگین خطای ۰/۵۱ پیش‌بینی کرد.

$$B_m n^5 \rho^5 + A_m n^4 \rho^4 + n\rho - \frac{2P}{RT} = 0 \quad (10-1)$$

فصل دوم

کمو متریکس

۲-۱- مقدمه

برای درک مکانیسم فرایندهای مختلف شیمیایی، کشف و توسعه‌ی مواد جدید، حفظ محیط زیست و زمینه‌های دیگر شیمی، هنوز توانایی حل مسائل به‌طور کامل وجود ندارد و برای عملی کردن بعضی از مسائل، نیاز به سیستم‌های بسیار پیچیده‌ای می‌باشد که انجام آن‌ها در گرو صرف هزینه‌های بسیار و مطالعات گسترده می‌باشد. در جهت حل این مشکل، روش‌های محاسباتی کمومتریکس می‌توانند مفید باشند.

۲-۲- کمومتریکس^۱

اصطلاح کمومتریکس اولین بار توسط اسوانت ولد^۲ مطرح گردید. همکاری ولد با بروس آر. کووالسکی^۳ منجر به تأسیس انجمن بین المللی کمومتریکس^۴ (ICS) در سال ۱۹۷۴ گردید. بنا به تعریف ICS، کمومتریکس عبارت است از کاربرد روش‌های ریاضی، آماری برای برقراری ارتباط بین سنجش‌های انجام شده روی یک سیستم یا فرایند شیمیایی به منظور درک بهتر اطلاعات شیمیایی. طبق یکی از جامع‌ترین تعاریف، کمومتریکس شاخه‌ای از شیمی است که از ریاضی و آمار و منطق برای رسیدن به اهداف زیر استفاده می‌کند:

الف) فرایندهای تجربی بهینه را طراحی و انتخاب می‌کند.

ب) حداکثر اطلاعات شیمیایی قابل حصول را از تحلیل داده‌های شیمیایی فراهم می‌کند.

ج) می‌تواند اطلاعات بیشتری درمورد سیستم‌های شیمیایی به دست آورد [۲۴].

کمومتریکس به عنوان یک شاخه‌ی علمی جوان در دو دهه‌ی اخیر به سرعت توسعه پیدا کرده است. این رشد سریع مدیون پیشرفت سریع دستگاه‌های هوشمند و خودکار آزمایشگاهی و همچنین

1- Chemometrics

2- Svante Wold

3- Bruce R. Kowaski

4- International Chemometrics Society

امکان استفاده از کامپیوترهای قدرتمند و نرم‌افزارهای ساده است. بنابراین امروزه کمومتریکس به عنوان ابزاری جدید در همه‌ی شاخه‌های شیمی مورد استفاده قرار می‌گیرد و هر شیمیدان در تحقیقات علمی‌اش به استفاده از این روش‌های آماری و ریاضی نیاز دارد.

۲-۳- رابطه‌ی کمی ساختار- خاصیت^۱ یا ساختار- فعالیت^۲

از کاربردهای مهم کمومتریکس می‌توان به رابطه‌ی کمی ساختار- خاصیت (QSPR) و رابطه‌ی کمی ساختار- فعالیت (QSAR) اشاره کرد. QSPR ارتباط ساختار با خاصیت را مورد بررسی قرار می‌دهد، ولی QSAR مطالعاتی است که فعالیت‌های بیولوژیکی ترکیبات را به ویژگی‌های ساختاری آن‌ها ارتباط می‌دهد. این مطالعات در واقع سعی در پیدا کردن رابطه‌ای هماهنگ میان فعالیت بیولوژیکی یا خاصیت مواد با ویژگی‌های مولکولی دارد، تا بتواند این قواعد را برای ارزیابی فعالیت یا خاصیت ترکیبات جدید به کار ببرد.

از جمله روش‌هایی که به طور وسیع در مطالعات مدل‌سازی ارتباط کمی ساختار-خاصیت استفاده می‌شوند، می‌توان از روش‌های خطی نام برد که مدل‌های خطی ساده و قابل تفسیری را می‌سازند، مانند روش حداقل مربعات متداول^۳ (OLS)، رگرسیون اجزاء اصلی^۴ (PCR)، تحلیل مولفه‌های اساسی (PCA)^۵، رگرسیون خطی چندگانه^۶ (MLR) و حداقل مربعات جزئی^۷ (PLS). دیگر روش‌ها مانند شبکه‌ی عصبی^۸ (ANN)، ماشین‌های بردار پشتیبان^۹ (SVM) و الگوریتم ژنتیک^{۱۰} (GA)

1- Quantitative Structure- Property Relationship

2- Quantitative Structure- Activity Relationship

3- Ordinary Least Squares

4- Principal Component Regression

5- Principle Component Analysis

6- Multiple Linear Regression

7- Partial Least Squares

8- Artificial Neural Networks

9- Support Vector Machines

10- Genetic Algorithm

مدلهایی غیرخطی بین فعالیت یا خاصیت یا ویژگی‌های مولکولی می‌سازند. مراحل کلی یک مدل-سازی QSPR یا QSAR در جدول (۱-۲) نشان داده شده است [۲۵].

جدول (۱-۲) - مراحل کلی یک مدل‌سازی QSPR یا QSAR

➤ فراهم کردن سری داده‌ها
➤ انتخاب و محاسبه توصیف‌کننده‌ها
➤ ارزیابی و تجزیه و تحلیل آماری توصیف‌کننده‌ها
➤ تقسیم‌بندی سری داده‌ها به سری‌های آموزش و ارزیابی و تست
➤ آنالیز آماری مدل‌ها و انتخاب مدل مناسب
➤ ارزیابی اعتبار مدل‌های انتخاب شده
➤ پیش‌بینی خصوصیت موردنظر

۲-۳-۱- فراهم کردن سری داده‌ها

اولین مرحله، انتخاب یک سری داده‌ی مناسب و معتبر می‌باشد. هر چه سری داده‌ها وسیع‌تر و متنوع‌تر باشد، مدل به دست آمده مدل کلی‌تر و معتبرتری بوده و محدود به یک گروه خاص از ترکیبات نخواهد بود.

۲-۳-۲- رسم ساختار و بهینه‌سازی

در این مرحله از مطالعه، ابتدا ساختار مولکولی هر ترکیب در نرم‌افزار HyperChem 7.0 ترسیم و سپس بهینه‌سازی می‌شود.

۲-۳-۳- محاسبه‌ی توصیف‌کننده‌ها

توصیف‌کننده‌ها پارامترهای عددی هستند که ویژگی‌های متفاوتی از یک مولکول را بیان می‌کنند. تعیین و انتخاب توصیف‌کننده‌ها از اهمیت بسزایی برخوردار است. در صورتی‌که توصیف‌کننده‌های انتخاب شده مناسب نباشند، عمل مدل‌سازی با مشکل مواجه شده و مدل مناسب به دست نخواهد آمد.

در این تحقیق ۱۴۸۱ توصیف‌کننده‌ی ساختاری توسط نرم‌افزار Dragon محاسبه شد که این توصیف‌کننده‌ها از نظر بیان چگونگی خصوصیات مولکول به ۱۸ دسته متفاوت تقسیم می‌شوند، که در جدول (۲-۲) نام این توصیف‌کننده‌ها آورده شده است.

۲-۳-۴- حذف توصیف‌کننده‌های نامناسب

پس از آن که توصیف‌کننده‌ها محاسبه شدند، باید توصیف‌کننده‌هایی که اطلاعات کمتری در مورد تغییرات خواص شیمیایی و یا بیولوژیکی مواد مورد نظر دارند، حذف شوند. برای این منظور لازم است که از فرایند کاهش متغیر استفاده گردد. در این فرایند توصیف‌کننده‌هایی حذف می‌شوند که معمولاً یک یا چند ویژگی زیر را داشته باشند:

- ۱- توصیف‌کننده‌هایی که دارای بیش از ۹۰٪ مقادیر یکسان باشند.
- ۲- توصیف‌کننده‌هایی که با توصیف‌کننده‌های دیگر همبستگی بیش از ۰/۹ دارند. برای این منظور می‌توان از نرم‌افزار SPSS استفاده کرد. اما انجام این کار با استفاده از این نرم‌افزار کاری بسیار وقت‌گیر است. اخیراً با استفاده از برنامه‌ی نوشته شده در محیط نرم افزار MATLAB، همبستگی توصیف‌کننده‌ها بررسی شده و توصیف‌کننده‌های حذفی مشخص می‌گردند.
- ۳- توصیف‌کننده‌هایی که با متغیر وابسته، همبستگی کمی دارند و ارتباط آن‌ها با خاصیت موردنظر مشکل است.

پس از حذف توصیف‌کننده‌های نامناسب، تعداد توصیف‌کننده‌ها و به دنبال آن پیچیدگی محاسبات به میزان زیادی کاهش می‌یابد.

جدول (۲-۲) - انواع توصیف‌کننده‌های محاسبه شده توسط نرم افزار Dragon

۱	توصیف‌کننده‌های زیرساختاری ^۱	۲	توصیف‌کننده‌های توپولوژی ^۲
۳	شمارنده‌ی مولکولی مورس ^۳	۴	توصیف‌کننده‌های BCUT ^۴
۵	شاخص بار ^۵	۶	خود ارتباطی‌های دو بعدی ^۶
۷	توصیف‌کننده‌ای بار ^۷	۸	شاخص‌های آروماتیسیت ^۸
۹	پروفایل‌های مولکولی راندیک ^۹	۱۰	توصیف‌کننده‌های هندسی ^{۱۰}
۱۱	توصیف‌کننده‌های RDF ^{۱۱}	۱۲	توصیف‌کننده‌های سه بعدی ^{۱۲}
۱۳	پروفایل‌های مولکولی راندیک ^{۱۳}	۱۴	توصیف‌کننده‌های GETAWAY ^{۱۴}
۱۵	گروه‌های عاملی ^{۱۵}	۱۶	اجزای میان اتمی ^{۱۶}
۱۷	توصیف‌کننده‌های تجربی ^{۱۷}	۱۸	خصوصیات مولکولی ^{۱۸}

-
- 1- Constitutional Descriptors
 - 2- Topological Descriptors
 - 3 - Molecular Walk Counts
 - 4 - BCUT Descriptors
 - 5 - Charge Indices
 - 6 - 2D Autocorrelations
 - 7 - Charge Descriptors
 - 8 - Aromatic Indices
 - 9 - Rancid Molecular Profiles
 - 10 - Geometrical Descriptors
 - 11 - Radial Distribution Function Descriptors
 - 12 - 3D Molecule Representation of Structure Backed on Electron Diffraction Descriptors(3D-MORSE Descriptors)
 - 13 - Weighted Holistic Invariant Molecular Descriptors
 - 14 - Geometry, Topology and Atom Weighted Assembly
 - 15 - Functional Groups
 - 16 - Atom- Central Fragments
 - 17 - Empirical Descriptors
 - 18 - Properties

۲-۳-۵- دسته‌بندی داده‌ها

قبل از ساختن مدل، سری داده‌ها به سه زیر مجموعه تقسیم می‌شوند. اولین زیر مجموعه، سری آموزش نام دارد که از آن برای ساخت مدل‌های خطی و غیرخطی مختلف استفاده می‌شود. دومین زیر مجموعه سری ارزیابی است که مدل‌های به دست آمده از سری آموزش توسط آن ارزیابی شده و به این ترتیب بهترین مدل انتخاب می‌شود. زیر مجموعه‌ی سوم سری تست یا آزمایش نام دارد که در طول فرایند مدل‌سازی دخالتی نداشته و از آن برای مقایسه‌ی مدل‌های متفاوت استفاده می‌شود. برای تقسیم‌بندی داده‌ها می‌توان از روش تقسیم تصادفی استفاده کرد و یا برای حصول نتیجه‌ی بهتر، از روش تحلیل مولفه‌های اساسی (PCA) بهره گرفت. روش PCA یکی از با ارزش‌ترین نتایج کاربرد جبرخطی است که به وفور در کلیه‌ی اشکال تحلیلی از علوم شبکه‌ی عصبی تا نمودارهای کامپیوتری استفاده شده است، به این دلیل که یک روش آسان و غیرپارامتریک برای استخراج اطلاعات مرتبط از یک مجموعه داده‌ی پیچیده است. اگر در هر مجموعه داده n متغیر وجود داشته باشد، هر متغیر می‌تواند دارای چند بعد باشد. با توجه به این‌که اغلب درک و شهود فضای چند بعدی دشوار است، روش تحلیل مولفه‌های اساسی ابعاد کلیه‌ی مشاهدات را براساس شاخص ترکیبی و دسته‌بندی مشاهدات مشابه کاهش می‌دهد. به طور کلی کاربرد عمده‌ی روش تحلیل مولفه‌های اساسی عبارت است از کاهش تعداد متغیرها و یافتن ساختار ارتباطی بین آن‌ها که در حقیقت همان دسته‌بندی متغیرها است.

برای انجام کار باید توسط نرم‌افزار SPSS و در نوار ابزار گزینه‌ی Analyze، ابتدا گزینه‌ی Dimension Reduction و سپس گزینه‌ی Factor را انتخاب نمود تا مولفه‌های اساسی برای مجموعه‌ی توصیف‌کننده‌ها به دست آید. معمولاً دو مولفه‌ی اول مقدار قابل توجهی از پراکندگی داده‌ها را در نظر می‌گیرد. بنابراین نموداری از مولفه‌ی اول بر حسب مولفه‌ی دوم ترسیم می‌شود. شکل این نمودار پراکندگی داده‌ها را نشان می‌دهد. براساس این نمودار نقطه داده‌های محیطی در

سری آموزش قرار می‌گیرند و سری ارزیابی و تست از داده‌های داخلی به طور تصادفی انتخاب می‌شوند.

۲-۳-۶- انتخاب بهترین توصیف‌کننده‌ها

با وجودی که با به کار بردن روش‌های کاهش متغیر، تعداد توصیف‌کننده‌ها به میزان زیادی کاهش می‌یابد، اما هنوز تعداد آن‌ها برای ساخت یک مدل مناسب زیاد است. بنابراین باید به ترتیبی، بهترین توصیف‌کننده‌ها را از میان توصیف‌کننده‌های باقی‌مانده، انتخاب نمود. از جمله‌ی این روش‌ها می‌توان به روش رگرسیون مرحله‌ای، الگوریتم ژنتیک اشاره کرد. در ادامه به اختصار راجع به هر کدام از این روش‌ها توضیحاتی ارائه می‌گردد.

۲-۳-۶-۱- انتخاب بهترین توصیف‌کننده‌ها توسط رگرسیون مرحله‌ای (SR)

یکی از متداول‌ترین و ساده‌ترین روش‌ها برای انتخاب متغیرهای مناسب در QSPR، استفاده از روش رگرسیون خطی مرحله‌ای می‌باشد. در این روش خصوصیت شیمیایی و یا بیولوژیکی مورد نظر به عنوان متغیر وابسته و توصیف‌کننده‌ها به عنوان متغیرهای مستقل در نظر گرفته می‌شوند. در روش مرحله‌ای، ابتدا با توجه به ضریب همبستگی، متغیر اولیه انتخاب می‌شود. سپس مدل دو متغیری ایجاد شده و پارامترهای آماری مدل حاصل برای سری ارزیابی مورد بررسی قرار می‌گیرند. در صورتی که مدل از نظر آماری بی‌معنا باشد، متغیر بکار گرفته شده در آن مدل حذف می‌گردد. سپس مدل سه متغیری ساخته می‌شود و دوباره پارامترهای آماری آن برای سری ارزیابی مورد بررسی قرار می‌گیرد. این کار تا جایی ادامه می‌یابد که تعداد توصیف‌کننده‌های مناسب انتخاب شوند.

به این ترتیب کلیه‌ی مدل‌های ممکن ساخته شده و برای هر یک از مدل‌ها پارامترهای آماری شامل ضریب همبستگی، مقدار میانگین مربعات خطا و آماره‌ی F تعیین می‌شود. مدلی که ضریب

همبستگی و آماره‌ی F آن بیشتر و مقدار میانگین مربعات خطای آن کمتر باشد انتخاب می‌شود و به این وسیله توصیف‌کننده‌های بهینه به روش رگرسیون مرحله‌ای به دست می‌آیند.

۲-۳-۶-۲- انتخاب بهترین توصیف‌کننده‌ها توسط الگوریتم ژنتیک

۲-۳-۶-۲- الف - مقدمه

امروزه روش‌های کلاسیک با بزرگ‌تر و پیچیده‌تر شدن مسائل و اهمیت یافتن سرعت رسیدن به پاسخ، دیگر جوابگوی حل بسیاری از مسائل نیستند، چرا که فضای جستجو با افزایش بعد مسئله به صورت نمایی زیاد می‌شود و این روش‌ها به دلیل محدودیت‌های هزینه‌ای، مقرون به صرفه نیستند. به همین دلیل در سال‌های اخیر استفاده از الگوریتم‌های جستجوی تصادفی مثل الگوریتم‌های تکاملی^۱ و الگوریتم‌های جستجوی ابتکاری^۲ به جای جستجوی همه جانبه مورد توجه قرار گرفته‌اند.

در سال ۱۹۷۵ جان هولند^۳ نظریه‌ی بنیادی ژنتیک را مطرح کرد. ایده‌ی اساسی این الگوریتم انتقال خصوصیات موروثی توسط ژن‌هاست. فرض کنید مجموعه خصوصیات انسان توسط کروموزوم‌های او به نسل بعدی منتقل می‌شوند. هر ژن در این کروموزوم‌ها نماینده‌ی یک خصوصیت است. حال اگر این کروموزوم کاملاً، به نسل بعد انتقال یابد، تمامی خصوصیات نسل بعدی شبیه به خصوصیات نسل قبل خواهد بود. بدیهی است که در عمل چنین اتفاقی رخ نمی‌دهد زیرا کروموزوم‌ها دستخوش تغییراتی از قبیل جهش و همببری می‌شوند [۲۶].

1- Evolutionary Algorithm
2- Heuristic
3- John Holland

۲-۳-۶-۲-ب- الگوریتم ژنتیک

الگوریتم ژنتیک (GA) یک تکنیک جستجو در علم رایانه برای یافتن راه حل بهینه و مسائل جستجو است. الگوریتم‌های ژنتیک یکی از انواع الگوریتم‌های تکاملی هستند که از علم زیست‌شناسی مثل وراثت، جهش، انتخاب ناگهانی، انتخاب طبیعی و ترکیب الهام گرفته‌اند.

در الگوریتم ژنتیک (GA)، متغیرهای مسئله، خصوصیات فردی فرض می‌شوند. این خصوصیات توسط ژن‌ها به صورت رمز درآمده‌اند و با کنار هم گذاشتن آن‌ها از طریق ایجاد یک جمعیت تصادفی، فرد بهینه و یا جمعیت بهینه با خصوصیات برتر جستجو می‌شود. بقای افراد با خصوصیات برتر یا ضعیف‌تر، به انتخاب طبیعی واگذار می‌شوند. به عبارت دیگر بقای افراد یک پدیده‌ی احتمالی، اما وابسته به میزان شایستگی آن‌ها می‌باشد.

مهم‌ترین و پرکاربردترین واژه‌هایی که در الگوریتم ژنتیک به کار برده می‌شوند، به صورت خلاصه در ادامه توضیح داده شده‌اند.

الف) کروموزوم: در الگوریتم ژنتیکی، هر فرد تنها یک کروموزوم دارد و هر کروموزوم به معنای یک راه حل مسئله به صورت رمزی شده می‌باشد. کروموزوم‌ها از ترکیب مقادیر مختلف متغیرها یا ژن‌ها (هر ژن مبین یک متغیر است) به وجود می‌آیند [۲۷ و ۲۸].

ب) جمعیت: به تعداد مشخصی از کروموزوم‌ها که در کنار هم قرار می‌گیرند، جمعیت گفته می‌شود و با نماد N معرفی می‌گردد. هر جمعیت توسط عملگرهای وراثتی (انتخاب، همبندی و جهش) جمعیت بعدی را می‌سازد.

ج) تابع شایستگی^۱: برتری کروموزوم‌ها (افراد) نسبت به یکدیگر با توجه به یک معیار سنجیده می‌شود. این معیار تابع شایستگی است.

1- Fitness function

د) تولید مثل: پس از آن که تمام افراد جمعیت توسط تابع شایستگی ارزیابی شدند، شایسته‌ترین آن‌ها طبق قوانین احتمالات انتخاب، و با یکدیگر تولید مثل می‌کنند. این عمل نیازمند عملگرهای وراثتی می‌باشد.

ه) عملگر انتخاب: انتخاب والدین جهت تولیدمثل و همچنین تعیین این که والدین چندبار حق زاد و ولد دارند، به کمک عملگر انتخاب که براساس اصل انتخاب طبیعی داروین عمل می‌کند، صورت می‌پذیرد. عملگر انتخاب با توجه به شایستگی کروموزوم‌ها، بطور تصادفی یک کروموزوم را انتخاب و یک کپی از آن را به حوضچه‌ی ازدواج^۱ منتقل می‌کند. کروموزوم انتخاب شده از جمعیت اولیه، حذف نمی‌شود و شانس انتخاب مجدد را دارد. عمل انتخاب و انتقال کروموزوم به تعداد جمعیت اولیه‌ی کروموزوم‌ها، انجام می‌شود. به عبارت دیگر N دفعه عمل انتخاب والد انجام شده و N والد در حوضچه‌ی ازدواج پدید می‌آید. ممکن است کروموزوم بیش از یک بار برای تولید مثل انتخاب شود. در این صورت حداقل تعداد کروموزومی که برای تولید مثل انتخاب نشده‌اند و از دور خارج شده و نسل و خصوصیاتش از بین رفته است، برابر با تعداد دفعات انتخاب آن کروموزوم منهای یک خواهد بود. جهت پیاده سازی این عملگر، معمولاً از تکنیک چرخ گردان^۲ استفاده می‌شود که در آن عنصری که عدد برازش (تناسب) بیشتری داشته باشد، انتخاب می‌شود.

و) عملگر همبری: همبری که در سلول‌های جنسی والدین رخ می‌دهد به معنی چسبیدن دو کروموزوم از طول به یکدیگر و تبادل برخی قطعات بین دو کروموزوم است. به عبارت دیگر، در همبری، دو کروموزومی که یکسان نیستند از محل‌هایی معین شکسته و سپس تکه‌های آن‌ها جابجا می‌شود. این عملگر ابتدا دو والد را بطور تصادفی از حوضچه‌ی ازدواج انتخاب می‌کند. سپس با یک احتمال معین که احتمال همبری^۳ (P_C) نامیده می‌شود، عمل همبری بین دو والد انتخاب شده را انجام می‌دهد و فرزندان نسل بعد را به وجود می‌آورد. احتمال این که عمل همبری انجام نشود

1- Mating pool

2- Roulette wheel

3- Probability of cross-over

(1- P_C) می‌باشد. اگر عمل همبری انجام نشود خود والدین به نسل بعدی منتقل می‌شوند. همبری انواع متفاوتی دارد که ساده‌ترین آن‌ها همبری تک نقطه‌ای^۱ بین دو والد است. نقطه‌ی برش^۲ در عمل همبری به صورت تصادفی تعیین می‌شود و فرزندی که پس از این عمل ایجاد می‌شوند برخی از خصوصیات هر دو والد را دارا هستند.

ز) عملگر جهش: در کروموزوم‌های فرزند بصورت کاملاً اتفاقی خصوصیت جزئی به منظور افزایش سازگاری با محیط صورت می‌پذیرد. این کار توسط عملگر جهش انجام می‌شود. احتمال جهش^۳ (P_m) در الگوریتم ژنتیکی بسیار پایین است (اغلب بین ۰/۰۱ تا ۰/۰۰۱). نحوه‌ی اجرای جهش به این صورت است که یک بیت از کروموزوم به صورت تصادفی انتخاب می‌شود. سپس مقدار این بیت اگر ۱ باشد با احتمال P_m به صفر و اگر صفر باشد با احتمال P_m به ۱ تغییر داده می‌شود. به این ترتیب تعداد کل بیت‌های جهش یافته در یک جمعیت N کروموزومی که طول هر کروموزوم آن L می‌باشد، برابر $P_m * N * L$ خواهد بود. مهم‌ترین مزیت جهش این است که خصوصیات را که در هیچ کدام از والدین وجود نداشته است، در فرزندان به‌وجود می‌آورد. اما اگر میزان جهش زیاد شود، احتمال از بین رفتن کروموزوم‌های شایسته پیش می‌آید. همچنین جهش به شدت در همگرایی الگوریتم ژنتیکی تأثیر دارد. به گونه‌ای که اگر احتمال جهش خیلی زیاد باشد، همگرایی از بین می‌رود و اگر خیلی کم باشد، سبب همگرایی زودرس می‌شود. بنابراین احتمال جهش باید متعادل و به گونه‌ای باشد که همگرایی مطلوبی را به‌وجود آورد [۲۹ و ۳۰].

ح) همگرایی^۴: پیشرفت به سوی یکنواختی را همگرایی گویند. در الگوریتم ژنتیکی، نسل به نسل جمعیت شایسته‌تر شده و جواب‌ها به سمت بهینه‌ی فرا محلی همگرا می‌شوند. زمانی که شایستگی افراد جمعیت و متوسط شایستگی جمعیت زیاد شود، شایستگی کروموزوم‌های جمعیت تقریباً یکسان و یکنواخت می‌شود و اصطلاحاً گفته می‌شود که الگوریتم همگرا شده است [۲۷ و ۲۸].

1- Single point cross-over
2- Cut point
3- Probability of mutation
4- convergence

۲-۳-۶-۲-ج: ساختار الگوریتم ژنتیکی

ساختار الگوریتم ژنتیک به صورت مراحل زیر می باشد:

۱- در الگوریتم ژنتیک متغیرهای طراحی را به صورت رشته‌هایی با طول معین که به صورت صفر و یک رمزگذاری^۱ شده‌اند، در نظر می گیرند که اولین گام الگوریتم ژنتیک می باشد. هریک از این رشته‌ها همانند کروموزوم‌هاست و رمزهای صفر و ۱ مشابه ژن‌ها در یک موجود زنده می باشد. فنون ژنتیکی سرشت تکراری دارند که با تعدادی طرح از پیش تعیین شده کار می کند. تکرار هر گام یک نسل^۲ نامیده می شود.

۲- جمعیت اولیه‌ای از کروموزوم‌ها با توجه به فضای مسئله به صورت کاملاً تصادفی ایجاد می گردد.

۳- کروموزوم‌ها رمزگشایی می شوند؛ یعنی اطلاعات دودویی ژن‌ها به اطلاعات ده‌دهی تبدیل می شود.

۴- هر کروموزوم براساس تابع شایستگی ارزیابی می گردد و رشته‌ی مطلوب‌تر برای نسل بعد انتخاب می شود.

۵- توسط عملگرهایی همچون پیوند، جهش و همبری نسل جدیدی با صلاحیت بالاتر تولید می گردد. با تکرار نسل‌ها رشته‌ها بهبود یافته و سرانجام جمعیت مطلوب که همان توصیف‌کننده‌ها هستند حاصل می گردد.

۲-۳-۶-۳- انتخاب توصیف‌کننده‌ها بر مبنای روش سهم گروه^۳ (GCM)

روش سهم گروه روشی ساده، کارآمد و در عین حال معتبر است که به طور گسترده‌ای در پیش‌بینی خواص فیزیکی و ترمودینامیکی مواد به کار می رود. این روش در واقع ارتباطی بین خاصیت یک ماده با ساختار شیمیایی آن برقرار می کند. مبنای به وجود آمدن روش سهم گروه‌ها این است که

1- Coding
2- Generation
3- Group Contribution Method

اگر چه هزاران ترکیب شیمیایی در دنیای علم و تکنولوژی وجود دارد اما تعداد گروه‌های ساختاری و عاملی سازنده‌ی همه‌ی این ترکیبات، بسیار کمتر از تعداد این ترکیبات می‌باشد. با این ایده، روش سهم گروه‌ها فرض می‌کند که خاصیت فیزیکی یا ترمودینامیکی هر ماده را می‌توان از سهم گروه‌های موجود در مولکول آن ماده در خاصیت موردنظر، به دست آورد. به این ترتیب یک روش عملی برای پیش‌بینی خواص تعداد زیادی ماده برحسب تعداد کمتری پارامتر به دست می‌آید که این پارامترها، سهم گروه‌های ساختاری را در خاصیت موردنظر مشخص می‌کنند.

لازم به توضیح است که تعریف گروه‌های تشکیل‌دهنده‌ی هر ماده، یک مقوله‌ی اختیاری است. به طور مثال ممکن است یک نفر پیشنهاد کند که باید اتم‌های اصلی شناسایی شوند. در حالی که ما می‌دانیم خواص کربن در الماس نسبت به کربن در گرافیت بسیار متفاوت است. از طرف دیگر، تجربه نشان داده است که اگر چه گروه کربونیل موجود در کتون‌ها نسبت به گروه کربونیل موجود در کربوکسیلیک اسیدها خواص متفاوتی دارد اما حداقل گروه کربونیل موجود در همه‌ی کتون‌ها مشابه می‌باشد. بدیهی است هر چه تمایز بین گروه‌ها بیشتر شود، صحت روش سهم گروه‌ها افزایش می‌یابد تا جایی که سرانجام هر ترکیب فقط شامل گروه‌های مربوط به خودش باشد. اما می‌توان با انتخاب صحیح گروه‌ها به روش سودمندی دست یافت که با وجود تعداد گروه‌های کم، صحت خواص پیش‌بینی شده قابل قبول باشد [۳۱].

در مطالعات QSPR که به طور گسترده‌ای برای مدل‌سازی خواص فیزیکی و شیمیایی مواد به کار می‌رود، خاصیت مورد نظر به توصیف‌کننده‌های مولکولی مرتبط می‌شود. توصیف‌کننده‌های مولکولی مشخصه‌های عددی ماده‌ی خالص هستند که مستقیماً از روی ساختار بهینه شده‌ی مولکول و با استفاده از الگوریتم‌ها و نرم‌افزارهای خاص محاسبه می‌شوند. اما مهم‌ترین عیب این روش، مسیر طولانی و پیچیده‌ی محاسبه‌ی توصیف‌کننده از روی ساختار مولکولی و سپس انتخاب بهترین آن‌هاست. در نتیجه مدل‌های حاصل از این مطالعات برای استفاده، چندان ساده نیستند. به همین دلیل در سال‌های اخیر استفاده از روش ساده و کارآمد سهم گروه‌ها برای انتخاب توصیف‌کننده‌های

مناسب جهت مطالعات QSPR پیشنهاد شده است. در این روش تعداد گروه‌های تشکیل‌دهنده‌ی هر ترکیب به عنوان توصیف‌کننده‌های آن ماده انتخاب می‌شود و در صورتی که یکی از گروه‌های تشکیل‌دهنده در ترکیبات مورد مطالعه وجود نداشته باشد توصیف‌کننده‌ی مرتبط با آن گروه، صفر در نظر گرفته می‌شود.

در طی چند سال اخیر از شبکه‌ی عصبی مصنوعی بر مبنای روش سهم گروه‌ها در پیش‌بینی خواص مختلفی از جمله برای پیش‌بینی نقطه‌ی اشتعال آلکان‌ها [۳۲]، دانسیته‌ی مایعات یونی [۳۳]، خصوصیات نقطه‌ی بحرانی تعدادی از هیدروکربن‌ها [۳۴]، هدایت یونی و ویسکوزیته‌ی مایعات یونی [۳۵] و دانسیته‌ی هیدروکربن‌ها [۳۶] استفاده شده است.

۲-۳-۷- ساختن مدل

در این مرحله با استفاده از روش‌های آماری مختلف، مدل‌سازی انجام می‌شود. در واقع مدل یک معادله‌ی ریاضی است که بیانگر رابطه‌ی بین متغیر وابسته و متغیرهای مستقل (توصیف‌کننده‌ها) می‌باشد. با داشتن مقادیر متغیرهای مستقل وارد شده در مدل، می‌توان مقدار متغیر وابسته را محاسبه کرد. در مطالعات رابطه‌ی کمی ساختار-خاصیت، مدل‌سازی به دو صورت خطی و یا غیرخطی انجام می‌شود. در مدل خطی با در نظر گرفتن این‌که رابطه‌ی بین توصیف‌کننده‌ها و متغیر وابسته، یک رابطه‌ی خطی است، مدل‌سازی صورت می‌گیرد. زمانی که مدل‌های خطی نتایج مطلوبی ارائه ندهند و یا روابط غیر خطی بین توصیف‌کننده‌ها و متغیر وابسته وجود داشته باشد، از مدل غیرخطی استفاده می‌شود. از جمله روش‌های خطی می‌توان از رگرسیون خطی چندگانه (MLR)، تحلیل مولفه‌های اساسی (PCR) و حداقل مربعات جزئی (PLS) و از روش‌های غیرخطی، از شبکه‌ی عصبی مصنوعی (ANN)، الگوریتم ژنتیک (GA) و ماشین بردار پشتیبان (SVM) نام برد.

چون در این پایان‌نامه از روش غیرخطی ANN برای مدل‌سازی استفاده شده است، لذا در ادامه فقط راجع به این روش توضیح داده خواهد شد.

۲-۳-۷-۱- شبکه‌ی عصبی مصنوعی

دانشمندان با بهره‌گیری از یافته‌های علمی درباره‌ی اصول و ساز و کارهای محاسباتی مغز انسان که عملکرد بسیار سریع و در عین حال دقیقی دارد، شبکه‌های عصبی مصنوعی (ANN) را به عنوان مدل بسیار ساده و ابتدایی از مغز، ارائه کردند. همان‌طور که مغز انسان می‌تواند با استفاده از تجربیات قبلی، مسائل جدید را تجزیه و تحلیل کند، شبکه‌های عصبی مصنوعی نیز در صورت آموزش قادر به ارائه‌ی جواب مناسب برای مسائل جدیدی، که قبلاً با آن روبرو نشده‌اند، می‌باشند [۳۷]. شبکه‌های عصبی مصنوعی جزء دسته‌ای از سیستم‌های دینامیکی هستند که با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای آن‌ها را کشف می‌کنند. به همین دلیل به این سیستم‌ها هوشمند^۱ گفته می‌شود زیرا بر اساس محاسبات روی داده‌های عددی یا مثال‌ها قوانین کلی را فرا می‌گیرند [۳۸].

۲-۳-۷-۱-الف- نرون زیستی

مغز به عنوان یک سیستم پردازش اطلاعات با ساختاری موازی، از حدود یک صد میلیارد سلول عصبی به نام نرون تشکیل شده که هر نرون به ده هزار نرون دیگر مرتبط می‌باشند. اکثر نرون‌ها از سه قسمت اساسی تشکیل شده‌اند:

۱- دندریت‌ها گیرنده‌های نرون هستند و سیگنال‌های الکتریکی را به قسمت بدنه‌ی سلول منتقل می‌کنند.

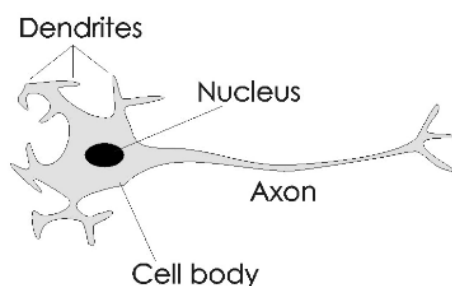
۲- بدنه‌ی سلول که به آن سوما^۲ نیز می‌گویند، انجام عملیات لازم بر روی اطلاعات ورودی را بر عهده دارد.

۳- آکسون رشته‌ای طولانی است که اطلاعات خروجی از سوما را دریافت کرده و به خارج نرون می‌فرستد.

1- Intelligent

2- Soma

محل اتصال آکسون یک سلول به دندریتهای سلولهای دیگر را سیناپس^۱ می‌گویند. پیامهای عصبی تنها به صورت یک طرفه از دندریته به بدنه سلول و از آنجا به آکسون حرکت می‌کنند. وقتی پیام به انتهای آکسون می‌رسد فیبرهای ماهیچه‌ای و یا نرونهای دیگر که به آن متصل هستند تحریک می‌شوند [۳۹]. شکل (۱-۲) ساختار یک نرون زیستی را نشان می‌دهد.



شکل (۱-۲) - ساختار یک نرون طبیعی

۲-۳-۷-۱-ب- نرون محاسباتی

شبکه‌های عصبی مصنوعی نیز مشابه مغز از تعداد زیادی نرون تشکیل شده‌اند که از طریق ارتباطاتی به هم متصل شده‌اند. نرون از تعدادی ورودی، یک خروجی، وزنهای مربوط به ایجاد سیگنال تحریک و تابع تبدیل^۲ یا انتقال تشکیل شده است. ارتباط بین خروجی و سیگنال تحریک در قسمتی به نام تابع انتقال به وجود می‌آید. بنابراین یک نرون از دو بخش تشکیل شده است.

بخش اول: وظیفه‌ی دریافت ورودی‌ها، وزندهی به آنها و محاسبه‌ی مجموع آنها به عنوان سیگنال تحریک را به عهده دارد.

بخش دوم: وظیفه‌ی اعمال تابع انتقال روی سیگنال تحریک و ساختن پیغام خروجی نرون به لایه‌ی بعد و یا خروجی شبکه را به عهده دارد.

1- Synapse
2- Transfer function

شکل (۲-۲) ساختمان یک نرون محاسباتی را نشان می‌دهد. بیان رابط‌های ورودی‌ها و خروجی‌های

شبکه در یک مدل ریاضی مطابق فرمولاسیون شکل (۲-۲) پروسپترون نامیده می‌شود.

با توجه به شکل، علاوه بر تمام ورودی‌های وزن‌دار، ورودی دیگری نیز در نظر گرفته شده و آن

عدد یک است. این عدد یک در بایاس^۱ ضرب شده و با مجموع حاصل ضرب ورودی‌ها در وزن‌ها جمع می‌شود تا ورودی خالص را ایجاد کند.

$$n = WX + b = W_{i,1}X_1 + W_{i,2}X_2 + \dots W_{i,n}X_n + b \quad (۲-۲)$$

سپس تابع محرک روی این ورودی خالص n اثر می‌کند و خروجی نرون را ایجاد می‌کند.

$$Y = f(n) = f(WX + b) \quad (۳-۲)$$

بایاس شبیه یک وزن است، با این تفاوت که روی ورودی ثابت یک، اعمال می‌شود و مزیت آن

افزودن یک متغیر اضافی به شبکه است. از این رو شبکه‌ای که دارای بایاس است نسبت به شبکه‌ای

فاقد بایاس قوی‌تر است. به علاوه یک نرون بدون بایاس همیشه در اثر ورودی صفر، خروجی را صفر

می‌دهد که این حالت مطلوب نیست و با اعمال بایاس برطرف می‌شود [۴۰].

۲-۳-۷-۱-پ-تابع انتقال

توابع انتقال می‌توانند به صورت خطی و یا غیرخطی باشند. انتخاب تابع انتقال براساس نیاز خاص

حل مسئله‌ای می‌باشد که قرار است توسط شبکه‌ی عصبی حل شود. در عمل تعداد محدودی از توابع

انتقال مورد استفاده قرار می‌گیرند. توابع انتقال باید در محدوده‌ی مشخصی تعریف شده باشند و به

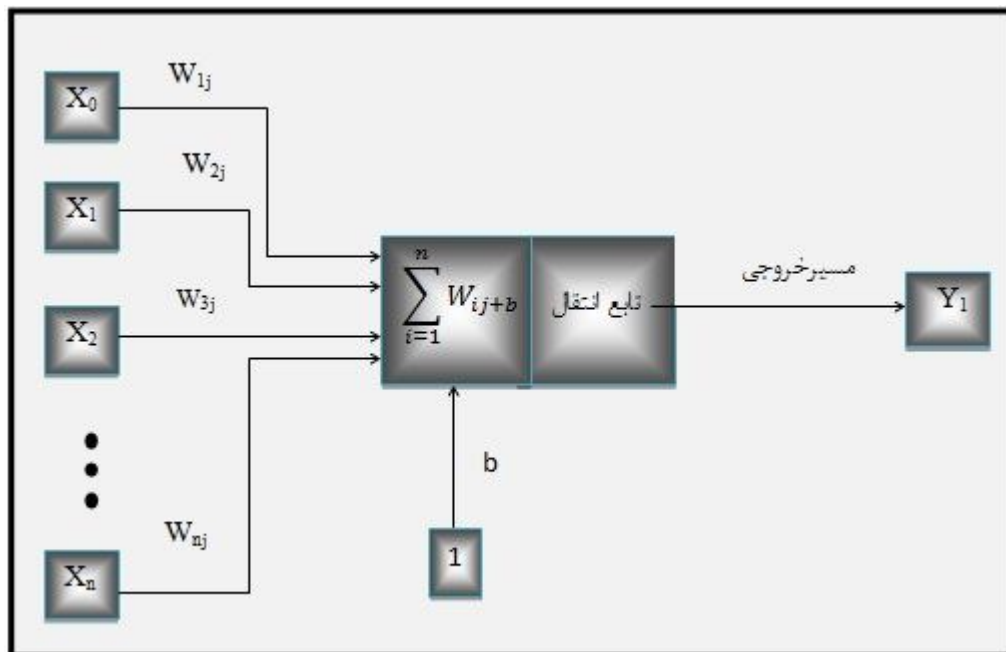
صورت تدریجی بین دو مقدار مجانب تغییر کنند. نمونه‌هایی از توابع انتقال به کار برده شده در

شبکه‌های عصبی مصنوعی عبارتند از: تابع انتقال pureline، تابع انتقال تانژانت سیگموئید و تابع

انتقال لگاریتم سیگموئید. تابع سیگموئیدی می‌تواند مقادیر ورودی را در محدوده‌ی $-\infty$ تا $+\infty$ دریافت

1- Bias

کند. اگر از تابع لگاریتم سیگموئیدی استفاده شود خروجی بین صفر و یک فشرده می‌شود، و اگر خروجی بین ۱ و -۱ فشرده شود، تابع تانژانت سیگموئیدی است [۴۱ و ۴۲].



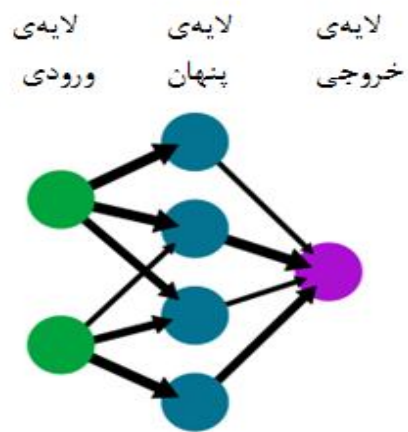
شکل (۲-۲) - شمای نرون مصنوعی نام

۲-۳-۷-۱-ت - ساختار لایه‌ای شبکه‌های عصبی

شبکه‌های عصبی مصنوعی به دو صورت تک لایه و چند لایه تفکیک می‌شوند. در شبکه‌های تک لایه، نرون‌های ورودی و خروجی یکی هستند. بنابراین به تعداد خروجی، نرون در شبکه وجود دارد، که همگی در یک لایه و به موازات هم کار می‌کنند.

گاهی دینامیک و قوانین شبکه‌های عصبی آن‌قدر پیچیده است که به پارامترهای قابل تنظیم بیشتری نیاز است، در این موارد از شبکه‌های چندلایه استفاده می‌شود. شبکه‌های چندلایه شامل لایه‌ی ورودی، لایه‌ی خروجی، و یک یا چند لایه‌ی پنهان یا لایه‌ی میانی می‌باشند. در این توپولوژی، نرون‌ها علاوه بر این‌که در لایه‌ی خود به شکل محدودی به یکدیگر اتصال داده شده‌اند، از طریق اتصال بین لایه‌ها نیز با یکدیگر ارتباط دارند [۴۳]. در یک شبکه چند لایه، هر لایه، ماتریس وزن، بردار بایاس،

بردار ورودی خالص و بردار خروجی مخصوص به خود را دارد که برای مشخص کردن لایه‌های مختلف از بالانویس استفاده می‌شود. در شکل (۳-۲) ساختار یک شبکه‌ی سه لایه نشان داده شده است [۴۰].



شکل (۳-۲) - ساختار یک شبکه‌ی سه لایه

۲-۳-۷-۱-ث - انواع شبکه‌ها از نظر ارتباطات بین نرونی

ارتباطات بین نرونی در یک شبکه‌ی عصبی از اهمیت بسیاری برخوردار می‌باشد و به نوعی قدرت شبکه را تعیین می‌کند. از نظر ارتباطات بین نرونی دو نوع تقسیم‌بندی در شبکه‌ها وجود دارد:

۱- شبکه‌های پیش‌خور^۱

۲- شبکه‌های پس‌خور^۲

در شبکه‌های پیش‌خور، اتصالات بین نرونی به گونه‌ای است که هر نرون لایه‌ی میانی، می‌تواند ورودی خود را از هر یک از نرون‌های لایه‌ی قبلی خود دریافت کند اما هیچ سیگنالی از نرون به خودش یا به نرون لایه‌ی قبل بر نمی‌گردد. بدین ترتیب در یک شبکه‌ی عصبی پیش‌خور، سیگنال‌ها به تدریج از یک لایه‌ی نرونی به لایه‌های بعدی حرکت می‌کنند و در نهایت به لایه‌ی آخر یا خروجی

1- Feed forward

2- Feed Back

شبکه می‌رسند. در نتیجه جریان اطلاعات در این نوع از شبکه‌ها همیشه در یک مسیر یک طرفه و از ورودی به خروجی می‌باشد.

نوع دیگری از ارتباط بین نرونی در شبکه‌های عصبی ارتباط بازخوردی است که در شبکه‌های پس‌خور وجود دارد. در این نوع از ارتباطات، خروجی یک نرون به همان نرون یا نرون‌های همان لایه یا به نرون‌های لایه‌ی قبلی اتصال داده می‌شود [۳۱]. بنابراین این نوع از شبکه‌ها می‌توانند با استفاده از حلقه‌های برگشتی، سیگنال‌هایی داشته باشند که در هر دو مسیر ورودی به خروجی و بالعکس حرکت کنند.

۲-۳-۷-۱-ج- آموزش شبکه عصبی

شبکه‌های عصبی می‌توانند براساس طراحی خود، سیگنال‌های ورودی را پردازش کنند و به سیگنال‌های خروجی مورد نظر تبدیل نمایند. به طور معمول پس از این که شبکه‌های عصبی طراحی و پیاده‌سازی شدند، باید پارامترهای بایاس و وزن‌ها (w و b) به ازای مجموعه‌هایی از سیگنال‌های ورودی، به گونه‌ای تنظیم شوند که سیگنال‌های خروجی شبکه، خروجی مطلوب را تشکیل دهند. چنین فرایندی را آموزش شبکه‌ی عصبی می‌نامند. به طور کلی برای آموزش شبکه دو روش وجود دارد: یادگیری با ناظر^۱ و یادگیری بدون ناظر^۲.

روش یادگیری با ناظر، شامل مراحل است که در آن سیستم از طریق در دست داشتن جواب واقعی آموزش می‌بیند. اما در روش بدون ناظر، شبکه‌ی عصبی باید خودسامان‌ده باشد و بدون در دست داشتن جواب واقعی، بتواند کار آموزش را انجام دهد. در این نوع یادگیری، شبکه می‌آموزد که الگوهای ورودی را به تعداد متناهی از گروه‌ها تقسیم‌بندی کند و به عبارتی عمل خوشه‌بندی را انجام

1- Supervised learning

2- Unsupervised learning

می‌دهد. باید توجه داشت در این حالت فرد طراح است که مقصد و هدف نهایی یعنی چیزی را که باید به آن رسید مشخص می‌کند [۳۱].

۲-۳-۷-۱-چ - تابع هدف

رفتار سیستم‌های یادگیر توسط الگوریتم‌های بازگشتی بیان می‌شود که به آن‌ها قوانین یادگیری می‌گویند. انواع مختلفی از قوانین یادگیری وجود دارد که یادگیری عملکردی^۱ یکی از آن‌هاست. قانون یادگیری در اینجا روندی است که توسط آن ماتریس وزن‌ها و بردارهای بایاس شبکه‌ی عصبی به نحوی تنظیم می‌شوند که عملکرد شبکه بهینه شود. منظور از بهینه شدن عملکرد شبکه، حداقل شدن خطایی است که بین مقادیر تجربی و پاسخ شبکه وجود دارد. برای بهینه‌سازی عملکرد، شبکه باید پارامترهایش را به گونه‌ای تنظیم کند که شاخص عملکرد^۲، کمترین مقدار را داشته باشد. شاخص عملکرد، معیاری است که به طور کمی، عملکرد شبکه را بهینه می‌کند و رابطه‌ی معکوسی با نحوه‌ی عملکرد شبکه دارد؛ یعنی هرچه مقدار این شاخص کوچک‌تر باشد، عملکرد شبکه بهتر خواهد بود و بالعکس. به شاخص عملکرد اصطلاحاً تابع هدف یا تابع خطا می‌گویند و در بیشتر الگوریتم‌ها، تابع هدف، متوسط مربعات خطا^۳ (MSE) می‌باشد [۳۱].

۲-۳-۷-۱-۵- معرفی روش‌های مختلف بهینه‌سازی عملکرد شبکه عصبی

الگوریتم‌های بهینه‌سازی تقریباً همگی دوری هستند؛ یعنی در تمام الگوریتم‌های بهینه‌سازی یک مقدار اولیه برای تابع حدس زده می‌شود و بعد این مقدار اولیه طی مراحل‌ی که از قاعده‌ی زیر پیروی می‌کند بهینه می‌شود [۳۱]:

1- Performance learning
2- Performance
3- Mean square error

$$X_{K+1} = X_K + \alpha P_K \quad (۴-۲)$$

که در آن، X_K تخمین نقطه‌ی مینیمم تابع در مرحله‌ی K ام است و X_{K+1} حدس مرحله‌ی بعدی نقطه‌ی مینیمم تابع می‌باشد. P_K نیز بردار جستجو می‌باشد؛ بدین معنی که از روی آن، تخمین مرحله‌ی بعد به دست می‌آید. α یک مقدار اسکالر بین صفر و یک بوده و سرعت یادگیری^۱ نامیده می‌شود. سرعت یادگیری، طول قدم در هر تکرار را در هر مسیر P_K تعیین می‌کند. انتخاب P_K و α روش‌های مختلف مینیمم‌سازی را از هم متمایز می‌کند که چند نمونه از آن‌ها در زیر ذکر شده است:

۱- روش نزول گرادیانی^۲

۲- الگوریتم پس انتشار خطا^۳

۳- روش لوببرگ-مارکوارت^۴

۴- الگوریتم تنظیم بایزین^۵

در این پایان‌نامه از الگوریتم‌های آموزشی لوببرگ-مارکوارت و تنظیم بایزین استفاده گردیده است که جزء روش‌های آموزش با ناظر می‌باشند.

الگوریتم آموزشی لوببرگ-مارکوارت در سال ۱۹۹۴ معرفی شد که روش مناسبی برای آموزش شبکه‌هایی هستند که شاخص عملکرد آن‌ها حداقل متوسط مربعات خطا می‌باشد. این روش تاکنون به عنوان سریع‌ترین و مؤثرترین روش آموزش شبکه‌های عصبی شناخته شده است. سرعت همگرایی این روش صدها برابر بیشتر از شیوه‌های استاندارد مشابه است [۳۱].

معادله‌ی مربوط به الگوریتم لوببرگ-مارکوارت برای بهینه‌سازی پارامترهای شبکه و کاهش تابع خطا در هر مرحله، طبق رابطه‌ی (۲-۵) می‌باشد:

1- Learning rate
2- Gradient descent algorithm
3- Error back propagation algorithm
4- Levenberg-Marquart
5- Bayesian regularization

$$X_{k+1} = X_k - [J^T(X_k)J(X_k) + \mu_k I]^{-1} J^T(X_k)E(X_k) \quad (5-2)$$

که در آن X برداری است که مولفه‌هایش پارامترهای تنظیم پذیر شبکه (وزن‌ها و بایاس‌ها) می‌باشند. k نشان‌دهنده‌ی مرحله‌ی آموزش، J معرف ماتریس ژاکوبین و I ماتریس یکانی است. μ نیز ضریب یادگیری مارکوات بوده و عددی غیر منفی است. ماتریس ژاکوبین چنین تعریف می‌شود:

$$\begin{bmatrix} \frac{\partial e_1(x)}{\partial x_1} & \frac{\partial e_1(x)}{\partial x_2} & \dots & \frac{\partial e_1(x)}{\partial x_n} \\ \frac{\partial e_2(x)}{\partial x_1} & \frac{\partial e_2(x)}{\partial x_2} & \dots & \frac{\partial e_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_n(x)}{\partial x_1} & \frac{\partial e_n(x)}{\partial x_2} & \dots & \frac{\partial e_n(x)}{\partial x_n} \end{bmatrix} \quad (6-2)$$

یعنی ماتریس ژاکوبین شامل مشتقات مرتبه‌ی اول خطای شبکه نسبت به پارامترهای شبکه است.

برای جلوگیری از آموزش بیش از حد شبکه و ارتقاء عمومیت شبکه، الگوریتم آموزشی تنظیم بایزین پیشنهاد شده است که در این الگوریتم تابع عملکرد متداول (MSE) از طریق افزایش عبارات میانگین مربعات وزن‌ها و بایاس‌های شبکه اصلاح شده است.

$$F = \beta E_D + \alpha E_W \quad (7-2)$$

که در آن α و β پارامترهای تابع کارایی، E_D و E_W به ترتیب میانگین مربعات خطاها و میانگین مربعات وزن‌ها می‌باشند که به صورت زیر تعریف می‌شوند:

$$E_D = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - t_i)^2 \quad (8-2)$$

$$E_W = \frac{1}{N} \sum_{j=1}^N (w_j)^2 \quad (9-2)$$

که در این روابط e_i بیانگر خطای مطلق، t_i مقادیر تجربی، y_i مقادیر محاسبه شده برای پارامتر تجربی مدنظر و N نشان‌دهنده‌ی تعداد نمونه‌ها می‌باشد [۴۴].

۲-۳-۸- ارزیابی قدرت پیش‌بینی مدل

برای اطمینان از این که مدل به دست آمده، مدل مناسبی است و توانایی پیش‌بینی نمونه‌های مختلفی از یک جمعیت را داراست، باید مدل را ارزیابی کرد. روش‌های مختلفی برای ارزیابی وجود دارد:

۱. با استفاده از پارامترهای آماری: به وسیله‌ی این شاخص‌های کمی صحت نتایج ارائه شده توسط مدل مورد سنجش قرار می‌گیرند. در ادامه راجع به مهم‌ترین پارامترهای آماری که برای انتخاب بهترین مدل و به تبع آن برای بهترین توصیف‌کننده‌ها مدنظر قرار دارند توضیحاتی ارائه خواهد شد.

ضریب همبستگی^۱: ساده‌ترین راه برای بررسی میزان همبستگی دو یا چند متغیر، محاسبه آماره‌ی ضریب همبستگی آن‌هاست. ضریب همبستگی دو متغیر x, y با رابطه‌ی (۱۰-۲) تعریف می‌شود. مقدار این آماره بین ۱ تا -۱ متغیر است. مقدار بزرگ‌تر آن نشان دهنده‌ی این است که ارتباط خطی بیشتری میان متغیر وابسته و متغیرهای مستقل وجود دارد.

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (10-2)$$

ضریب تعیین^۲: به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و نشان‌دهنده‌ی نسبت تغییرات متغیر وابسته به متغیر مستقل است. رابطه‌ی ریاضی مربوط به ضریب تعیین به صورت زیر است:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (11-2)$$

که در آن SSR طبق رابطه‌ی (۱۲-۲) بیانگر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی

1- Correlation coefficient

2 - Determination coefficient

3- Sum Square Regression

متغیر وابسته از میانگین مقادیر آن است:

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (12-2)$$

و SST^1 طبق رابطه (2-13) نشانگر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (13-2)$$

همچنین SSE^2 نیز مبین مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است:

$$SST = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (14-2)$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15-2)$$

طبق رابطه‌ی (2-15) اگر تمام مشاهدات بر روی خط برازش شده قرار گرفته باشند، یعنی به ازای تمام نقاط $y_i = \hat{y}_i$ باشد، مقدار R^2 برابر یک می‌شود و هرگونه انحرافی از این حالت باعث می‌شود که مقدار R^2 از یک کوچکتر شود.

ضریب تعیین تصحیح شده³: ضریب تعیین تصحیح شده یکی از پارامترهایی است که در و مقایسه مدل‌های مختلف با تعداد متفاوت متغیرهای پیشگو استفاده می‌شود و از آن جایی که R^2 (ضریب تعیین)، تعداد پارامترهای موجود در مدل را به حساب نمی‌آورد، بدین منظور از ضریب تعیین تصحیح شده طبق رابطه زیر برای انتخاب مدل برتر استفاده می‌شود:

1- Sum Square Total
2- Sum Square Errore
3- Adjusted determination coefficient

$$R_{adj}^2 = 1 - \frac{n-1}{n-p-1} \cdot \frac{SSE}{SST} = 1 - (1-R^2) \frac{n-1}{n-p-1} \quad (۱۶-۲)$$

که در این رابطه، p تعداد متغیرهای مستقل و n تعداد ترکیبات مورد بررسی می باشد.

آماره F : آزمون F یا آزمون فشر در واقع آزمون معنی دار بودن آماری در تحلیل رگرسیون ساده و چند متغیره است و برابر با نسبت میانگین مربعات رگرسیون (MSR) به میانگین مربعات باقیمانده‌ها (MSE) است. بیان ریاضی آن به صورت زیر می باشد:

$$F = \frac{MSR}{MSE} = \frac{SSR / df_m}{SSE / df_{res}} \quad (۱۷-۲)$$

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

مجموع مربع باقیمانده‌ها^۱ ($PRESS$): برابر مجموع مربعات تفاوت بین مقدار کمیت مشاهده شده (y_i) و مقدار تخمین زده شده ($\hat{y}_i$) است.

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (۱۸-۲)$$

خطای استاندارد پیش‌بینی^۲ (SEP):

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (۱۹-۲)$$

میانگین مربع خطاها^۳ (MSE):

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (۲۰-۲)$$

1- Predictive Residual Sum of Squares
2- Standard Error of Prediction
3- Mean Square Error

خطای مطلق میانگین^۱ (MAE):

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}|}{n} \quad (21-2)$$

میانگین درصد انحراف مطلق^۲ (AAD):

$$AAD = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right| \quad (22-2)$$

میانگین انحراف^۳ ($bias$):

$$bias = \frac{100}{n} \sum_{i=1}^n \frac{y_i - \hat{y}}{y_i} \quad (23-2)$$

بیشترین انحراف^۴ (D_{Max}):

$$D_{Max} = \text{Max}(100 \left| \frac{y_i - \hat{y}}{y_i} \right|) \quad (24-2)$$

FIT: به عنوان یک پارامتر آماری برای مقایسه مدل‌های مختلف با تعداد متفاوت توصیف‌کننده به کار می‌رود که هر چه مقدار این پارامتر برای یک مدل بیشتر باشد، آن مدل مناسب‌تر می‌باشد. مقدار آن از رابطه‌ی زیر به دست می‌آید.

$$FIT = \frac{R^2(n-p-1)}{(n+p^2)(1-R^2)} \quad (25-2)$$

1- Mean Absolute Error
 2- Absolute Average percent deviation
 3 - average deviation
 4 - Maximum Deviation

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

۲. با استفاده از نمودار برگشتی: در این نمودار مقادیر پیش‌بینی شده‌ی کمیت مورد نظر بر حسب مقادیر تجربی رسم می‌شود و به کمک ضریب همبستگی (R^2) به‌دست آمده از نمودار، پراکندگی نقاط اطراف خط برگشت تعیین می‌شود. محدوده‌ی تغییرات بین صفر و ۱ است. اگر $R^2=1$ باشد، همبستگی کاملی بین داده‌ها وجود دارد و نتایج پیش‌بینی شده به واقعیت نزدیک‌ترند. اما اگر $R^2=0$ باشد، بین داده‌ها هیچ‌گونه همبستگی وجود ندارد.

۳. با استفاده از نمودار خطای باقی‌مانده: منظور از عبارت خطای باقی‌مانده، اختلاف بین مقادیر پیش‌بینی‌شده و مقادیر تجربی است. اگر پراکندگی مقادیر در دو طرف نمودار صفر باشد، این امر نشان‌دهنده‌ی تصادفی بودن خطاهاست. ولی اگر عمده‌ی نقاط، در این نمودار، در یک طرف صفر باشد، این بدان معناست که خطای جهت‌داری رخ داده است.

۴. با استفاده از آزمون Y - تصادفی: در این آزمون، مقادیر تجربی به صورت تصادفی در محدوده‌ی همان مقادیر، تغییر داده شده و سپس همبستگی متغیرهای مستقل با متغیرهای وابسته با استفاده از یکی از شاخص‌های آماری که معمولاً R^2 است، مورد بررسی قرار می‌گیرد. اختلاف زیاد بین شاخص آماری به دست آمده از این روش و شاخص آماری به‌دست آمده از مدل اصلی، نشان‌دهنده‌ی عدم وجود همبستگی شانسی بین متغیر مستقل اصلی و مقادیر تصادفی می‌باشد. به طور معمول این فرایند چندین بار تکرار می‌گردد.

۵. استفاده از تست بیرونی: اعتبار و اهمیت مدل‌های پیش‌بینی‌شده وقتی مشخص می‌شود که بتواند خواص مولکول‌های مجهولی که در دسته‌ی آموزش موجود نیست را پیش‌بینی کند. بدین منظور از ابتدای کار ۱۵٪ از مولکول‌ها به عنوان سری تست بیرونی کنار گذاشته

می‌شود و در نهایت از مدل‌های منتخب بهینه شده، جهت پیش‌بینی خاصیت مورد نظر مولکول‌های تست بیرونی که در مدل‌سازی استفاده نشده‌اند، به کار گرفته می‌شود.

فصل سوم

محاسبات

۳-۱- مطالعه‌ی ارتباط کمی ساختار-خاصیت دانسیته‌ی سیالات تجمعی در

محدوده‌ی وسیعی از دما و فشار

در مطالعه‌ی سیستم‌های شیمیایی، بیوشیمیایی و طبیعی، دانشی از خواص مختلف فیزیکی و شیمیایی مواد خالص و یا مخلوط‌ها تحت شرایط مختلف ترمودینامیکی ضروری است. یکی از مهم‌ترین این خواص، دانسیته می‌باشد. اما همیشه یافتن مقادیر تجربی در شرایط مختلف دما و فشار، خصوصاً از منابع علمی امکان‌پذیر نیست، به‌علاوه اندازه‌گیری تجربی این خواص مشکل است و به صرف هزینه و وقت زیادی نیاز دارد. در این شرایط استفاده از روش‌های مختلف علمی برای پیش‌بینی دانسیته کارآمدتر و به‌صرفه‌تر است.

در این تحقیق، از تکنیک‌های رگرسیون مرحله‌ای (SR)، و الگوریتم ژنتیک (GA) و سهم گروه (GCM) به عنوان روش‌های انتخاب توصیف‌کننده و از روش شبکه عصبی مصنوعی جهت مدل‌سازی غیرخطی برای پیش‌بینی دانسیته‌ی سیالات تجمعی در محدوده‌ی وسیعی از دما و فشار استفاده گردید. آنچه در ادامه می‌آید مراحل طراحی و بهینه‌سازی و همچنین ارزیابی مدل‌های مختلف با استفاده از این روش‌ها می‌باشد.

۳-۲- مراحل مدل‌سازی

۳-۲-۱- جمع‌آوری سری داده‌ها

دانسیته‌ی ۲۸ الکل نوع اول، دوم و سوم، ۱۱ کربوکسیلیک اسید و ۲۰ آمین نوع اول و دوم، شامل ۲۰۹۳ نقطه داده‌ی دانسیته در محدوده وسیعی از دما و فشار به عنوان سری داده‌ها جمع‌آوری گردید. نام این ترکیبات به همراه محدوده‌ی دما، فشار و دانسیته، مرجع مورد استفاده برای هر ترکیب و تعداد نقاط داده مورد استفاده از هر ترکیب در مدل‌سازی در جدول (۳-۱) آمده است.

جدول (۱-۳) - محدوده‌ی دما و فشار و تعداد داده‌های استفاده شده در مدلسازی برای سیالات تجمعی مختلف به همراه مرجع مورد استفاده برای هر یک از آن‌ها

مرجع	تعداد داده‌ی استفاده شده	$\Delta\rho(kg.m^{-3})$	$\Delta P (MPa)$	$\Delta T (K)$	نام ترکیب
۱۰	۵۶	۷۲۹/۵-۸۴۱/۸	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	اتانول
۱۴	۵۶	۱۰۶۶/۱-۱۱۴۳/۳	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	اتان دیول
۱۰	۵۶	۷۴۷/۷-۸۵۲/۱	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱-پروپانول
۴۵	۳۳	۷۵۹/۷-۸۸۷/۴	۰/۱ - ۱۷۳/۹	۲۷۸/۱۵ - ۳۲۳/۱۵	۲-پروپانول
۴۵	۵۲	۷۵۷/۳-۸۹۹/۴	۰/۱ - ۲۰۶/۱	۲۸۳/۱۵ - ۳۴۸/۱۵	۲-متیل-۱-پروپانول
۴۵	۲۳	۷۲۶/۶-۸۰۸/۵	۰/۱ - ۹۵/۷	۳۲۳/۱۵ - ۳۴۸/۱۵	۲-متیل-۲-پروپانول
۱۴	۵۶	۹۸۵/۸-۱۰۷۱/۴	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱و۲-پروپان دیول
۱۰	۵۶	۷۵۶/۶-۸۵۵/۶	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱-بوتانول
۴۵	۵۲	۷۵۷/۳-۹۰۳/۷	۰/۱ - ۲۰۵/۸	۲۸۳/۱۵ - ۳۴۸/۱۵	۲-بوتانول
۴۶	۳۳	۷۹۹/۶-۸۷۰/۸	۰/۱ - ۱۰۰	۲۹۳/۱۵ - ۳۱۸/۱۵	۲-متیل-۱-بوتانول
۱۴	۵۶	۹۵۱/۳-۱۰۹۳/۲	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱و۲-بوتان دیول
۴۷	۳۷	۷۰۳/۴-۹۰۰/۵	۱۰ - ۲۰۰	۲۲۸/۱۵ - ۴۴۳	۱-پنتانول
۴۷	۳۸	۶۸۰/۶-۸۹۸/۹	۱۰ - ۲۰۰	۲۳۴ - ۴۳۳	۲-پنتانول
۴۷	۴۱	۶۹۱/۹-۹۱۹/۰	۱۰ - ۲۰۰	۲۳۳/۶ - ۴۳۳/۱	۳-پنتانول
۱۵	۴۴	۷۵۳/۵-۸۷۶/۸	۰/۱ - ۱۰۰	۳۰۳/۱۵ - ۳۶۸/۱۵	۲-متیل-۳-پنتانول
۱۵	۴۴	۷۷۰/۳-۸۸۸/۱	۰/۱ - ۱۰۰	۳۰۳/۱۵ - ۳۶۸/۱۵	۳-اتیل-۳-پنتانول
۴۸	۳۳	۹۰۳/۸-۹۷۱/۹	۰/۱ - ۱۰۰	۲۹۳/۱۵ - ۳۱۸/۱۵	۲-متیل-۱و۲-پنتان دیول
۱۰	۵۶	۷۶۹/۴-۸۶۰/۹	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱-هگزانول
۴۹	۶۶	۷۵۴/۹-۸۱۴/۴	۱/۰۱ - ۲۲/۰۰۴	۳۱۳/۱۳ - ۳۶۲/۷۳	۲-هگزانول
۱۰	۵۶	۷۷۴/۲-۸۶۳/۴	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱-هپتانول
۵۰	۶۵	۷۶۱/۲-۸۱۷/۲	۱/۳۰۰ - ۲۲/۰۰۶۳	۳۱۳/۱۶ - ۳۶۲/۷۲	۲-هپتانول
۱۰	۵۶	۷۷۸-۸۶۵/۵	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱-اوکتانول
۱۰	۵۶	۷۸۰/۹-۸۶۶/۴	۰/۱ - ۶۰	۲۷۸/۱۵ - ۳۵۸/۱۵	۱-نونانول
۱۰	۵۶	۷۸۳/۶-۸۶۲/۷	۰/۱ - ۶۰	۲۸۸/۱۵ - ۳۵۸/۱۵	۱-دکانول
۱۰	۴۲	۷۸۶-۸۵۸/۷	۰/۱ - ۶۰	۲۹۸/۱۵ - ۳۵۸/۱۵	۱-اندکانول
۵۱	۲۳	۷۹۵/۲-۸۳۶/۱	۰/۱ - ۴۰	۳۲۳/۱۵ - ۳۴۸/۱۵	۱-دودکانول
۵۱	۲۴	۷۹۸/۰-۸۳۷/۹	۰/۱ - ۳۹/۶۵	۳۲۳/۱۵ - ۳۴۸/۱۵	۱-تترادکانول
۵۱	۱۱	۸۰۰/۲-۸۲۴/۹	۰/۱ - ۴۰/۱	۳۴۸/۱۵	۱-هگزادکانول
۵۲	۱۰	۷۱۵/۷-۷۴۰/۴	۱ - ۳۳۹	۲۹۸/۱۵	پروپیل آمین
۵۲	۱۰	۷۳۲/۲-۷۵۵/۴	۱ - ۳۳۹	۲۹۸/۱۵	n-بوتیل آمین
۵۲	۵۸	۶۲۳/۹-۸۰۱/۵	۰/۱ - ۱۴۰	۲۹۳/۱۵ - ۴۰۳/۱۵	۲-آمینو بوتان
۵۴	۲۳	۷۰۹/۹-۸۰۳/۸	۰/۱ - ۱۰۰	۲۹۳/۱۵ - ۳۵۳/۱۵	۲-آمینو پنتان
۵۴	۲۴	۷۱۳/۲-۸۱۶/۷	۰/۱ - ۱۰۰	۲۹۳/۱۵ - ۳۵۳/۱۵	۲-آمینوهپتان

ادامه جدول (۳-۱)

۲-آمینوآکتان	۲۹۳/۱۵ - ۴۰۳/۱۵	۰/۱ - ۱۴۰	۶۸۰/۱-۸۳۷/۶	۶۰	۵۳
پنتیل آمین	۲۹۳/۱۵-۳۵۳/۱۵	۰/۱ - ۱۴۰	۶۹۸/۰-۸۲۲/۰	۶۰	۵۵
هگزیل آمین	۲۹۳/۱۵-۳۵۳/۱۵	۰/۱ - ۱۴۰	۷۱۲/۴-۸۳۰/۵	۶۰	۵۵
هپتیل آمین	۲۹۳/۱۵-۳۵۳/۱۵	۰/۱ - ۱۴۰	۷۲۴/۴-۸۳۸/۳	۶۰	۵۵
۱-آمینو-۲-پروپانول	۲۹۳/۱۵-۳۲۳/۱۵	۱	۹۳۶/۴-۹۶۱/۲	۷	۵۶
۳-آمینو-۱-پروپانول	۲۹۳/۱۵-۳۲۳/۱۵	۱	۹۶۶/۸-۹۹۱/۳	۷	۵۶
۲-آمینو-۲-متیل-۱-پروپانول	۲۹۳/۱۵-۳۲۳/۱۵	۱	۹۱۱/۴-۹۳۶/۷	۷	۵۶
دی اتانول آمین	۲۹۳/۱۵-۳۲۳/۱۵	۱	۱۰۷۷/۴-۱۰۹۷/۲	۷	۵۶
اکتیل آمین	۲۹۳/۱۵-۳۲۳/۱۵	۱	۷۵۶/۸-۷۷۹/۳	۷	۵۶
دسیل آمین	۳۲۸/۱۵-۲۹۸/۱۵	۱	۷۶۹/۳-۷۸۷/۳	۶	۵۶
دودسیل آمین	۳۲۸/۱۵-۲۹۸/۱۵	۱	۷۷۷/۴-۷۹۴/۸	۶	۵۶
دی اتیل آمین	۳۲۸/۱۵-۲۹۸/۱۵	۱	۶۷۶/۹-۷۰۰/۰	۵	۵۶
دی پروپیل آمین	۳۲۸/۱۵-۲۹۸/۱۵	۱	۷۰۸/۶-۷۳۵/۶	۷	۵۶
دی بوتیل آمین	۳۲۸/۱۵-۲۹۸/۱۵	۱	۷۳۱/۸-۷۵۵/۹	۷	۵۶
دی هگزیل آمین	۳۲۳/۱۵-۲۹۸/۱۵	۱	۷۶۸/۵-۷۳۲/۹	۵	۵۶
دی اکتیل آمین	۳۲۸/۱۵-۳۰۳/۱۵	۱	۷۷۹/۰-۷۹۶/۶	۴	۵۶
استیک اسید	۲۹۳/۱۵ - ۳۲۳/۱۵	۰/۱ - ۲۵	۱۰۱۴/۵-۱۰۶۷/۷	۴۰	۵۸
پروپانویک اسید	۲۹۳/۱۵ - ۳۲۳/۱۵	۰/۱ - ۲۵	۱۰۱۳/۵-۹۶۱/۳	۴۰	۵۸
بوتانویک اسید	۲۹۳/۱۵ - ۳۲۳/۱۵	۰/۱ - ۲۵	۹۲۷/۲-۹۷۸/۲	۴۰	۵۸
پنتانویک اسید	۲۹۳/۱۵ - ۳۲۳/۱۵	۰/۱ - ۲۵	۹۱۱/۱-۹۵۸/۴	۴۰	۵۸
هگزانویک اسید	۲۹۳/۱۵ - ۳۲۳/۱۵	۰/۱ - ۲۵	۸۹۹/۶-۹۴۴/۹	۴۰	۵۸
هپتانویک اسید	۲۹۳/۱۵ - ۳۲۳/۱۵	۰/۱ - ۲۵	۸۹۱/۶-۹۳۵/۷	۴۰	۵۸
اکتانویک اسید	۲۹۳/۱۵ - ۳۲۳/۱۵	۰/۱ - ۲۵	۸۸۷/۰-۹۲۹/۹	۴۰	۵۸
دکانویک اسید	۳۴۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۹	۸۲۶/۱-۸۶۸/۲	۲۴	۵۹
دودکانویک اسید	۳۴۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۹	۸۳۳/۵-۸۶۲/۲	۲۴	۵۹
تترادکانویک اسید	۳۵۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۹	۸۲۹/۸-۸۵۱/۳	۲۴	۵۹
هگزادکانویک اسید	۳۵۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۹	۸۲۶/۹-۸۴۸/۱	۲۴	۵۹

۳-۲-۲- بهینه‌سازی ساختار مولکول‌ها

پس از جمع‌آوری سری داده‌ها، اولین مرحله جهت مدل‌سازی، رسم و بهینه‌سازی ساختار

مولکول‌های مورد بررسی می‌باشد. به این منظور ساختار تمام ترکیبات با استفاده از نرم افزار

HyperChem رسم و سپس با احتساب اتم‌های هیدروژن و استفاده از روش نیمه تجربی AM1

ساختار سه‌بعدی ترکیبات بهینه گردید. بهینه‌سازی تا حد گرادیانی $0/001$ کیلوکالری بر مول انجام شد.

۳-۲-۳- محاسبه‌ی توصیف‌کننده‌ها

در این تحقیق برای محاسبه‌ی توصیف‌کننده‌ها از دو روش سهم گروه و استفاده از نرم افزار Dragon به‌کار برده شد که در ادامه در مورد هریک توضیح داده خواهد شد.

۳-۲-۳-۱- محاسبه‌ی توصیف‌کننده‌ها با استفاده از نرم‌افزار Dragon

برای محاسبه‌ی اغلب توصیف‌کننده‌ها، نیاز به ساختار سه‌بعدی و حالت پایدار مولکول می‌باشیم. برای این منظور ساختارهای بهینه شده‌ی مولکول‌ها به عنوان ورودی به نرم‌افزار Dragon وارد گردید و برای هر مولکول ۱۴۸۱ توصیف‌کننده‌ی ساختاری در ۱۸ طبقه‌ی مختلف به‌دست آمد. با توجه به این‌که دانسیته به دما و فشار وابسته است، لذا دو متغیر تجربی دما و فشار هم به آن‌ها اضافه شد. به این ترتیب ۱۴۸۳ توصیف‌کننده برای هر ترکیب به‌دست آمد.

۳-۲-۳-۲- محاسبه‌ی توصیف‌کننده‌ها به روش سهم گروه

در محاسبه و انتخاب توصیف‌کننده‌ها بر مبنای روش سهم گروه‌ها، ابتدا باید گروه‌های ساختاری مولکول‌ها را مشخص نمود. با توجه به اختیاری بودن تعریف گروه‌های تشکیل‌دهنده‌ی مواد در روش سهم گروه‌ها، در این پایان‌نامه شکستن مولکول‌ها از محل پیوندهای کربن-کربن صورت گرفته است. به این ترتیب ۱۰ گروه عاملی به ترتیبی که در جدول (۳-۲) آمده است در ترکیبات مورد بررسی شناسایی شدند. تعداد هر یک از گروه‌های عاملی موجود در مولکول‌های مورد بررسی به عنوان توصیف‌کننده‌های حاصل از روش سهم گروه‌ها انتخاب شدند که مقادیر آن برای تمامی مولکول‌های

موجود در سری داده‌ها در جدول (۳-۳) آورده شده است. در صورتی که هر گروه عاملی در مولکول مورد نظر وجود نداشته باشد، برای آن عدد صفر در نظر گرفته می‌شود.

مراحل حذف و انتخاب بهترین توصیف‌کننده‌ها برای این نوع از توصیف‌کننده‌ها انجام نمی‌شود. بلکه داده‌های مربوط به این توصیف‌کننده‌ها پس از دسته‌بندی و مرتب‌سازی CR برای مدل‌سازی وارد شبکه‌ی عصبی مصنوعی می‌شود. منظور از مرتب‌سازی CR این است که همبستگی توصیف‌کننده‌ها با خاصیت موردنظر یعنی دانسیته توسط نرم افزار SPSS محاسبه و به ترتیب کاهشی این مقدار مرتب می‌شوند. مقادیر همبستگی توصیف‌کننده‌های حاصل از روش سهم گروه با دانسیته در جدول (۳-۴) ارائه شده است.

جدول (۳-۲)- گروه‌های عاملی موجود در مولکول‌های مورد بررسی به عنوان توصیف‌کننده‌های روش سهم گروه

	گروه عاملی	فرمول شیمیایی
۱	متیل	-CH ₃
۲	کربن نوع دوم	-CH ₂ -
۳	کربن نوع سوم	>CH-
۴	الکل نوع اول	-CH ₂ -OH
۵	الکل نوع دوم	>CH-OH
۶	الکل نوع سوم	$\begin{array}{c} \\ -C- OH \\ \end{array}$
۷	کربوکسیلیک اسید	-COOH
۸	آمین نوع اول به فرم ۱	-CH ₂ -NH ₂
۹	آمین نوع اول به فرم ۲	>CH-NH ₂
۱۰	آمین نوع دوم	>NH

جدول (۳-۳) - تعداد گروه‌های عاملی موجود در مولکول‌های مختلف سری داده‌ها به عنوان توصیف‌کننده‌های روش سهم

گروه

ردیف	نام مولکول	گروه‌های عاملی									
		-CH ₃	-CH ₂ -	>CH-	-CH ₂ -OH	>CH-OH	-C-OH	-COOH	-CH ₂ -NH ₂	>CH-NH ₂	>NH
۱	اتانول	۱	۰	۰	۱	۰	۰	۰	۰	۰	۰
۲	اتان دیول	۰	۰	۰	۲	۰	۰	۰	۰	۰	۰
۳	۱-پروپانول	۱	۱	۰	۱	۰	۰	۰	۰	۰	۰
۴	۲-پروپانول	۲	۰	۰	۰	۱	۰	۰	۰	۰	۰
۵	۲-متیل-۱-پروپانول	۲	۰	۱	۱	۰	۰	۰	۰	۰	۰
۶	۲-متیل-۲-پروپانول	۳	۰	۰	۰	۱	۰	۰	۰	۰	۰
۷	۱ و ۲-پروپان دیول	۱	۰	۰	۱	۱	۰	۰	۰	۰	۰
۸	۱-بوتانول	۱	۲	۰	۱	۰	۰	۰	۰	۰	۰
۹	۲-بوتانول	۲	۱	۰	۰	۱	۰	۰	۰	۰	۰
۱۰	۲-متیل-۱-بوتانول	۲	۱	۱	۱	۰	۰	۰	۰	۰	۰
۱۱	۱ و ۲-بوتان دیول	۱	۱	۰	۱	۱	۰	۰	۰	۰	۰
۱۲	۱-پنتانول	۱	۳	۰	۱	۰	۰	۰	۰	۰	۰
۱۳	۲-پنتانول	۲	۲	۰	۰	۱	۰	۰	۰	۰	۰
۱۴	۳-پنتانول	۲	۲	۰	۰	۱	۰	۰	۰	۰	۰
۱۵	۲-متیل-۳-پنتانول	۳	۱	۱	۰	۱	۰	۰	۰	۰	۰
۱۶	۳-اتیل-۳-پنتانول	۳	۳	۰	۰	۰	۱	۰	۰	۰	۰
۱۷	۲-متیل-۲ و ۴-پنتان دیول	۳	۱	۰	۰	۱	۱	۰	۰	۰	۰
۱۸	۱-هگزانول	۱	۴	۰	۱	۰	۰	۰	۰	۰	۰
۱۹	۲-هگزانول	۲	۳	۰	۰	۱	۰	۰	۰	۰	۰
۲۰	۱-هپتانول	۱	۵	۰	۱	۰	۰	۰	۰	۰	۰
۲۱	۲-هپتانول	۲	۴	۰	۰	۱	۰	۰	۰	۰	۰
۲۲	۱-اوکتانول	۱	۶	۰	۱	۰	۰	۰	۰	۰	۰
۲۳	۱-نونانول	۱	۷	۰	۱	۰	۰	۰	۰	۰	۰
۲۴	۱-دکانول	۱	۸	۰	۱	۰	۰	۰	۰	۰	۰
۲۵	۱-اندکانول	۱	۹	۰	۱	۰	۰	۰	۰	۰	۰
۲۶	۱-دودکانول	۱	۱۰	۰	۱	۰	۰	۰	۰	۰	۰

ادامه جدول (۳-۳)

ردیف	نام مولکول	گروه‌های عاملی									
		-CH ₃	-CH ₂ -	>CH-	-CH ₂ -OH	>CH-OH	>CHOH	->C-OH	-CH ₂ -NH ₂	>CH-NH ₂	>NH
۲۷	۱-تترادکانول	۱	۱۲	۰	۱	۰	۰	۰	۰	۰	۰
۲۸	۱-هگزادکانول	۱	۱۴	۰	۱	۰	۰	۰	۰	۰	۰
۲۹	پروپیل آمین	۱	۱	۰	۰	۰	۰	۰	۱	۰	۰
۳۰	n-بوتیل آمین	۱	۲	۰	۰	۰	۰	۰	۱	۰	۰
۳۱	۲-آمینو بوتان	۲	۱	۰	۰	۰	۰	۰	۰	۱	۰
۳۲	۲-آمینو پنتان	۲	۲	۰	۰	۰	۰	۰	۰	۱	۰
۳۳	۲-آمینوهپتان	۲	۴	۰	۰	۰	۰	۰	۰	۱	۰
۳۴	۲-آمینواکتان	۲	۵	۰	۰	۰	۰	۰	۰	۱	۰
۳۵	پنتیل آمین	۱	۳	۰	۰	۰	۰	۰	۱	۰	۰
۳۶	هگزیل آمین	۱	۴	۰	۰	۰	۰	۰	۱	۰	۰
۳۷	هپتیل آمین	۱	۵	۰	۰	۰	۰	۰	۱	۰	۰
۳۸	اکتیل آمین	۱	۶	۰	۰	۰	۰	۰	۱	۰	۰
۳۹	دسیل آمین	۱	۸	۰	۰	۰	۰	۰	۱	۰	۰
۴۰	دودسیل آمین	۱	۱۰	۰	۰	۰	۰	۰	۱	۰	۰
۴۱	دی اتیل آمین	۲	۰	۰	۰	۰	۰	۰	۰	۰	۱
۴۲	دی پروپیل آمین	۲	۲	۰	۰	۰	۰	۰	۰	۰	۱
۴۳	دی بوتیل آمین	۲	۴	۰	۰	۰	۰	۰	۰	۰	۱
۴۴	دی هگزیل آمین	۲	۸	۰	۰	۰	۰	۰	۰	۰	۱
۴۵	دی اکتیل آمین	۲	۱۲	۰	۰	۰	۰	۰	۰	۰	۱
۴۶	۱-آمینو-۲-پروپانول	۱	۱	۰	۰	۱	۰	۰	۱	۰	۰
۴۷	۳-آمینو-۱-پروپانول	۰	۱	۰	۱	۰	۰	۰	۱	۰	۰
۴۸	۲-آمینو-۱-متیل-۱-پروپانول	۲	۰	۰		۱	۰	۰	۰	۱	۰
۴۹	دی اتانول آمین	۰	۰	۰	۳	۰	۰	۰	۰	۰	۱
۵۰	استیک اسید	۱	۰	۰	۰	۰	۰	۱	۰	۰	۰
۵۱	پروپانویک اسید	۱	۱	۰	۰	۰	۰	۱	۰	۰	۰
۵۲	بوتانویک اسید	۱	۲	۰	۰	۰	۰	۱	۰	۰	۰
۵۳	پنتانویک اسید	۱	۳	۰	۰	۰	۰	۱	۰	۰	۰
۵۴	هگزانویک اسید	۱	۴	۰	۰	۰	۰	۱	۰	۰	۰
۵۵	هپتانویک اسید	۱	۵	۰	۰	۰	۰	۱	۰	۰	۰

ادامه جدول (۳-۳)

ردیف	نام مولکول	گروه‌های عاملی									
		-CH ₃	-CH ₂ -	>CH-	-CH ₂ -OH	>CH-OH	>-C-OH	-COOH	-CH ₂ -NH ₂	>CH-NH ₂	>NH
۵۶	اکتانویک اسید	۱	۶	۰	۰	۰	۰	۱	۰	۰	۰
۵۷	دکانویک اسید	۱	۸	۰	۰	۰	۰	۱	۰	۰	۰
۵۸	دودکانویک اسید	۱	۱۰	۰	۰	۰	۰	۱	۰	۰	۰
۵۹	تترادکانویک اسید	۱	۱۲	۰	۰	۰	۰	۱	۰	۰	۰
۶۰	هگزادکانویک اسید	۱	۱۴	۰	۰	۰	۰	۱	۰	۰	۰

جدول (۴-۳)- مقادیر همبستگی توصیف‌کننده‌های روش GCM با دانسیته

مقدار همبستگی با دانسیته	توصیف‌کننده
۰/۳۸۵	-COOH
-۰/۳۶۶	-CH ₃
۰/۳۲۱	-CH ₂ -NH ₂
-۰/۳۴۹	T
۰/۳۱۴	-CH ₂ -OH
-۰/۲۵۶	-CH ₂ -
۰/۱۳۰	>CH-OH
-۰/۱۱۳	P
-۰/۱۱۳	>CH-NH ₂
۰/۰۵۶	>NH
-۰/۰۳۲	>CH-
۰/۰۲۹	$\begin{array}{c} \\ -C- OH \\ \end{array}$

۳-۲-۴- حذف توصیف‌کننده‌های نامناسب

با توجه به این‌که تعداد زیاد توصیف‌کننده‌های حاصل از نرم‌افزار Dragon باعث پیچیدگی محاسبات شده و تعدادی از توصیف‌کننده‌ها حاوی اطلاعات یکسانی هستند، در این مرحله از روش‌های کاهش متغیر استفاده می‌شود. به همین منظور، ابتدا توصیف‌کننده‌هایی که برای تمام

مولکول‌ها، مقادیر ثابت و یا تقریباً ثابت داشتند، حذف شدند. سپس به وسیله‌ی برنامه‌ی نوشته شده در نرم‌افزار MATLAB توصیف‌کننده‌هایی که همبستگی بزرگتر از ۰/۹ داشتند، نیز حذف گردید. در پایان این مرحله ۱۴۲ توصیف‌کننده از مجموع ۱۴۸۳ توصیف‌کننده باقی ماند.

۳-۲-۵- دسته‌بندی داده‌ها

قبل از شروع مرحله‌ی انتخاب بهترین توصیف‌کننده‌ها باید داده‌ها به سه سری آموزش، ارزیابی و تست تقسیم شوند. بنابراین در ابتدا از بین ۵۹ ترکیب مورد مطالعه، داده‌های *PVT*، ۳ ترکیب به‌طور کامل کنار گذاشته شد تا از آن‌ها در سری اعتبارسنجی بیرونی استفاده شود. سپس داده‌های *PVT* ۵۶ ترکیب باقی‌مانده به‌طور تصادفی به سه سری آموزش، ارزیابی و تست تقسیم شدند.

۳-۲-۶- انتخاب بهترین توصیف‌کننده‌ها

پس از دسته‌بندی داده‌ها، باید از بین ۱۴۲ توصیف‌کننده‌ی باقی‌مانده، بهترین توصیف‌کننده‌ها برای ساختن مدل انتخاب شوند. در این پایان‌نامه از روش خطی رگرسیون مرحله‌ای و روش غیرخطی الگوریتم ژنتیک برای انتخاب توصیف‌کننده‌های مناسب استفاده شد که در ادامه در مورد هریک از این روش‌ها توضیح داده شده است.

۳-۲-۶-۱- انتخاب بهترین توصیف‌کننده‌ها به روش رگرسیون مرحله‌ای (SR)

برای انتخاب توصیف‌کننده‌های مناسب و معتبر به این روش، ابتدا از داده‌های سری آموزش توسط نرم‌افزار SPSS رگرسیون مرحله‌ای گرفته شد. برای این منظور با انتخاب سربرگ Analyze در این نرم‌افزار، گزینه‌ی Regression و سپس گزینه‌ی Linear انتخاب و ۱۵ مدل مختلف به‌دست آمد.

سپس این ۱۵ مدل توسط سری ارزیابی مورد بررسی قرار گرفتند. میانگین مربعات خطا (MSE) و ضریب رگرسیون (R^2) و آماره F محاسبه شده برای سری ارزیابی مدل‌های ۴ تا ۱۵، در جدول (۳-۵) آمده است. مدلی که تعداد توصیف‌کننده‌های کمتر، مقدار ضریب رگرسیون و آماره F بیشتر و مقدار میانگین مربعات خطا کمتری برای سری ارزیابی داشته باشد، به عنوان بهترین مدل انتخاب می‌شود. با توجه به پارامترهای آماری به‌دست آمده در این مرحله، مدل ۸ به عنوان مدل بهینه انتخاب گردید. برای ارزیابی اهمیت و سهم هریک از توصیف‌کننده‌ها در مدل، مقدار اثر متوسط^۱ (ME) برای از توصیف‌کننده‌ها از رابطه‌ی زیر محاسبه می‌شود [۶۰]:

$$ME_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j \beta_j \sum_{i=1}^n d_{ij}} \quad (۳-۱)$$

که در این رابطه ME_j ، اثر متوسط توصیف‌کننده‌ی j ، β_j ، ضریب توصیف‌کننده‌ی j ، d_{ij} ، مقدار توصیف‌کننده برای هر مولکول و m تعداد توصیف‌کننده‌ها در مدل می‌باشد.

نام کامل توصیف‌کننده‌های وارد شده در مدل بهینه به همراه ضریب رگرسیون و اثر متوسط آن‌ها در جدول (۳-۶) آمده است. همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۳-۷) که عدم همبستگی بین آن‌ها را نشان می‌دهد و مقادیر همبستگی توصیف‌کننده‌ها با دانسیته در جدول (۳-۸) ارائه شده است. مقادیر توصیف‌کننده‌های به‌دست آمده از روش رگرسیون مرحله‌ای برای مولکول‌های مختلف موجود در سری داده‌ها در جدول (۳-۹) ذکر شده است.

1 - Mean Effect

جدول (۳-۵) - مقایسه‌ی آماره‌های سری ارزیابی مدل‌های به دست آمده از روش رگرسیون مرحله‌ای

مدل	توصیف‌کننده‌ها	R^2	MSE	F
۴	X0AV, R3m, T, P	۰/۹۲۱	۶۳۴/۳۰۳	۸۰۴/۸۶۶
۵	X0AV, R3m, T, P, RDF035e	۰/۹۵۳	۳۷۴/۲۲۵	۱۱۴۳/۹۵۷
۶	X0AV, R3m, T, P, RDF035e, GATS4e	۰/۹۵۹	۳۶۰/۱/۹۰۰	۱۲۶۹/۵۷۸
۷	X0AV, R3m, T, P, RDF035e, GATS4e, RBF	۰/۹۷۵	۳۳۳/۸۴۵	۱۶۲۶/۷۵۴
۸	X0AV, R3m, T, P, RDF035e, GATS4e, RBF, ATS2e	۰/۹۸۲	۱۳۱/۲۱۴	۱۹۸۷/۸۴۹
۹	X0AV, R3m, T, P, RDF035e, GATS4e, RBF, ATS2e, R4p	۰/۹۸۳	۱۳۱/۵۸۱	۱۸۷۸/۲۸۵
۱۰	X0AV, R3m, T, P, RDF035e, GATS4e, RBF, ATS2e, R4p, qneg_A	۰/۹۸۵	۱۱۹/۸۷۹	۱۸۴۵/۵۶۵
۱۱	X0AV, R, T, P, RDF035e, GATS4e, RBF, ATS2e, R4p, qneg_A, HATSv	۰/۹۸۶	۱۱۰/۱۴۶	۱۸۵۷/۲۸۵
۱۲	X0AV, R3m, T, P, RDF035e, GATS4e, RBF, ATS2e, R4p, qneg_A, HATSv, G2m	۰/۹۸۷	۱۰۵/۸۳۳	۱۷۷۸/۲۰۴
۱۳	X0AV, R3m, T, P, RDF035e, GATS4e, RBF, ATS2e, R4p, qneg_A, HATSv, G2m, R4m_A	۰/۸۴۵	۳۶۳۰/۰/۱۲	۱۶۴۳/۲۵۰
۱۴	X0AV, R3m, T, P, RDF035e, GATS4e, RBF, ATS2e, R4p, qneg_A, HATSv, G2m, R4m_A, IC5	۰/۹۸۷	۱۰۳/۵۹۴	۱۵۴۳/۰/۲۶
۱۵	X0AV, R3m, T, P, RDF035e, RBF, ATS2e, R4p, qneg_A, HATSv, G2m, R4m_A, IC5	۰/۹۸۷	۱۰۳/۶۰۸	۱۶۶۵/۹۰۸

جدول (۳-۶) - توصیف‌کننده‌های انتخاب شده توسط روش رگرسیون مرحله‌ای

شماره	علامت	اثر متوسط	گروه	نام کامل
۱	X0AV	-۳۰/۷۹۹	Connectivity	Average valence connectivity index of order -0
۲	R3m	۱۳/۸۹۰	GETAWAY	Autocorrelation of lag 3/weighted by atomic masses
۳	RDF035e	۱۷/۹۵۲	RDF	Radial Distribution Function-3.5/weighted by atomic Sanderson electronegativities
۴	GATS4e	-۴/۶۶۲	2D autocorrelations	Geary autocorrelation-lag4/weighted by atomic Sanderson electronegativities
۵	RBF	۳۴/۰۰۰	Constitutional	Rotatable bond fraction
۶	ATS2e	۲۳/۴۷۳	2D autocorrelation	Broto-Moreau autocorrelation of topological structure-lag2/Weighted by atomic Sanderson electronegativities
۷	T	-۶۴/۱۹۳	Experimental	Temperature
۸	P	۵۲/۷۹۲	Experimental	Pressure

جدول (۳-۷) - ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای

	X0AV	R3m	T	P	RDF035e	GATS4e	RBF	ATS2e
X0AV	۱							
R3m	۰/۲۵۲	۱						
T	۰/۱۵۹	۰/۰۳۲	۱					
P	۰/۳۹۴	۰/۰۱۹	۰/۰۹۲	۱				
RDF035e	۰/۳۴۵	۰/۴۹۹	۰/۰۶۴	۰/۱۳۲	۱			
GATS4e	۰/۰۵۶	۰/۵۲۵	۰/۰۰۶	۰/۰۶۵	۰/۲۹۷	۱		
RBF	۰/۵۸۳	۰/۲۵۲	۰/۰۲۲	-۰/۳۶۰	۰/۰۹۳	۰/۱۰۹	۱	
ATS2e	۰/۸۰۸	۰/۵۵۳	۰/۱۸۵	-۰/۲۲۶	۰/۲۷۶	۰/۰۳۲	۰/۲۷۶	۱

جدول (۳-۸) - مقادیر همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون مرحله‌ای با دانسیته

	X0AV	ATS2e	RBF	R3m	T	P	GATS4e	RDF035e
مقدار همبستگی با دانسیته	-۰/۸۶۵	۰/۸۶۵	۰/۶۰۰	۰/۵۴۵	۰/۳۴۳	-۰/۱۶۴	۰/۱۴۹	۰/۰۰۶

جدول (۳-۹) - مقادیر توصیف کننده های انتخاب شده به روش SR

	نام مولکول	X0Av	RBF	ATS2e	R3m	RDF035e	GATS4e
۱	اتانول	۰/۷۱۸	۰/۲۵۰	۱/۰۰۵	۰/۰۴۰	۱/۸۵۲	۰
۲	اتان دیول	۰/۵۷۷	۰/۳۳۰	۱/۰۸۱	۰/۱۴۰	۴/۴۹۸	۱/۹۷۷
۳	۱-پروپانول	۰/۷۱۵	۰/۲۷۰	۰/۹۸۶	۰/۰۷۰	۲/۵۳۳	۳/۱۰۶
۴	۲-پروپانول	۰/۷۵۶	۰/۲۷۰	۰/۹۸۷	۰/۰۶۰	۴/۶۴۵	۰
۵	۲-متیل-۱-پروپانول	۰/۷۴۶	۰/۲۹۰	۰/۹۷۶	۰/۱۲۰	۵/۷۷۰	۳/۰۷۳
۶	۲-متیل-۲-پروپانول	۰/۷۸۹	۰/۲۹۰	۰/۹۷۸	۰/۰۸۰	۸/۶۷۶	۰
۷	۱ و ۲-پروپان دیول	۰/۶۳۶	۰/۳۳۰	۱/۰۴۱	۰/۱۹۰	۲/۷۵۸	۲/۰۸۳
۸	۱-بوتانول	۰/۷۱۴	۰/۲۹۰	۰/۹۷۶	۰/۰۵۰	۲/۶۴۴	۲/۰۷۶
۹	۲-بوتانول	۰/۷۴۶	۰/۲۹۰	۰/۹۷۷	۰/۱۱۰	۴/۵۱۲	۲/۰۰۱
۱۰	۲-متیل-۱-بوتانول	۰/۷۴۰	۰/۲۹۰	۰/۹۷۰	۰/۱۱۰	۶/۸۱۲	۲/۷۸۵
۱۱	۱ و ۲-بوتان دیول	۰/۶۴۸	۰/۳۳۰	۱/۰۱۹	۰/۲۰۰	۳/۴۸۲	۲/۶۲۴
۱۲	۱-پنتانول	۰/۷۱۳	۰/۲۹۰	۰/۹۷۰	۰/۰۷۰	۲/۵۹۳	۱/۷۷۳
۱۳	۲-پنتانول	۰/۷۴۰	۰/۲۹۰	۰/۹۷۱	۰/۰۲۰	۵/۱۵۵	۱/۶۱۲
۱۴	۳-پنتانول	۰/۷۴۰	۰/۲۹۰	۰/۹۷۱	۰/۱۳۰	۸/۷۷۳	۳/۴۲۷
۱۵	۲-متیل-۳-پنتانول	۰/۷۶۶	۰/۳۰۰	۰/۹۶۷	۰/۱۵۰	۱۰/۴۰۷	۲/۵۱۰
۱۶	۳-اتیل-۳-پنتانول	۰/۷۵۹	۰/۳۰۰	۰/۹۶۴	۰/۱۷۰	۱۴/۰۰۱	۳/۶۱۹
۱۷	۲-متیل-۲ و ۴-پنتان دیول	۰/۷۱۰	۰/۳۳۰	۰/۹۶۶	۰/۰۸۰	۱۳/۷۶۰	۰/۶۹۳
۱۸	۱-هگزانول	۰/۷۳۵	۰/۳۰۰	۰/۹۶۶	۰/۰۵۰	۴/۵۷۳	۱/۵۰۱
۱۹	۲-هگزانول	۰/۷۳۵	۰/۳۰۰	۰/۹۶۶	۰/۰۵۰	۴/۵۷۳	۱/۵۰۱
۲۰	۱-هپتانول	۰/۷۱۱	۰/۳۰۰	۰/۹۶۳	۰/۰۳۰	۲/۵۹۳	۱/۵۲۳
۲۱	۲-هپتانول	۰/۷۳۲	۰/۳۰۰	۰/۹۶۳	۰/۰۵۰	۴/۴۵۸	۱/۴۳۴
۲۲	۱-اوکتانول	۰/۷۱۱	۰/۳۱۰	۰/۹۶۱	۰/۰۳۰	۲/۵۹۳	۱/۴۶۳
۲۳	۱-نونانول	۰/۷۱۰	۰/۳۱۰	۰/۹۵۹	۰/۰۳۰	۲/۵۹۳	۱/۴۲۳
۲۴	۱-دکانول	۰/۷۱۰	۰/۳۱۰	۰/۹۵۷	۰/۰۳۰	۲/۵۹۲	۲/۳۹۵
۲۵	۱-اندکانول	۰/۷۱۰	۰/۳۱۰	۰/۹۵۶	۰/۰۳۰	۲/۵۹۲	۱/۳۷۵
۲۶	۱-دودکانول	۰/۷۱۰	۰/۳۲۰	۰/۹۵۵	۰/۰۳۰	۲/۵۹۲	۱/۳۶۱
۲۷	۱-تترادکانول	۰/۷۰۹	۰/۳۲۰	۰/۹۵۴	۰/۰۳۰	۲/۵۹۲	۱/۳۴۵
۲۸	۱-هگزادکانول	۰/۷۰۹	۰/۳۲۰	۰/۹۵۳	۰/۰۲۰	۲/۵۹۳	۱/۳۳۸
۲۹	پروپیل آمین	۰/۷۴۸	۰/۲۵۰	۰/۹۵۷	۰/۰۵۰	۱/۰۷۵	۲/۶۲۰
۳۰	n-بوتیل آمین	۰/۷۴۰	۰/۲۷۰	۰/۹۵۴	۰/۰۴۰	۱/۰۹۳	۱/۸۷۱

ادامه جدول (۳-۹)

	نام مولکول	X0Av	RBF	ATS2e	R3m	RDF035e	GATS4e
۳۱	۲-آمینو بوتان	۰/۷۷۲	۰/۲۷۰	۰/۹۵۴	۰/۰۸۰	۳/۸۱۳	۰/۷۵۱
۳۲	۲-آمینو پنتان	۰/۷۶۱	۰/۲۸۰	۰/۹۵۳	۰/۰۶۰	۳/۵۸۵۰	۱/۴۳۴
۳۳	۲-آمینوهپتان	۰/۷۴۸	۰/۲۹۰	۰/۹۵۰	۰/۰۴۰	۳/۹۲۳	۱/۳۹۴
۳۴	۲-آمینواکتان	۰/۷۴۳	۰/۳۰۰	۰/۹۵۰	۰/۰۳۰	۴/۵۵۵	۱/۳۹۳
۳۵	پنتیل آمین	۰/۷۳۴	۰/۲۸۰	۰/۹۵۲	۰/۰۲۰	۱/۰۹۳	۱/۶۹۴
۳۶	هگزیل آمین	۰/۷۳۰	۰/۲۹۰	۰/۹۵۱	۰/۰۳۰	۱/۰۹۳	۱/۶۰۶
۳۷	هپتیل آمین	۰/۷۲۷	۰/۲۹۰	۰/۹۵۰	۰/۰۲۰	۱/۰۶۷	۱/۵۵۸
۳۸	اکتیل آمین	۰/۷۲۵	۰/۳۰۰	۰/۹۴۹	۰/۰۲۰	۱/۰۶۸	۱/۵۳۲
۳۹	دسیل آمین	۰/۷۲۲	۰/۳۰۰	۰/۹۴۸	۰/۰۲۰	۱/۰۹۳	۱/۵۱۲
۴۰	دودسیل آمین	۰/۷۲۰	۰/۳۱۰	۰/۹۴۸	۰/۰۲۰	۱/۰۶۸	۱/۵۱۳
۴۱	دی اتیل آمین	۰/۷۸۳	۰/۲۷۰	۰/۹۶۹	۰/۰۲۰	۱/۲۵۸	۰/۴۶۵
۴۲	دی پروپیل آمین	۰/۸۰۸	۰/۲۹۰	۰/۹۶۲	۰/۰۳۰	۷/۶۰۲	۰/۳۹۲
۴۳	دی بوتیل آمین	۰/۷۴۹	۰/۳۰۰	۰/۹۵۸	۰/۰۲۰	۱/۷۳۱	۲/۵۵۰
۴۴	دی هگزیل آمین	۰/۷۳۶	۰/۳۱۰	۰/۹۵۳	۰/۰۲۰	۱/۷۳۱	۲/۲۸۳
۴۵	دی اکتیل آمین	۰/۷۲۹	۰/۳۱۰	۰/۹۵۱	۰/۰۲۰	۱/۷۳۱	۲/۱۷۳
۴۶	۱-آمینو-۲-پروپانول	۰/۶۶۲	۰/۳۱۰	۱/۰۰۸	۰/۱۴۰	۳/۰۳۴	۱/۰۰۸
۴۷	۳-آمینو-۱-پروپانول	۰/۶۲۹	۰/۳۱۰	۱/۰۰۷	۰/۰۵۰	۱/۸۷۰	۱/۹۰۳
۴۸	۲-آمینو-۱-متیل-۱-پروپانول	۰/۶۹۷	۰/۳۱۰	۰/۹۹۵	۰/۱۸۰	۶/۱۷۶	۲/۵۳۷
۴۹	دی اتانول آمین	۰/۶۰۳	۰/۳۵۰	۱/۰۴۵	۰/۰۴۰	۳/۰۱۶	۱/۴۵۰
۵۰	استیک اسید	۰/۵۸۹	۰/۲۹۰	۱/۰۸۸	۰/۰۹۰	۰/۵۴۰	۰
۵۱	پروپانویک اسید	۰/۶۱۳	۰/۳۰۰	۱/۰۳۴	۰/۱۲۰	۲/۴۸۰	۴/۳۹۰
۵۲	بوتانویک اسید	۰/۶۲۸	۰/۳۱۰	۱/۰۱۰	۰/۰۷۰	۲/۴۸۹	۲/۴۹۰
۵۳	پنتانویک اسید	۰/۶۴۰	۰/۳۱۰	۰/۹۹۶	۰/۰۵۰	۲/۵۰۹	۲
۵۴	هگزانویک اسید	۰/۶۴۸	۰/۳۲۰	۰/۹۸۷	۰/۰۴۰	۲/۶۴۷	۱/۷۵۹
۵۵	هپتانویک اسید	۰/۶۵۵	۰/۳۲۰	۰/۹۸۰	۰/۰۴۰	۲/۶۴۷	۱/۶۱۸
۵۶	اکتانویک اسید	۰/۶۶۰	۰/۳۲۰	۰/۹۷۶	۰/۰۵۰	۲/۶۴۷	۱/۵۲۶
۵۷	دکانویک اسید	۰/۶۶۸	۰/۳۲۰	۰/۹۶۹	۰/۰۵۰	۲/۶۴۷	۱/۴۱۷
۵۸	دودکانویک اسید	۰/۶۷۳	۰/۳۲۰	۰/۹۶۵	۰/۰۵۰	۰/۶۴۷	۱/۳۵۷
۵۹	تترادکانویک اسید	۰/۶۷۸	۰/۳۳۰	۰/۹۶۲	۰/۰۵۰	۲/۶۴۷	۱/۳۲۱
۶۰	هگزادکانویک اسید	۰/۶۸۱	۰/۳۳۰	۰/۹۶۰	۰/۰۵۰	۲/۶۴۷	۱/۲۹۹

۲-۳-۲- انتخاب توصیف‌کننده‌های معتبر به روش الگوریتم ژنتیک (GA)

در روش انتخاب توصیف‌کننده‌های نهایی به روش الگوریتم ژنتیک، نیازی به مرحله‌ی رمزگشایی نیست زیرا الگوریتم ژنتیک فقط نقش انتخاب توصیف‌کننده را ایفا می‌کند، بنابراین تمام ژن‌ها تک‌بیتی در نظر گرفته شده و هر ژن معرف یک توصیف‌کننده است. به منظور اجرای برنامه‌ی الگوریتم ژنتیک، تمام توصیف‌کننده‌های باقی‌مانده به عنوان ورودی به این برنامه داده شدند. برنامه M-file استفاده شده در این تحقیق در بسته‌ی نرم‌افزاری MATLAB دارای مقادیر پیش‌فرض احتمال ترکیب ۰/۵، احتمال جهش ۰/۰۱، اندازه‌ی جمعیتی برابر با ۳۰ و با تکرار ۵۰۰ می‌باشد. این برنامه ۱۲ بار تکرار گردید تا احتمال انتخاب توصیف‌کننده‌های تصادفی به حداقل برسد. توصیف‌کننده‌هایی که بیش از ۵۰٪ در این الگوریتم تکرار شده‌اند عبارتند از: دما، فشار، X0AV، ATS2e و RBF. این توصیف‌کننده‌ها قبلاً در توصیف‌کننده‌های حاصل از رگرسیون مرحله‌ای نیز بودند بنابراین، نام کامل و مقدار آن‌ها در این‌جا تکرار نمی‌شود. تعداد دفعات تکرار توصیف‌کننده‌ها به ازای ۵۰۰ با تکرار، در ۱۲ بار اجرای الگوریتم ژنتیک در جدول (۳-۱۰) گزارش شده است. همچنین مقادیر همبستگی این توصیف‌کننده‌های انتخابی با دانسیته در جدول (۳-۱۱) ارائه شده است.

جدول (۳-۱۰) - تعداد تکرار توصیف‌کننده‌های حاصل از GA به ازای ۱۲ بار اجرا

دفعات اجرا	تعداد تکرار توصیف‌کننده به ازای اجرای ۵۰۰ بار تکرار											
	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
T	۴۵۹	۴۴۴	۴۳۸	۴۴۵	۴۴۶	۴۴۲	۴۳۹	۴۴۱	۳۹۸	۳۶۵	۳۷۷	۳۸۲
P	۴۰۲	۳۹۵	۳۸۹	۳۹۸	۴۰۷	۴۰۲	۳۹۸	۳۹۲	۳۱۲	۳۱۶	۳۴۳	۳۴۳
X0Av	۳۶۸	۳۸۳	۳۷۶	۳۷۸	۴۰۶	۳۸۶	۳۷۷	۳۸۸	۲۸۷	۲۹۸	۲۹۲	۲۷۰
ATS2e	۲۹۱	۲۹۸	۲۸۲	۲۸۳	۲۹۳	۲۸۲	۲۷۰	۳۹۲	۲۷۳	۲۴۶	۲۴۸	۲۲۹
RBF	۲۷۲	۲۶۴	۲۵۸	۲۷۲	۲۵۱	۲۵۶	۲۵۸	۲۷۹	۱۹۶	۲۲۹	۱۹۵	۲۰۸

جدول (۳-۱۱) - مقادیر همبستگی توصیف‌کننده‌های انتخابی روش GA با دانسیته

	X0Av	ATS2e	RBF	T	P
مقدار همبستگی با دانسیته	-۰/۸۵۱	۰/۸۴۹	۰/۵۹۰	۰/۳۴۹	-۰/۱۱۳

۳-۲-۷- مدل‌سازی غیرخطی

در این مرحله از تحقیق به منظور انجام مدل‌سازی غیرخطی، از شبکه‌ی عصبی مصنوعی استفاده شد. برای مدل‌سازی از توصیف‌کننده‌های مرتب شده به روش CR حاصل از سه روش SR، GA و GCM به عنوان ورودی استفاده گردید که در ادامه راجع به مدل‌سازی با هریک از این روش‌ها توضیحاتی ارائه خواهد شد.

۳-۲-۷-۱- مدل‌سازی غیرخطی با استفاده از شبکه عصبی مصنوعی

با توجه به پیچیدگی داده‌ها و احتمال وجود روابط غیرخطی بین متغیرهای مستقل و متغیر وابسته، روش غیرخطی شبکه‌ی عصبی مصنوعی به عنوان تکنیکی برتر در مطالعات QSPR برای مدل‌سازی دانسیته مورد استفاده قرار گرفت. در تمامی مطالعات مربوط به شبکه‌ی عصبی از دو تابع آموزشی لونیبرگ-مارکوارت (trainlm) و بایزین (trainbr) و دو تابع انتقال لگاریتم سیگموئیدی (logsig) و تانژانت سیگموئیدی (tansig) استفاده گردید. به این ترتیب با ترکیب دو تابع آموزشی با دو تابع انتقال گفته شده، چهار شبکه عصبی مصنوعی طراحی گردید که هر کدام جداگانه بهینه شود. در هریک از شبکه‌های طراحی شده، توصیف‌کننده‌هایی که با روش‌های مختلف انتخاب شده بودند به عنوان ورودی به شبکه داده شدند. بنابراین در هریک از شبکه‌ها، به ازای هر توصیف‌کننده‌ی انتخابی یک نرون وجود دارد. لایه‌ی خروجی نیز شامل یک نرون می‌باشد که دانسیته‌ی متناظر با ورودی‌ها را نشان می‌دهد. در هر کدام از شبکه‌های طراحی شده، تعداد نرون لایه‌ی پنهان، تعداد ورودی شبکه، تعداد دور آموزشی و تعداد دورهای آموزشی باید بهینه شوند. مقدار بهینه برای هریک از این کمیت‌ها

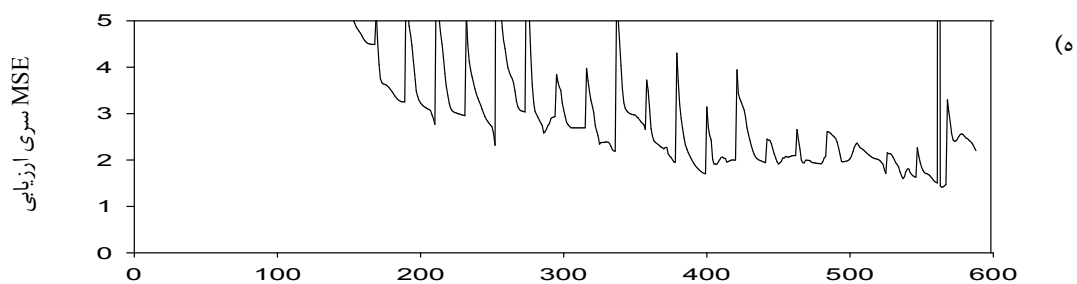
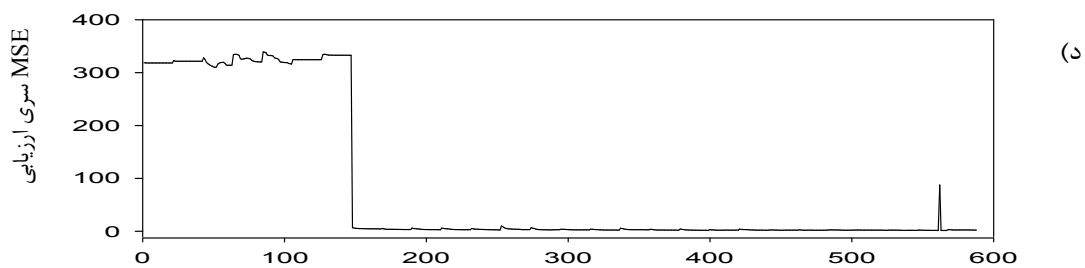
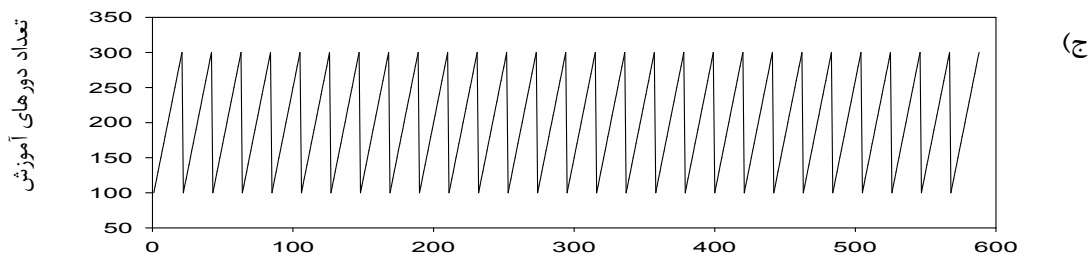
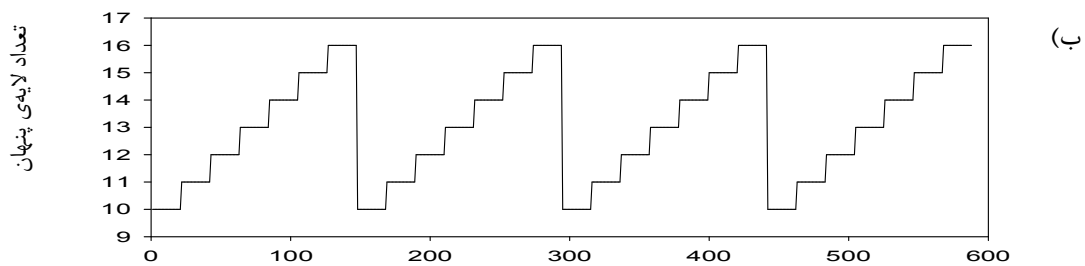
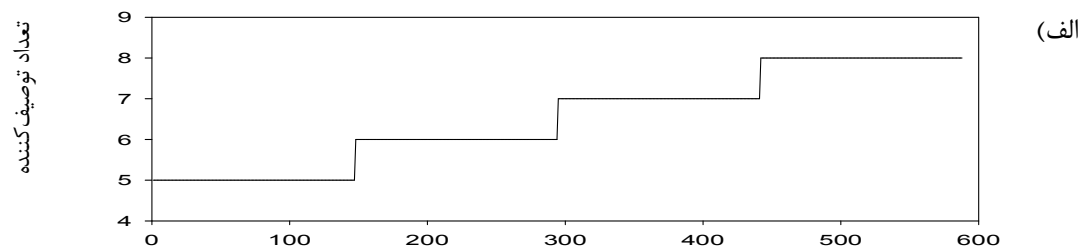
براساس کمترین مقدار میانگین مربعات خطا به عنوان تابع عملکرد شبکه برای سری ارزیابی انتخاب شد.

در ادامه، چگونگی مدل سازی شبکه عصبی مصنوعی با توصیف کننده های بهینه ی به دست آمده از روش رگرسیون مرحله ای، الگوریتم ژنتیک و سهم گروه توضیح داده خواهد شد.

۳-۲-۷-۱-الف - بهینه سازی شبکه ی عصبی مصنوعی با استفاده از

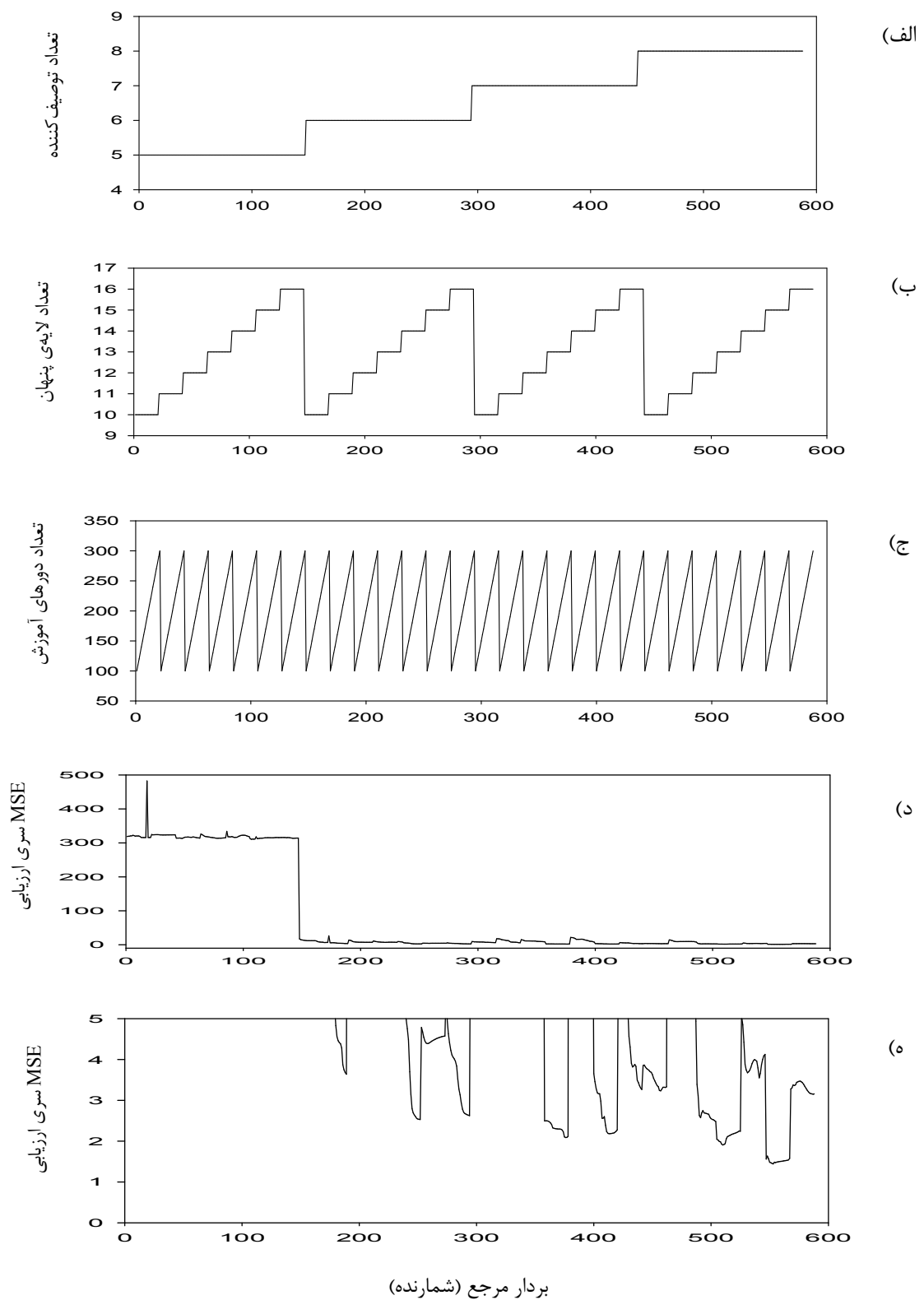
توصیف کننده های حاصل از رگرسیون مرحله ای (SR-ANN)

شبکه ی طراحی شده در این پروژه، شبکه ای با سه لایه شامل یک لایه ی ورودی، یک لایه ی پنهان و یک لایه ی خروجی است که تعداد گره های لایه ی ورودی برابر توصیف کننده های بهینه ی حاصل از روش رگرسیون مرحله ای یعنی تعداد ۸ توصیف کننده ی ذکر شده در بخش (۳-۲-۵-۱) می باشد. لایه ی خروجی نیز دارای یک نرون است که نشان دهنده ی دانسیته ی متناظر با هر ورودی می باشد. برای بهینه سازی تعداد ورودی های شبکه، تعداد گره های لایه ی پنهان و تعداد دور آموزشی از روش بهینه سازی همزمان استفاده گردید. برای این منظور تعداد ورودی های شبکه از ۲ تا ۸ با گام ۱، تعداد گره ی لایه ی پنهان شبکه از ۲ تا ۱۶ با گام ۱ و تعداد دور آموزشی شبکه از ۱۰ تا ۳۰۰ با گام ۱۰ انتخاب گردید و به ازای همه ی ترکیبات ممکن این سه پارامتر، چهار شبکه ی عصبی مصنوعی طراحی شده به طور جداگانه بهینه گردید. در لایه ی خروجی تمام شبکه های مورد مطالعه از تابع انتقال خطی (pureline) استفاده گردید. به دلیل این که رسم نمودارهای مربوط به کل محدوده ی تغییرات پارامترها بسیار شلوغ و غیر واضح خواهد شد، بنابراین قسمتی از روند تغییرات پارامترهای شبکه در حین بهینه سازی همزمان پارامترها به همراه مقادیر MSE به دست آمده برای سری ارزیابی حول مقادیر بهینه برحسب یک بردار مرجع در شکل های (۳-۱) تا (۳-۴) نشان داده شده است. بردار مرجع، تعداد ترکیب های ممکن سه پارامتر، تعداد توصیف کننده، تعداد لایه ی پنهان و تعداد دور آموزشی را نشان می دهد.

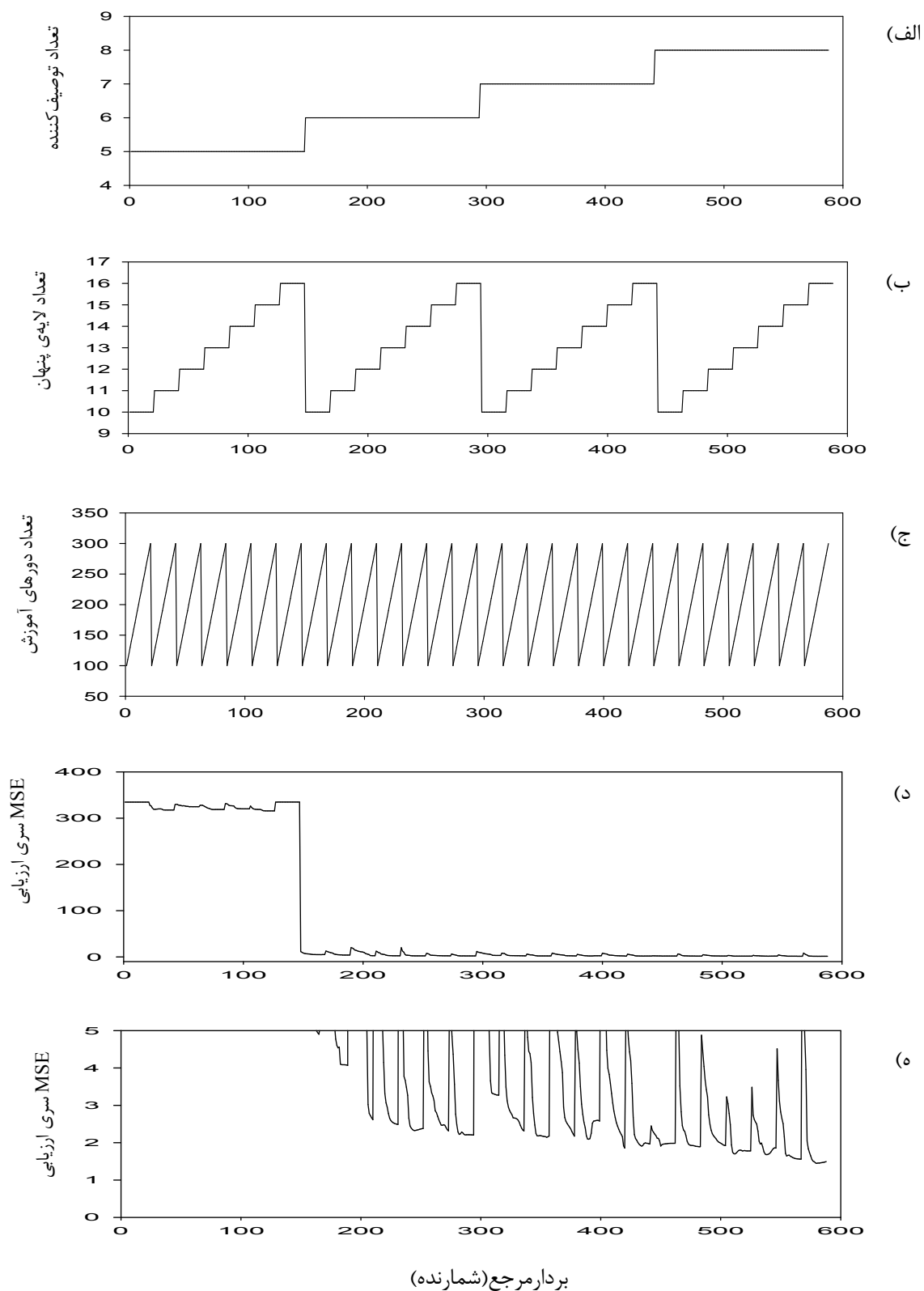


بردار مرجع (شمارنده)

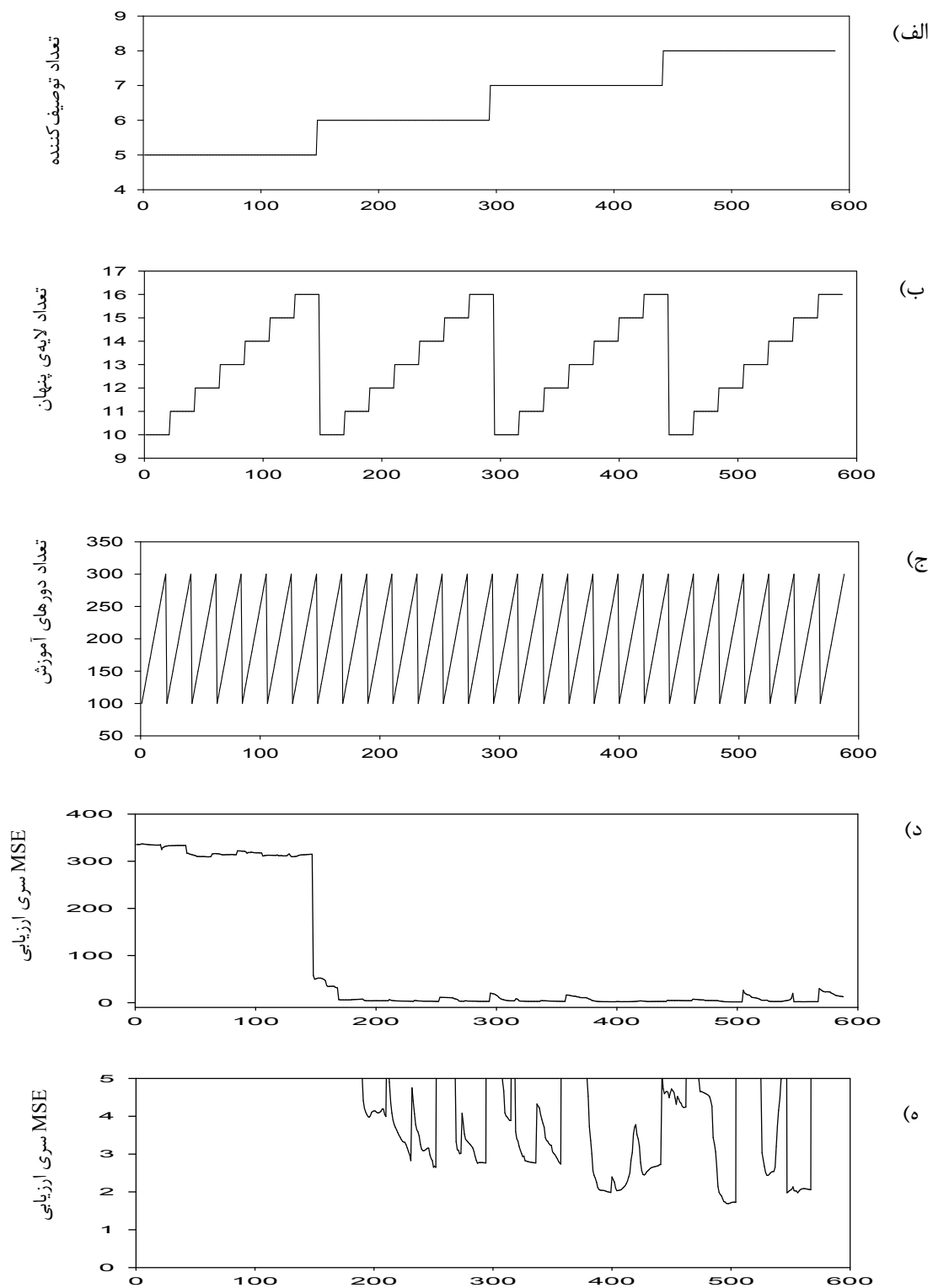
شکل (۳-۱) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) سری MSE ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف کننده‌های حاصل از SR



شکل (۳-۲) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت و با استفاده از توصیف‌کننده‌های حاصل از SR



شکل (۳-۳) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف کننده‌های حاصل از SR



بردار مرجع (شمارنده)

شکل (۳-۴) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت و با استفاده از توصیف‌کننده‌های حاصل از SR

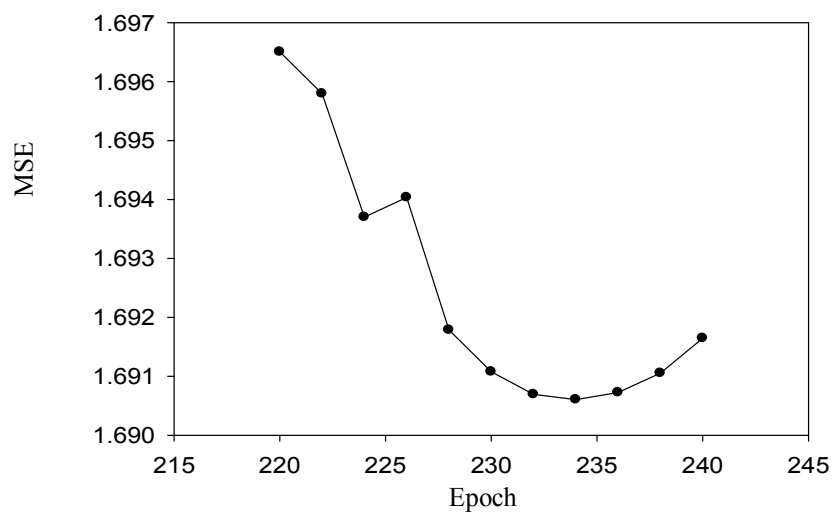
مقادیر بهینه‌ی به‌دست آمده برای هریک از شبکه‌های طراحی شده در جدول (۳-۱۲) گردآوری شده است. نتایج موجود در این جدول نشان می‌دهد که MSE هر چهار شبکه‌ی بهینه شده تقریباً یکسان هستند. بنابراین شبکه‌ای که پیچیدگی کمتری دارد به عنوان شبکه‌ی بهینه انتخاب می‌شود. برای این منظور شبکه‌ای با تابع آموزشی تانژانت سیگموئیدی و تابع انتقال لونیگ-مارکوارت با ۸ توصیف‌کننده در لایه‌ی ورودی و ۱۲ گره در لایه‌ی پنهان و دور آموزشی ۲۳۰ پیچیدگی کمتری دارد.

جدول (۳-۱۲)- توابع و پارامترهای بهینه‌ی شبکه‌های طراحی شده با توصیف‌کننده‌های حاصل از SR

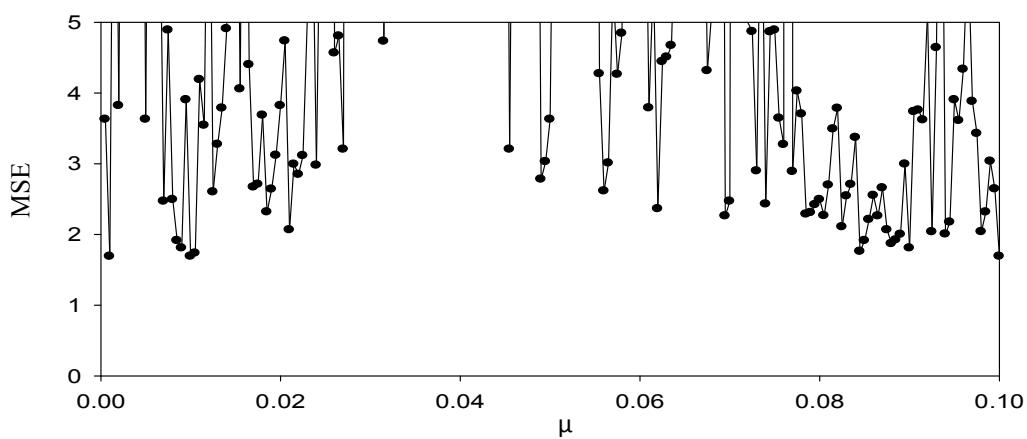
تابع انتقال	تابع آموزش	تعداد توصیف‌کننده	تعداد نرون لایه‌ی پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموئید	تنظیم بایزین	۸	۱۵	۲۷۰	۱/۴۱۲
لگاریتم سیگموئید	لونیگ-مارکوارت	۸	۱۵	۱۸۰	۱/۴۴۴
تانژانت سیگموئید	تنظیم بایزین	۸	۱۶	۲۲۰	۱/۴۴۶
تانژانت سیگموئید	لونیگ-مارکوارت	۸	۱۲	۲۳۰	۱/۶۹۱

برای به‌دست آوردن مقدار بهینه‌ی دقیق دور آموزشی، شبکه‌ای با همان مقادیر بهینه‌ی تعداد نرون لایه‌ی ورودی، تعداد نرون لایه‌ی پنهان، تابع آموزش و تابع انتقال اما با تعداد دورهای آموزشی ۲۲۰ تا ۲۴۰ با گام ۲ مورد بررسی قرار گرفت. که طبق شکل (۳-۵) دور آموزشی ۲۳۴ کمترین مقدار MSE را دارد و به عنوان تعداد دور آموزشی بهینه انتخاب شد.

جهت بهینه کردن مقدار پارامتر ممنوم، شبکه‌ی بهینه را با مقدار μ از ۰/۰۰۰۵ تا ۰/۱ با گام ۰/۰۰۰۵ انتخاب شد. نتایج این بررسی نشان داد که مقدار ۰/۰۱ دارای کمترین مقدار MSE بوده و به عنوان مقدار بهینه‌ی μ انتخاب شد. نمودار میانگین مربعات خطا برای سری ارزیابی برحسب μ حول نقطه بهینه در شکل (۳-۶) رسم شده است.



شکل (۳-۵) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب تعداد دور آموزشی



شکل (۳-۶) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ

با توجه به روند بهینه‌سازی ارائه شده، مقدار پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های منتخب حاصل از روش رگرسیون مرحله‌ای (SR-ANN) در جدول (۳-۱۳) ارائه شده است.

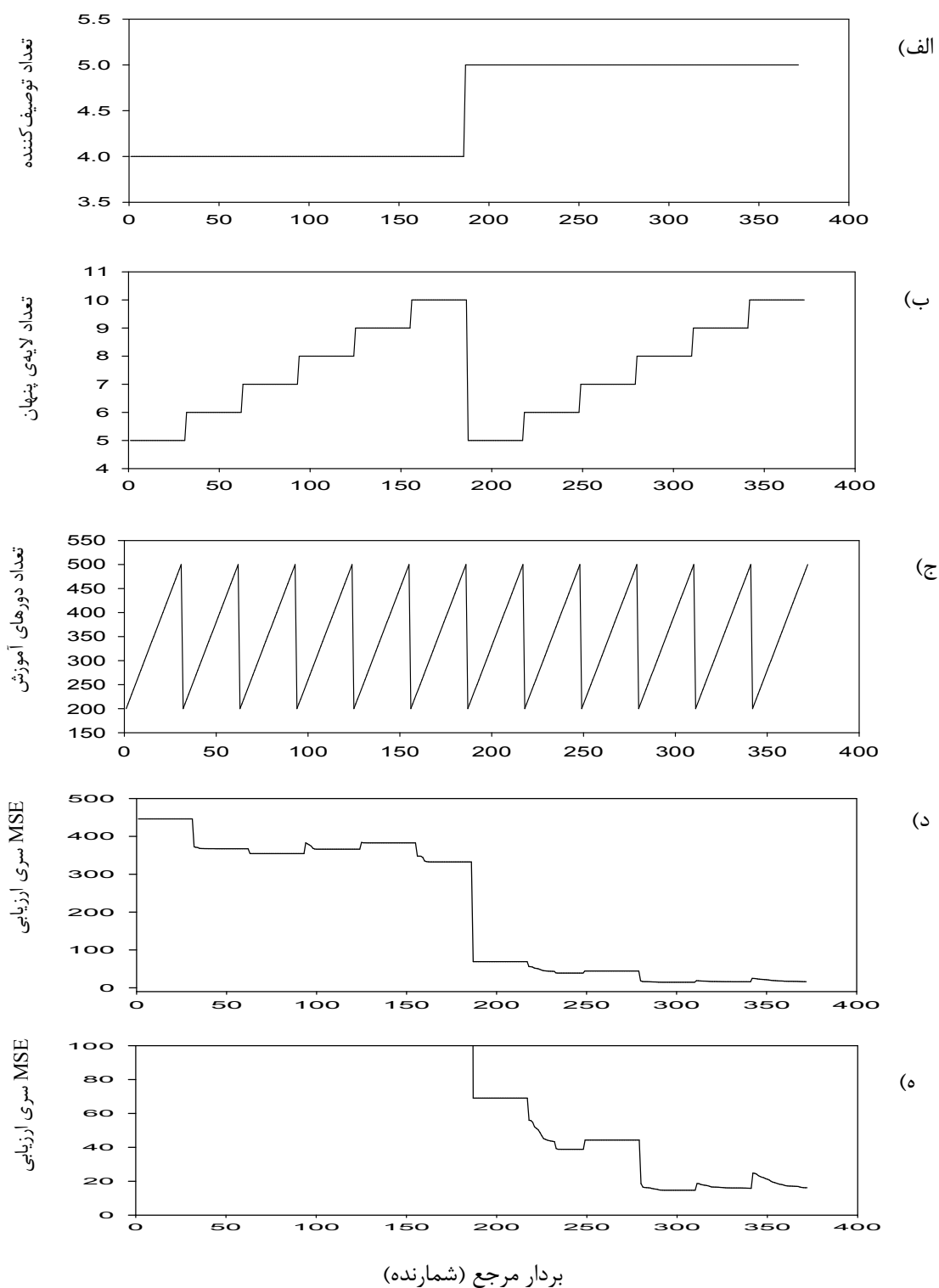
جدول (۳-۱۳) - توابع و مقادیر پارامترهای بهینه‌ی شبکه‌ی (SR-ANN)

تابع آموزش	Train lm
تابع انتقال لایه‌ی پنهان	tansig
تابع انتقال لایه‌ی خروجی	pure line
تعداد نرون لایه‌ی پنهان	۱۲
تعداد متغیرهای ورودی	۸
تعداد دورهای آموزش	۲۳۴
پارامتر μ	۰/۰۱
مقدار MSE سری ارزیابی	۱/۶۹۱

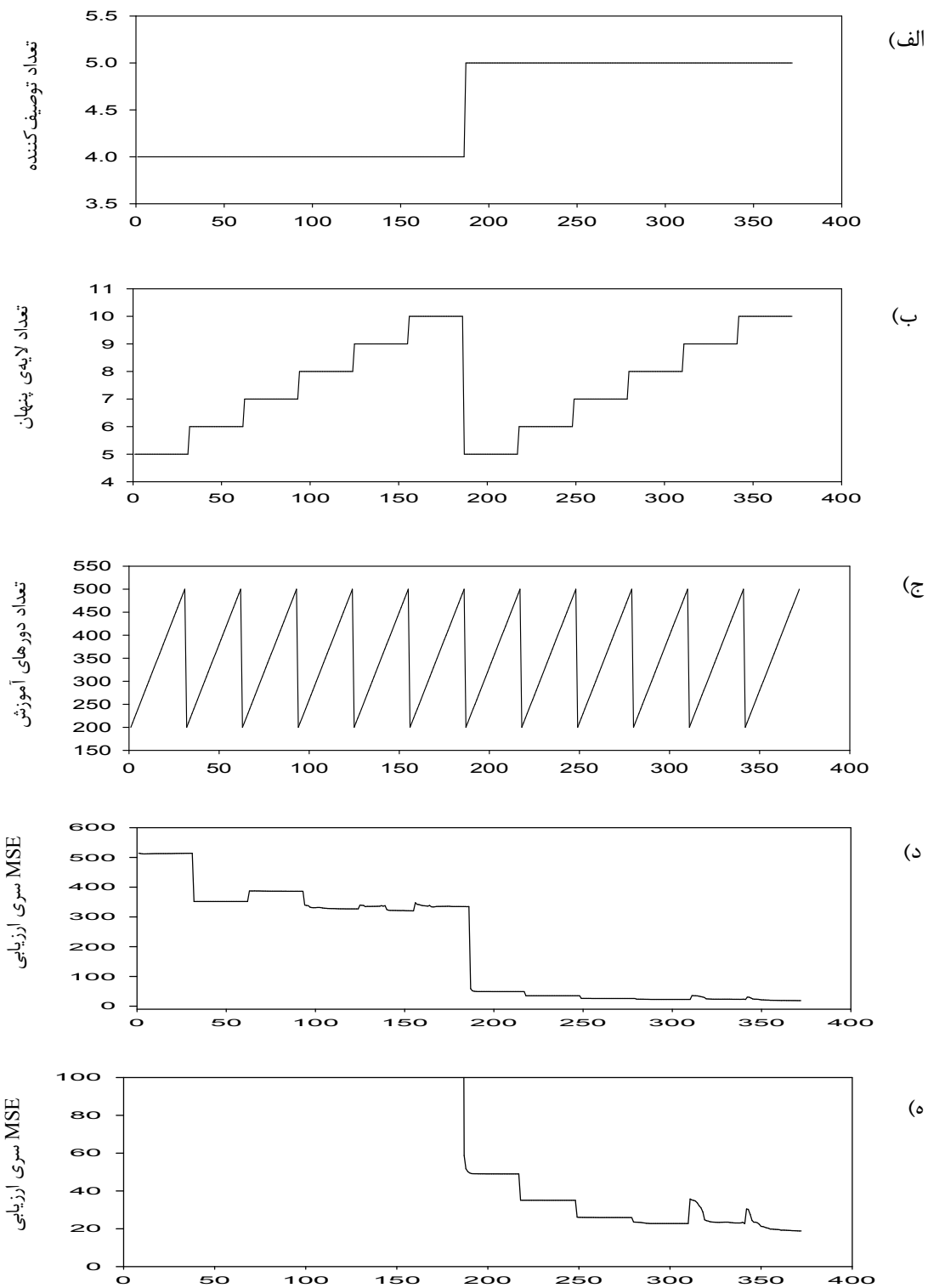
۳-۲ - ۷-۱-ب - بهینه‌سازی مدل شبکه‌ی عصبی مصنوعی با توصیف‌کننده‌های

الگوریتم ژنتیک (GA-ANN)

شبکه‌ی عصبی مورد مطالعه در این بخش نیز، شبکه‌ای سه لایه می‌باشد. به منظور به حداقل رساندن تابع عملکرد MSE در سری ارزیابی، شبکه‌هایی با مقدار μ ثابت، تعداد ورودی ۲ تا ۵ متغیرهای مستقل ذکر شده در بخش (۳-۲-۵)، تعداد متفاوت گره‌های لایه‌ی پنهان از ۲ تا ۱۰ و دور آموزشی متغیر ۱۰ تا ۵۰۰ با گام ۱۰ طراحی گردید. شبکه‌های طراحی شده به ازای همه‌ی ترکیب‌های ممکن از این سه پارامتر بهینه‌سازی گردید که نتایج این بررسی‌ها را تنها حول نقطه‌ی بهینه برای هریک از شبکه‌های طراحی شده، در شکل‌های (۳-۷) تا (۳-۱۰) نشان داده شده است.



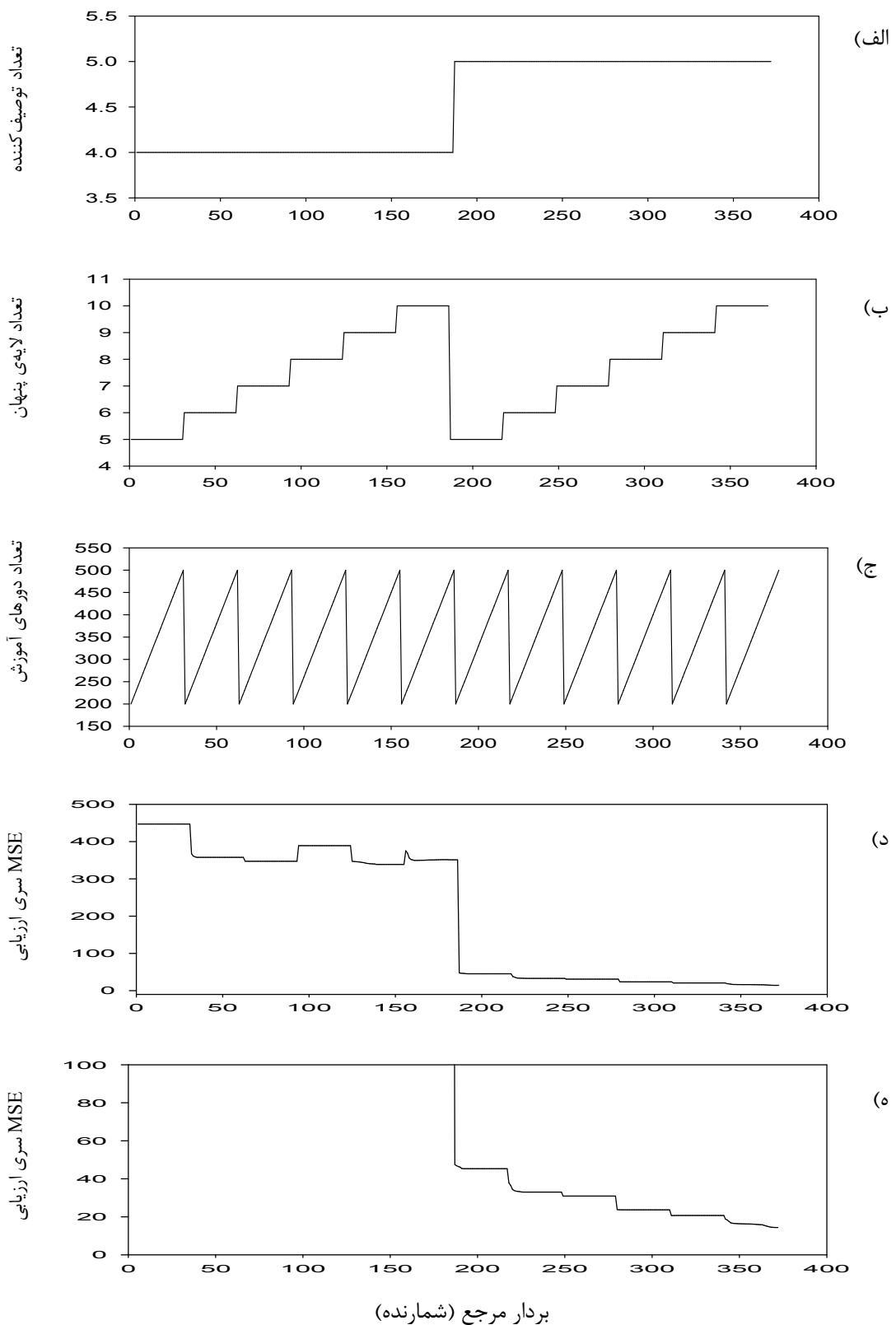
شکل (۳-۷) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف کننده‌های حاصل از GA



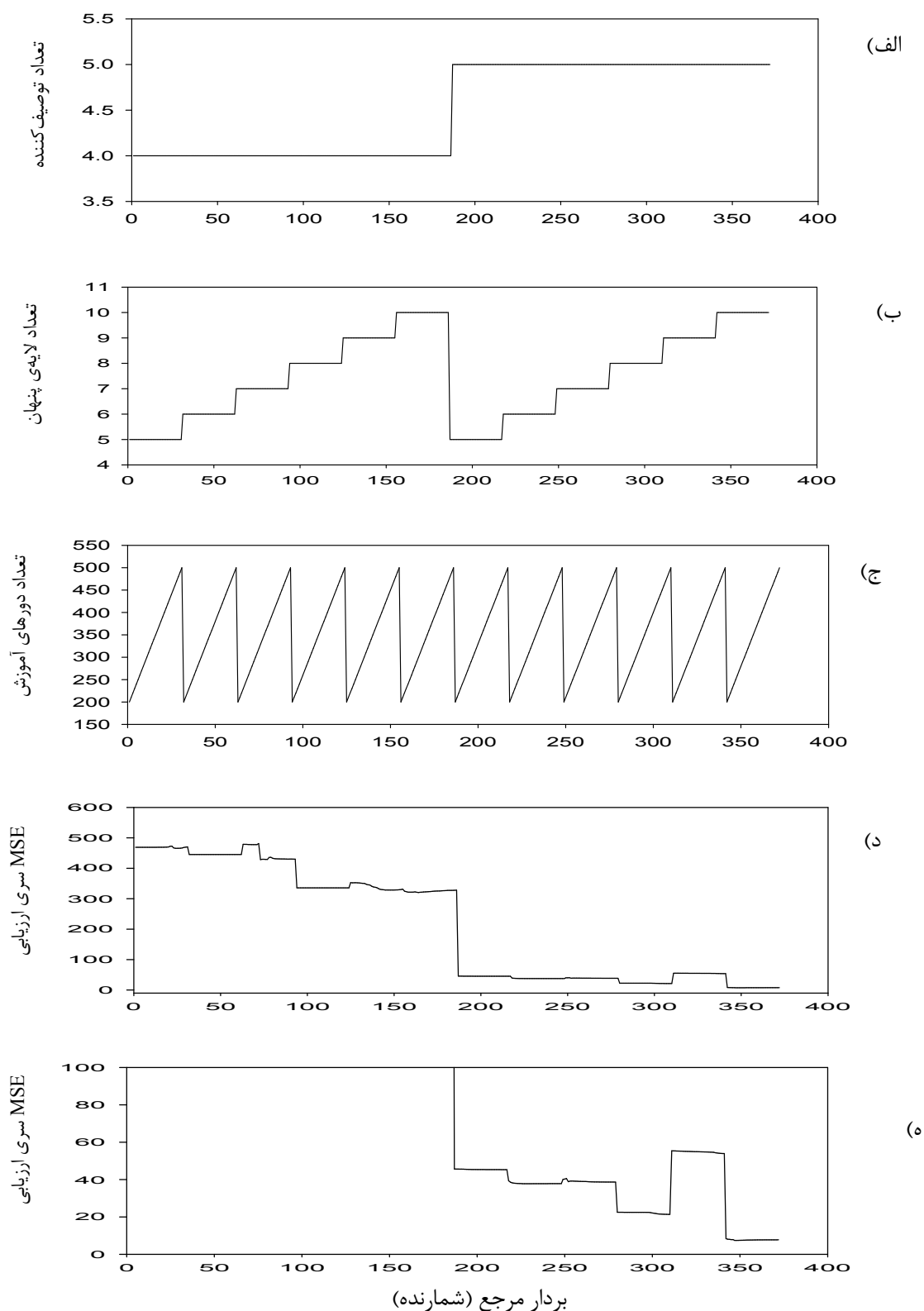
بردار مرجع (شمارنده)

شکل (۳-۸) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال لگاریتم سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت و با استفاده از توصیف کننده‌های حاصل از

GA



شکل (۳-۹) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه‌ی مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی بایزین و با استفاده از توصیف‌کننده‌های حاصل از GA



شکل (۳-۱۰) - نمودارهای (الف) تعداد توصیف کننده، (ب) تعداد نرون لایه مخفی، (ج) تعداد دور آموزش، (د) MSE سری ارزیابی و (ه) ناحیه‌ای از نمودار MSE سری ارزیابی که مینیمم آن را بهتر نشان می‌دهد، بر حسب بردار مرجع برای شبکه‌ی عصبی مصنوعی با تابع انتقال تانژانت سیگموئیدی، الگوریتم آموزشی لونبرگ-مارکوارت با استفاده از توصیف کننده‌های حاصل از

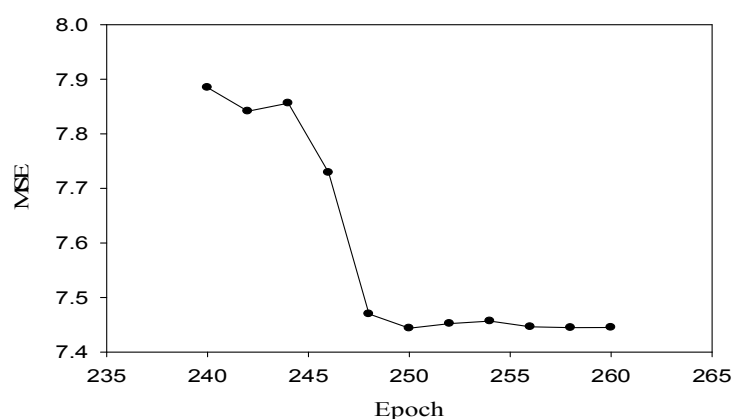
GA

توابع و پارامترهای بهینه شده‌ی شبکه‌های GA-ANN در جدول (۳-۱۴) گزارش شده است. نتایج موجود در این جدول نشان می‌دهد که شبکه‌ای با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئیدی با ۵ توصیف‌کننده در لایه‌ی ورودی، ۱۰ گره در لایه‌ی پنهان و دور آموزشی بهینه‌ی ۲۵۰ کمترین مقدار MSE را دارد و به عنوان شبکه‌ی بهینه انتخاب گردید.

جدول (۳-۱۴) - توابع و پارامترهای بهینه‌ی شبکه‌های GA-ANN

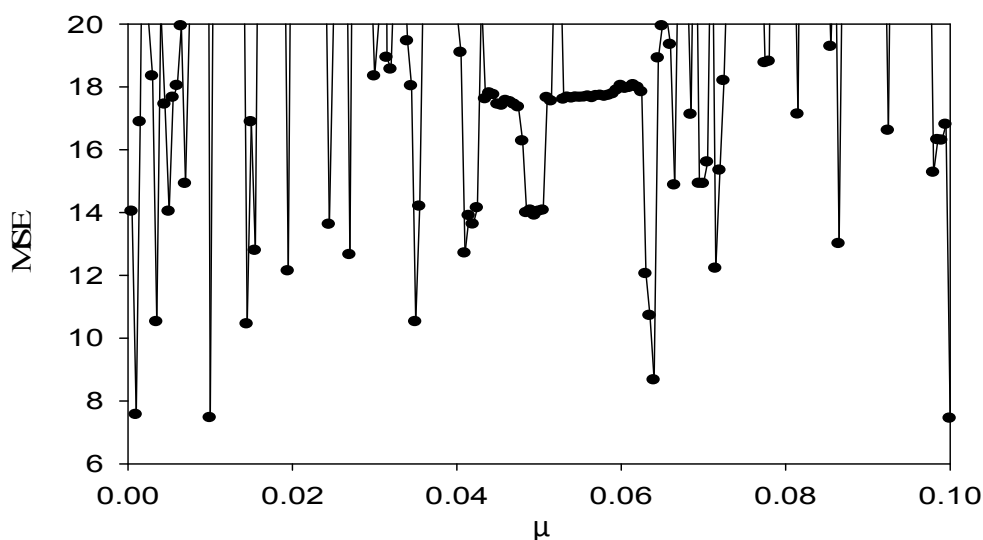
MSE	تعداد دور آموزش	تعداد نرون لایه‌ی پنهان	تعداد توصیف‌کننده	تابع آموزش	تابع انتقال
۱۴/۶۸۲	۳۸۰	۸	۵	تنظیم بایزین	لگاریتم سیگموئید
۱۸/۸۲۸	۴۶۰	۱۰	۵	لونبرگ-مارکوارت	لگاریتم سیگموئید
۱۴/۳۶۵	۵۰۰	۱۰	۵	تنظیم بایزین	تانژانت سیگموئید
۷/۴۴۳	۲۵۰	۱۰	۵	لونبرگ-مارکوارت	تانژانت سیگموئید

برای تعیین دقیق‌تر دور آموزش، شبکه‌ای با همان مقادیر بهینه تعداد نرون لایه ورودی، تعداد لایه‌ی پنهان، تابع آموزش و تابع انتقال اما با تعداد دورهای ۲۴۰ تا ۲۶۰ با گام ۲ مورد بررسی قرار گرفت که طبق شکل (۳-۱۱) دور آموزشی ۲۵۰ با مقدار MSE قبلی، ۷/۴۴۳ به عنوان بهترین مقدار برای تعداد دور آموزشی در شبکه‌ی بهینه انتخاب گردید.



شکل (۳-۱۱) - نمودار میانگین مربعات خطا برای سری ارزیابی بر حسب تعداد دورهای آموزش متفاوت با گام ۲

در ادامه بهینه کردن شبکه مورد بررسی، مقدار پارامتر μ نیز بایستی بهینه شود که بدین منظور، در ساختار شبکه‌ی بهینه با ۵ متغیر ورودی، ۱۰ گره در لایه‌ی پنهان، الگوریتم آموزشی لونبرگ-مارکوارت، تابع انتقال تانژانت سیگموئیدی و دور آموزشی ۲۵۰، مقدار μ از ۰/۰۰۰۵ تا ۰/۱ با گام‌های ۰/۰۰۰۵ تغییر داده شد و آن‌گاه برای هر مورد مقدار میانگین مربع خطا برای سری ارزیابی محاسبه و ثبت گردید که مقدار ۰/۰۱ به عنوان مقدار بهینه‌ی μ انتخاب شد. نمودار مقدار میانگین مربع خطا بر حسب μ حول نقطه‌ی بهینه در شکل (۳-۱۲) رسم شده است.



شکل (۳-۱۲)- نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ

توابع و مقادیر پارامترهای بهینه شده شبکه‌ی عصبی برای پیش بینی دانسیته‌ی سیالات تجمعی، در جدول (۳-۱۵) ارائه شده است:

جدول (۳-۱۵) - توابع و مقادیر پارامترهای بهینه شده‌ی GA-ANN

تابع آموزش	Train lm
تابع انتقال لایه‌ی پنهان	tansig
تابع انتقال لایه‌ی خروجی	pure line
تعداد نرون لایه‌ی پنهان	۱۲
تعداد متغیرهای ورودی	۵
تعداد دوره‌های آموزش	۲۵۰
پارامتر μ	۰/۰۱
مقدار MSE سری ارزیابی	۷/۴۴۳۷

۳-۲-۷-۱-ج - مدل‌سازی شبکه‌ی عصبی مصنوعی با استفاده از

توصیف‌کننده‌های حاصل از روش سهم گروه (GCM-ANN)

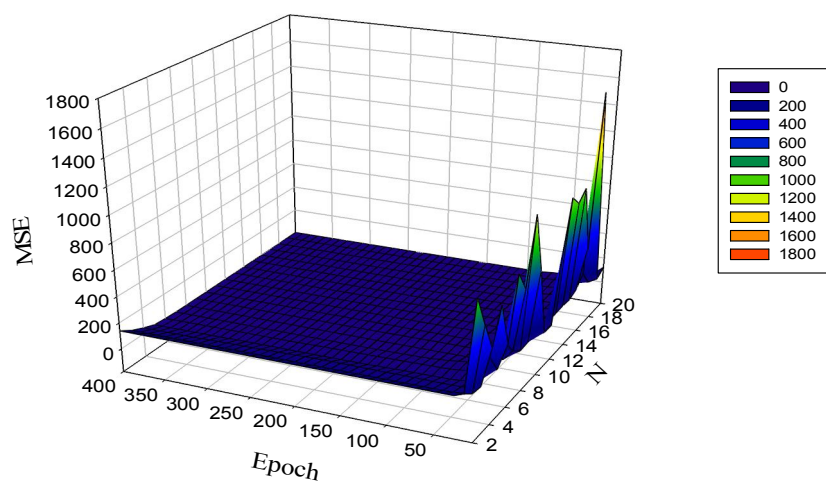
همان‌طور که در قسمت (۲-۳-۲-۳) گفته شد، ۱۰ توصیف‌کننده بر مبنای سهم گروه‌ها به همراه دو متغیر تجربی دما و فشار به عنوان توصیف‌کننده‌های ورودی مدل انتخاب شدند. لایه‌ی خروجی نیز یک نرون دارد که دانسیته‌ی متناظر با ورودی‌ها را نشان می‌دهد. برای آموزش شبکه دو الگوریتم آموزشی تنظیم بایزین و لونبرگ-مارکوارت و برای انتخاب تابع انتقال مناسب دو تابع انتقال تانژانت سیگموئید و لگاریتم سیگموئید مورد استفاده قرار گرفتند، که از ترکیب این دو الگوریتم و دو تابع انتقال، چهار نوع شبکه‌ی عصبی مختلف طراحی گردید. سپس برای هر یک از شبکه‌ها، از روش تغییر هم‌زمان برای تغییر تعداد نرون‌های لایه‌ی مخفی و تعداد دور آموزش جهت آموزش شبکه استفاده گردید و به ازای همه‌ی ترکیب‌های ممکن از این دو پارامتر شبکه‌های طراحی شده بهینه‌سازی شدند.

در تمامی این شبکه‌ها مقدار μ ثابت در نظر گرفته شد و تعداد توصیف‌کننده‌ها ثابت و برابر ۱۲ و تعداد نرون‌های لایه‌ی مخفی از ۲ تا ۲۰ و دور آموزشی از ۱۰ تا ۴۰۰ با گام ۱۰ در نظر گرفته شدند. در این قسمت نیز مانند بخش گذشته به دلیل تعداد زیاد داده‌های به دست آمده بخشی از نتایج حاصل از این بررسی‌ها برای سری ارزیابی در جداول (۱۶-۳) تا (۱۹-۳) گردآوری شده است و نتایج حاصل از کل محاسبات در نمودارهای سه بعدی (۱۳-۳) تا (۱۶-۳) آمده است.

جدول (۳-۱۶) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی تنظیم بایزین و تابع

انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت

		تعداد گره‌ی لایه‌ی پنهان								
		۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
دور آموزشی	۱۵۰	۲/۵۳۲	۲/۰۴۳	۱/۶۸۹	۲/۵۲۰	۱/۷۹۴	۱/۶۹۵	۱/۵۲۶	۱/۵۱۰	۲/۲۱۸
	۱۶۰	۲/۴۶۹	۲/۰۸۳	۱/۶۰۰	۲/۴۴۳	۱/۷۸۴	۱/۷۵۱	۱/۵۳۱	۱/۴۶۸	۲/۲۳۹
	۱۷۰	۲/۳۹۷	۲/۱۲۷	۱/۵۶۵	۲/۳۷۳	۱/۷۵۶	۱/۷۶۰	۱/۵۶۶	۱/۴۲۹	۲/۲۵۹
	۱۸۰	۲/۳۲۸	۲/۱۵۹	۱/۵۳۵	۲/۲۸۳	۱/۷۳۷	۱/۷۴۳	۱/۵۸۶	۱/۳۹۶	۲/۲۲۲
	۱۹۰	۲/۲۷۱	۲/۱۷۶	۱/۵۱۳	۲/۱۷۷	۱/۷۲۳	۱/۷۲۵	۱/۵۷۱	۱/۳۹۷	۲/۱۷۱
	۲۰۰	۲/۲۳۰	۲/۱۳۷	۱/۵۰۳	۲/۰۸۹	۱/۷۰۸	۱/۷۰۹	۱/۵۱۶	۱/۴۴۸	۲/۰۹۶
	۲۱۰	۲/۲۱۱	۲/۱۱۲	۱/۴۸	۲/۰۲۹	۱/۶۸۹	۱/۶۸۳	۱/۴۴۷	۱/۵۱۹	۱/۹۸۴
	۲۲۰	۲/۲۱۹	۲/۰۹۹	۱/۴۹۷	۱/۹۸۱	۱/۶۷۰	۱/۶۶۳	۱/۴۰۲	۱/۵۸۴	۱/۸۳۵
	۲۳۰	۲/۱۸۵	۲/۰۵۲	۱/۵۰۲	۱/۹۰۶	۱/۶۶۷	۱/۶۵۵	۱/۳۸۴	۱/۶۲۳	۱/۷۰۶
	۲۴۰	۲/۱۴۹	۱/۹۹۱	۱/۵۲۱	۱/۷۹۴	۱/۶۷۹	۱/۶۶۱	۱/۳۸۰	۱/۶۲۲	۱/۶۳۶
	۲۵۰	۲/۱۲۱	۱/۹۳۵	۱/۵۳۹	۱/۷۴۵	۱/۶۹۸	۱/۶۸۴	۱/۳۸۰	۱/۶۱۷	۱/۶۲۶
	۲۶۰	۲/۰۹۹	۱/۸۸۹	۱/۵۵۱	۱/۷۴۶	۱/۷۱۹	۱/۷۱۵	۱/۳۷۸	۱/۶۴۰	۱/۶۵۱
	۲۷۰	۲/۰۸۰	۱/۸۵۲	۱/۵۵۵	۱/۸۲۸	۱/۷۴۵	۱/۷۴۳	۱/۳۷۳	۱/۶۶۷	۱/۶۷۶
	۲۸۰	۲/۰۶۳	۱/۸۱۹	۱/۵۵۶	۱/۹۴۱	۱/۷۷۱	۱/۷۵۳	۱/۳۶۵	۱/۶۹۳	۱/۶۹۱
	۲۹۰	۲/۰۴۶	۱/۷۸۹	۱/۵۶۵	۱/۹۴۹	۱/۷۸۲	۱/۷۴۹	۱/۳۵۶	۱/۷۱۱	۱/۶۹۸
	۳۰۰	۲/۰۲۴	۱/۷۶۷	۱/۶۰۴	۱/۹۴۰	۱/۷۸۱	۱/۷۲۱	۱/۳۴۸	۱/۷۲۳	۱/۷۰۳
	۳۱۰	۱/۷۹۳	۱/۷۵۵	۱/۵۵۷	۲/۰۲۸	۱/۷۷۴	۱/۷۰۷	۱/۳۹۹	۱/۷۳۴	۱/۷۱۲
	۳۲۰	۱/۶۸۵	۱/۷۴۶	۱/۵۵۳	۲/۰۲۰	۱/۷۶۲	۱/۷۰۰	۱/۴۸۶	۱/۸۶۴	۱/۷۳۴
	۳۳۰	۱/۶۷۳	۱/۷۴۴	۱/۵۵۱	۱/۹۶۳	۱/۷۵۴۵	۱/۶۹۴	۱/۵۰۲	۱/۳۲	۱/۷۸۵
	۳۴۰	۱/۷۰۴	۱/۷۴۲	۱/۵۵۱	۱/۹۳۰	۱/۷۲۸	۱/۶۸۸	۱/۴۹۸	۲/۰۲۳	۱/۸۴۷
	۳۵۰	۱/۷۰۱	۱/۷۳۹	۱/۵۵۰	۱/۹۱۵	۱/۷۱۰	۱/۶۷۳	۱/۴۸۵	۲/۱۳۶	۱/۹۰۰

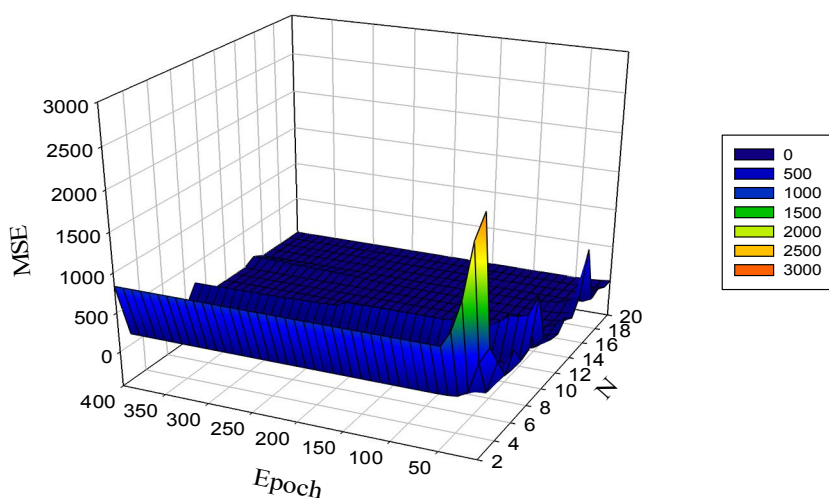


شکل (۳-۱۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۶)

جدول (۳-۱۷) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی تنظیم

لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت

		تعداد گره‌ی لایه‌ی پنهان								
		۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
دور آموزشی	۱۵۰	۲/۸۵۵	۱/۷۸۷	۱/۵۶۱	۳/۸۶۱	۳/۹۰۲	۱/۷۷۱	۲/۵۴۰	۱/۶۹۵	۱/۷۹۹
	۱۶۰	۲/۷۲۴	۱/۷۴۸	۱/۶۰۳	۳/۰۳۱	۳/۸۱۱	۱/۹۰۹	۲/۵۶۳	۱/۶۴۴	۱/۷۵۴
	۱۷۰	۲/۵۹۳	۱/۷۱۴	۱/۶۳۲	۳/۲۲۴	۳/۶۹۳	۲/۰۵۰	۲/۵۸۴	۱/۶۰۲	۱/۴۴۷
	۱۸۰	۲/۵۰۶	۱/۶۸۹	۱/۵۹۷	۳/۳۵۵	۳/۳۷۲	۲/۱۴۰	۲/۵۸۶	۱/۵۸۰	۱/۵۵۱
	۱۹۰	۲/۴۵۵	۱/۶۷۲	۱/۵۳۹	۳/۷۳۸	۳/۲۶۰	۲/۱۹۶	۲/۵۶۵	۱/۵۷۰	۱/۶۵۰
	۲۰۰	۲/۴۰۶	۱/۶۶۵	۱/۵۰۳	۴/۲۸۵	۲/۹۹۷	۲/۲۹۰	۲/۵۲۳	۱/۵۶۵	۱/۶۹۸
	۲۱۰	۲/۳۳۴	۱/۶۶۲	۱/۵۲۶	۶/۳۹۲	۲/۸۴۴	۲/۵۴۷	۲/۴۵۶	۱/۵۶۱	۱/۸۳۸
	۲۲۰	۲/۲۳۴	۱/۶۵۸	۱/۵۴۱	۸/۴۸۴	۲/۶۸۳	۲/۸۶۷	۲/۳۶۱	۱/۵۵۷	۱/۹۵۹
	۲۳۰	۲/۱۴۹	۱/۶۵۳	۱/۵۴۵	۱۱/۷۴۱	۲/۶۶۳	۳/۲۶۲	۲/۲۰۳	۱/۵۵۲	۱/۹۷۹
	۲۴۰	۲/۰۴۸	۱/۶۴۷	۱/۵۳۵	۱۳/۵۰۹	۲/۳۹۴	۳/۴۷۰	۲/۰۶۲	۱/۵۴۷	۲/۱۲۳
	۲۵۰	۱/۸۶۳	۱/۶۴۲	۱/۵۲۱	۱۵/۱۰۷	۲/۲۴۹	۳/۶۳۲	۲/۰۰۹	۱/۵۴۱	۲/۲۷۶
	۲۶۰	۱/۷۳۲	۱/۶۳۸	۱/۵۰۸	۲۵/۱۲۸	۲/۱۲۷	۳/۷۳۸	۲/۰۰۵	۱/۵۳۳	۲/۴۷۲
	۲۷۰	۱/۶۴۹	۱/۶۳۵	۱/۴۸۳	۳۴/۹۱۰	۲/۱۲۰	۳/۸۶۲	۲/۰۰۸	۱/۵۲۶	۲/۶۹۸
	۲۸۰	۱/۵۶۶	۱/۶۳۵	۱/۴۵۶	۳۶/۸۵۰	۲/۲۱۰	۳/۸۴۲	۲/۰۰۳	۱/۵۱۹	۲/۷۸۶
	۲۹۰	۱/۴۴۴	۱/۶۴۲	۱/۴۵۴	۴۵/۰۴۷	۲/۰۷۸	۳/۲۷۷	۱/۹۸۷	۱/۵۱۱	۲/۷۹۷
	۳۰۰	۱/۳۸۳	۱/۶۳۳	۱/۴۶۱	۴۶/۲۵۹	۲/۰۷۶	۳/۳۵۸	۱/۹۹۳	۱/۵۰۴	۲/۷۷۵
	۳۱۰	۱/۳۵۴	۱/۶۳۴	۱/۴۷۵	۴۹/۰۶۴	۲/۰۶۲	۳/۵۷۷	۱/۰۲۳	۱/۴۹۷	۲/۷۴۸
	۳۲۰	۱/۳۴۰	۱/۶۴۴	۱/۵۰۱	۴۹/۳۵۳	۲/۰۵۱	۳/۷۱۹	۱/۰۶۳	۱/۴۹۱	۲/۷۲۱
	۳۳۰	۱/۳۲۸	۱/۶۴۴	۱/۵۴۳	۵۵/۹۳۳	۲/۰۵۴	۳/۸۵۴	۲/۱۰۱	۱/۴۸۹	۲/۶۹۰
	۳۴۰	۱/۳۰۰	۱/۶۹۱	۱/۶۰۵	۶۱/۶۹۲	۲/۰۰۵	۳/۸۷۷	۲/۱۲۸	۱/۴۹۱	۲/۶۶۱
	۳۵۰	۱/۳۱۱	۱/۶۸۶	۱/۶۸۶	۵۷/۳۱۳	۲/۰۷۳	۳/۹۳۰	۲/۱۴۶	۱/۴۹۵	۲/۶۳۵

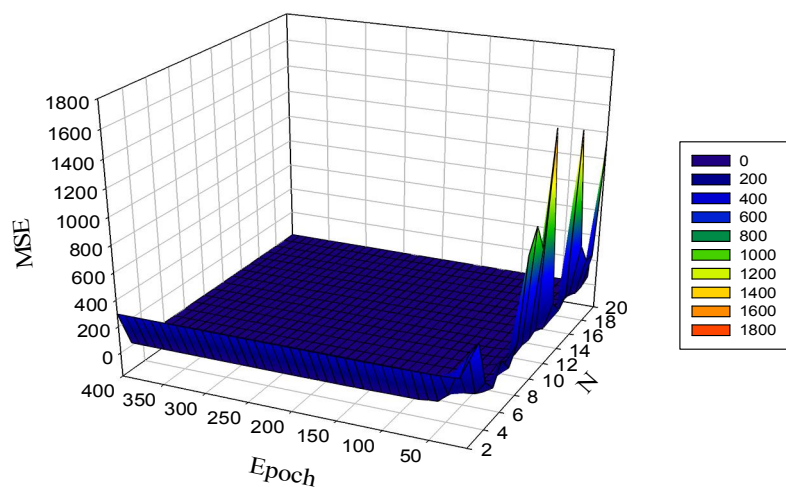


شکل (۳-۱۴) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۷)

جدول (۳-۱۸) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی تنظیم بایزین و تابع انتقال

تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت

		تعداد گره‌ی لایه‌ی پنهان								
		۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
دور آموزشی	۱۵۰	۵/۸۰۰	۲/۲۷۱	۲/۳۲۸	۱/۷۷۱	۱/۴۸۴	۱/۶۵۲	۱/۴۰۳	۱/۸۱۹	۱/۷۶۸
	۱۶۰	۵/۵۶۸	۲/۱۳۱	۲/۱۰۰	۱/۸۴۷	۱/۵۳۷	۱/۵۴۷	۱/۱۵۳	۱/۷۲۹	۱/۸۰۳
	۱۷۰	۵/۳۵۴	۲/۰۱۸	۱/۹۲۰	۱/۷۶۸	۱/۵۸۳	۱/۴۹۷	۱/۱۰۵	۱/۷۷۱	۱/۸۳۳
	۱۸۰	۵/۱۲۶	۱/۹۳۷	۱/۷۸۷	۱/۷۹۲	۱/۶۵	۱/۵۸۷	۱/۱۱۱	۱/۷۵۹	۱/۸۳۹
	۱۹۰	۴/۸۷۵	۱/۹۰۵	۱/۶۸۸	۱/۷۸۸	۱/۹۵۱	۱/۶۲۲	۱/۲۳۸	۱/۷۵۲	۱/۸۱۴
	۲۰۰	۴/۶۰۶	۱/۹۸۳	۱/۶۱۳	۱/۷۸۰	۱/۸۳۸	۱/۵۹۴	۱/۵۴۲	۱/۷۴۵	۱/۷۱۶
	۲۱۰	۴/۳۴۰	۱/۹۵۸	۱/۵۵۶	۱/۷۷۷	۱/۷۸۷	۱/۶۰۳	۱/۶۰۴	۱/۷۳۲	۱/۷۵۱
	۲۲۰	۴/۶۴۲	۱/۷۸۱	۱/۵۲۱	۱/۷۸۲	۱/۴۱۳	۱/۵۹۹	۱/۵۵۸	۱/۷۱۷	۱/۷۷۴
	۲۳۰	۲/۷۴۲	۱/۶۵۳	۱/۵۰۸	۱/۷۸۷	۱/۲۵۷	۱/۵۸۸	۱/۴۶۸	۱/۷۰۹	۱/۸۴۳
	۲۴۰	۲/۵۶۰	۱/۵۳۲	۱/۵۱۴	۱/۷۸۷	۱/۳۷۸	۱/۵۷۴	۱/۴۲۸	۱/۷۶۷	۱/۹۵۸
	۲۵۰	۲/۲۳۸	۱/۴۷۶	۱/۷۳۱	۱/۷۷۹	۱/۵۳۵	۱/۵۵۷	۱/۳۸۶	۱/۴۶۹	۲/۱۲۷
	۲۶۰	۲/۱۹۸	۱/۴۴۱	۱/۹۱۵	۱/۷۶۲	۱/۷۹۸	۱/۵۳۷	۱/۳۶۲	۱/۴۶۶	۲/۱۸۶
	۲۷۰	۱/۸۷۴	۱/۴۱۸	۱/۷۳۸	۱/۷۴۱	۲/۲۲۹	۱/۵۱۵	۱/۳۷۳	۱/۶۲۴	۲/۳۵۳
	۲۸۰	۱/۶۲۵	۱/۳۹۲	۱/۵۸۰	۱/۷۱۸	۲/۴۱	۱/۴۹۲	۱/۴۸۵	۲/۶۶۱	۲/۵۸۳
	۲۹۰	۱/۴۶۴	۱/۳۵۵	۱/۵۲۱	۱/۶۷	۲/۴۶۸	۱/۴۶۸	۱/۶۵۱	۳/۰۷	۲/۹۰۸
	۳۰۰	۱/۵۰۳	۱/۳۴۷	۱/۵۷۵	۱/۶۷۶	۲/۵۲۲	۱/۴۴۶	۱/۷۸۱	۳/۱۸۲	۲/۹۹۴
	۳۱۰	۱/۶۴۵	۱/۳۴۹	۱/۵۶۹	۱/۴۳۵	۲/۵۴۲	۱/۴۲۳	۱/۸۶۸	۲/۷۲۴	۳/۰۸۰
	۳۲۰	۱/۷۲۷	۱/۳۵۸	۱/۷۶۱	۱/۱۶۵	۲/۵۸۳	۱/۳۹۸	۱/۹۰۳	۲/۳۴۳	۲/۸۶۸
	۳۳۰	۱/۷۸۷	۱/۴۰۳	۱/۷۳۸	۱/۲۷۲	۲/۵۲۹	۱/۳۷۰	۱/۹۱۱	۱/۹۶۴	۲/۸۱۱
	۳۴۰	۱/۸۱۹	۱/۵۵۷	۱/۶۷۶	۱/۳۶۶	۲/۵۳۱	۱/۳۴۳	۱/۹۱۵	۱/۶۵۵	۲/۸۲۸
	۳۵۰	۱/۸۴۶	۱/۶۸۴	۱/۸۲۰	۱/۴۵۴	۲/۵۳۷	۱/۳۱۸	۱/۹۲۹	۱/۴۹۶	۲/۹۴۳

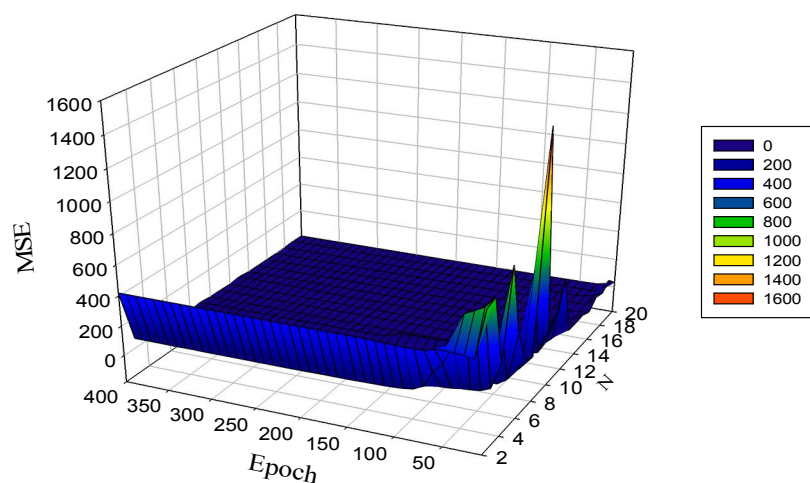


شکل (۳-۱۵) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۸)

جدول (۳-۱۹) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی تنظیم

لونبرگ - مارکوارت و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت

		تعداد گرهی لایه پنهان								
		۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
دور آموزشی	۱۵۰	۲/۰۹۰	۲/۶۲۳	۲/۵۶۹	۱/۵۹۵	۱/۱۲۹	۱/۱۷۴	۱/۵۰۵	۱/۳۶۶	۱/۴۵۰
	۱۶۰	۲/۰۶۶	۲/۷۱۱	۲/۴۶۱	۱/۵۳۷	۱/۱۱۲	۱/۱۵۶	۱/۴۱۹	۱/۳۴۴	۱/۴۹۱
	۱۷۰	۲/۰۵۴	۲/۷۶۶	۲/۳۴۵	۱/۴۸۷	۱/۱۰۱	۱/۱۵۵	۱/۳۸۲	۱/۳۳۸	۱/۵۲۹
	۱۸۰	۲/۰۵۲	۲/۸۶۰	۲/۰۷۶	۱/۵۰۰	۱/۰۹۶	۱/۱۳۷	۱/۳۹۵	۱/۳۴۵	۱/۵۶۵
	۱۹۰	۲/۰۵۹	۲/۹۶۶	۱/۸۹۳	۱/۵۵۳	۱/۰۹۴	۱/۱۳۵	۱/۴۴۳	۱/۳۵۰	۱/۶۰۲
	۲۰۰	۲/۰۶۲	۳/۰۰۸	۱/۸۱۹	۱/۶۵۹	۱/۰۹۰	۱/۱۴۱	۱/۵۰۳	۱/۳۴۷	۱/۵۲۱
	۲۱۰	۲/۰۶۶	۳/۰۸۰	۴/۸۰۱	۱/۷۲۷	۱/۰۸۴	۱/۱۴۰	۱/۵۷۵	۱/۳۳۷	۱/۳۷۳
	۲۲۰	۲/۰۷۲	۳/۱۶۲	۱/۷۴۷	۱/۴۷۹	۱/۰۷۴	۱/۱۳۱	۱/۶۷۳	۱/۳۴۰	۱/۲۲۲
	۲۳۰	۲/۰۷۹	۳/۳۶۸	۱/۶۹۸	۱/۳۱۸	۱/۰۵۵	۱/۱۱۷	۱/۷۷۷	۱/۳۴۲	۱/۳۶۲
	۲۴۰	۲/۰۸۸	۳/۲۸۴	۱/۶۶۳	۱/۲۴۸	۱/۰۳۶	۱/۱۰۱	۱/۸۸۱	۱/۳۴۲	۱/۶۵۱
	۲۵۰	۲/۰۹۷	۲/۸۴۹	۱/۶۴۱	۱/۱۹۹	۱/۰۲۷	۱/۰۸۳	۱/۹۸۱	۱/۳۳۵	۲/۰۳۵
	۲۶۰	۲/۱۰۴	۲/۲۷۰	۱/۶۳۳	۱/۱۷۹	۱/۰۲۸	۱/۰۶۵	۲/۰۷۴	۱/۳۵۳	۲/۳۵۵
	۲۷۰	۲/۱۰۹	۳/۰۹۰	۱/۶۷۵	۱/۱۷۶	۱/۰۳۵	۱/۰۴۶	۲/۱۴۹	۱/۴۰۸	۲/۵۸۶
	۲۸۰	۲/۱۲۹	۳/۹۲۰	۱/۷۱۳	۱/۱۸۴	۱/۰۴۶	۱/۰۲۷	۲/۲۰۳	۱/۴۹۱	۲/۷۷۳
	۲۹۰	۲/۱۴۵	۴/۴۹۹	۱/۷۳۸	۱/۲۰۳	۱/۰۵۹	۱/۰۰۸	۲/۲۴۳	۱/۸۶۵	۲/۷۴۲
	۳۰۰	۲/۱۵۰	۵/۰۶۴	۱/۸۵۷	۱/۲۳۴	۱/۰۷۴	۰/۹۸۹	۲/۲۷۹	۳/۲۱۴	۲/۷۶۱
	۳۱۰	۲/۱۶۵	۵/۵۴۴	۱/۹۲	۱/۲۷۶	۱/۰۹۰	۰/۹۷۴	۲/۳۲۴	۴/۳۸۴	۲/۸۸۳
	۳۲۰	۲/۱۷۱	۸/۸۰۱	۲/۰۳۱	۱/۳۱۶	۱/۱۰۵	۰/۹۶۳	۲/۳۸۵	۴/۴۳۰	۳/۰۱۸
	۳۳۰	۲/۱۸۰	۹/۳۵۴	۲/۰۷۱	۱/۳۲۹	۱/۱۲۱	۰/۹۵۶	۲/۴۴۷	۴/۶۳۱	۳/۱۱۳
	۳۴۰	۲/۱۹۱	۹/۳۶۶	۲/۱۰۰	۱/۲۹۷	۱/۱۳۵	۰/۹۵۴	۲/۵۱۱	۴/۷۷۴	۳/۱۶۴
	۳۵۰	۲/۱۹۹	۹/۴۲۲	۲/۱۱۴	۱/۲۶۱	۱/۱۴۶	۰/۹۵۶	۲/۵۴۳	۵/۶۰۴	۳/۲۰۰



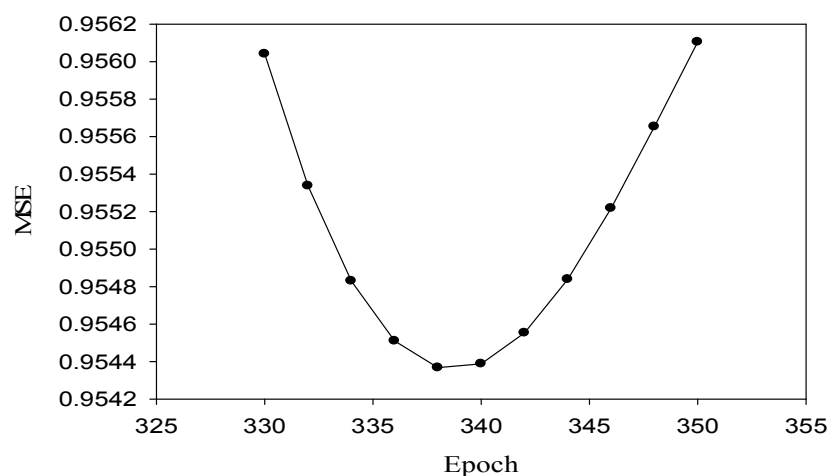
شکل (۳-۱۶) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۹)

جدول (۳-۲۰) - توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN

تابع انتقال	تابع آموزش	تعداد نرون لایه‌ی پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموئید	تنظیم بایزین	۱۸	۳۰۰	۱/۳۴۸
لگاریتم سیگموئید	لونبرگ-مارکوارت	۱۲	۳۴۰	۱/۳۰۰
تانژانت سیگموئید	تنظیم بایزین	۱۸	۱۷۰	۱/۱۰۵
تانژانت سیگموئید	لونبرگ-مارکوارت	۱۷	۳۴۰	۰/۹۵۴

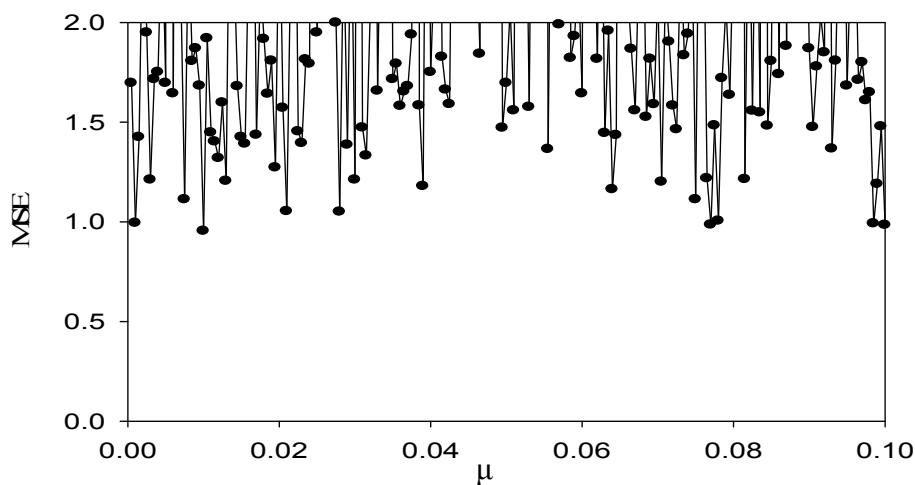
نتایج موجود در این جداول و شکل‌ها نشان می‌دهد که شبکه‌ی با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید در لایه‌ی پنهان ۱۷ و دور آموزشی ۳۴۰ که مقدار MSE کمتری نسبت به سایر شبکه‌ها دارد، به عنوان شبکه‌ی بهینه انتخاب گردید.

برای تعیین دقیق‌تر دور آموزش (Epoch)، شبکه‌ای با همان مقادیر بهینه تعداد نرون لایه ورودی، تعداد نرون لایه‌ی پنهان، تابع آموزش و تابع انتقال اما با تعداد دورهای ۳۳۰ تا ۳۵۰ با گام ۲ مورد بررسی قرار گرفت که طبق شکل (۳-۱۷) دور آموزشی ۳۳۸ با مقدار $MSE = ۰/۹۵۴۳$ به عنوان تعداد دور بهینه برای شبکه انتخاب گردید.



شکل (۳-۱۷) - نمودار میانگین مربعات خطا برای سری ارزیابی بر حسب تعداد دور آموزشی متفاوت با گام ۲

در مرحله‌ی بعد برای تعیین مقدار بهینه‌ی μ ، با استفاده از شبکه‌ی بهینه، مقدار μ از ۰/۰۰۰۵ تا ۰/۱ با گام ۰/۰۰۰۵ تغییر داده شد و بر اساس کمترین مقدار MSE برای هر گام، مقدار بهینه‌ی μ ، ۰/۰۱ به دست آمد. نمودار مقدار میانگین مربعات خطا بر حسب μ حول نقطه‌ی بهینه نیز در شکل (۳-۱۸) رسم شده است. توابع و مقادیر پارامترهای بهینه شده‌ی شبکه‌ی عصبی- روش سهم گروه برای پیش بینی دانسیته‌ی سیالات تجمعی، در جدول (۳-۲۱) ارائه شده است.



شکل (۳-۱۸)- نمودار میانگین مربعات خطا بر حسب μ حول نقطه بهینه برای سری ارزیابی

جدول (۳-۲۱)- توابع و مقادیر پارامترهای بهینه شده GCM-ANN

Train lm	تابع آموزش
tansig	تابع انتقال لایه‌ی پنهان
pure line	تابع انتقال لایه‌ی خروجی
۱۷	تعداد نرون لایه‌ی پنهان
۱۲	تعداد متغیرهای ورودی
۳۳۸	تعداد دوره‌های آموزش
۰/۰۱	پارامتر μ
۰/۹۵۴۳	مقدار MSE سری ارزیابی

۳-۲-۸- ارزیابی قدرت پیش بینی مدل های غیر خطی

در این مرحله باید مدل های به دست آمده از شبکه ی عصبی مصنوعی که با سه روش متفاوت انتخاب توصیف کننده، طراحی و بهینه سازی شده اند، مورد ارزیابی قرار گیرند. برای این منظور پارامترهای آماری، نمودار برگشتی، نمودار خطای باقی مانده و روش آزمون Y-تصادفی استفاده گردید که در ادامه راجع به هر یک توضیح داده خواهد شد.

۳-۲-۸-۱- ارزیابی مدل های غیر خطی بهینه شده با استفاده از پارامترهای

آماري

برای بررسی میزان کارایی مدل های بهینه شده به روش های مختلف شامل SR-ANN، GA-ANN و GCM-ANN برخی از پارامترهای آماری حاصل از این مدل ها که در بخش (۲-۳-۸) روابط آن ها ارائه شده است، برای سری های ارزیابی، تست با هم مقایسه شده است. این مقادیر در جدول (۳-۲۲) گزارش شده است.

جدول (۳-۲۲)- پارامترهای آماری مربوط به هر یک از شبکه های بهینه شده برای سری های ارزیابی و تست

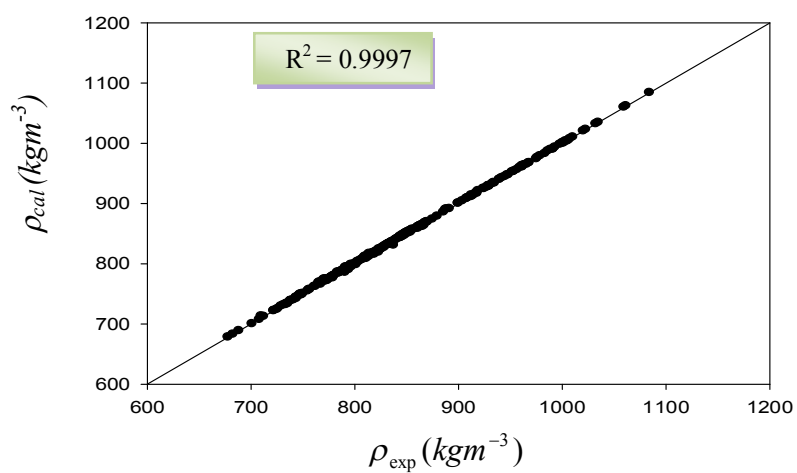
SR-ANN		GA-ANN		GCM-ANN		
ارزیابی	تست	ارزیابی	تست	ارزیابی	تست	
۱/۶۹۴۷	۰/۰۳۱۸	۷/۴۴۳۷	۶/۸۵۴۰	۰/۹۴۷۳	۱/۲۴۳۲	MSE
۰/۹۹۹۷	۰/۹۹۹۸	۰/۹۹۸۹	۰/۹۹۹۰	۰/۹۹۹۹	۰/۹۹۹۸	R^2
۰/۸۲۳۷	۷/۹۹۹۲	۲/۰۴۵۳	۱/۷۰۱۸	۰/۶۳۶۳	۱/۱۲۷۹	D_{max}
۰/۱۰۵۹	۰/۱۹۸۹	۰/۲۴۴۳	۰/۲۳۲۹	۰/۰۷۹۸	۰/۰۸۶۷	AAD
۰/۰۰۰۶	۰/۱۹۸۹	۰/۰۳۴۹	۰/۰۴۰۴	-۰/۰۰۲۸	۰/۰۰۶۹	$bias$
۴۷۲/۴۲	۳۰/۸۳۲۵	۱۹۶۵/۱۴	۱۵۵۵/۸۶	۲۵۰/۰۹	۲۸۲/۲۲	$PRESS$
۰/۷۸۶۵	۱/۵۹۰۹	۱/۹۹۴۶	۱/۹۱۵۲	۰/۶۵۴۹	۰/۷۱۱۰	MAE

۳-۲-۸-۲- با استفاده از نمودار برگشتی

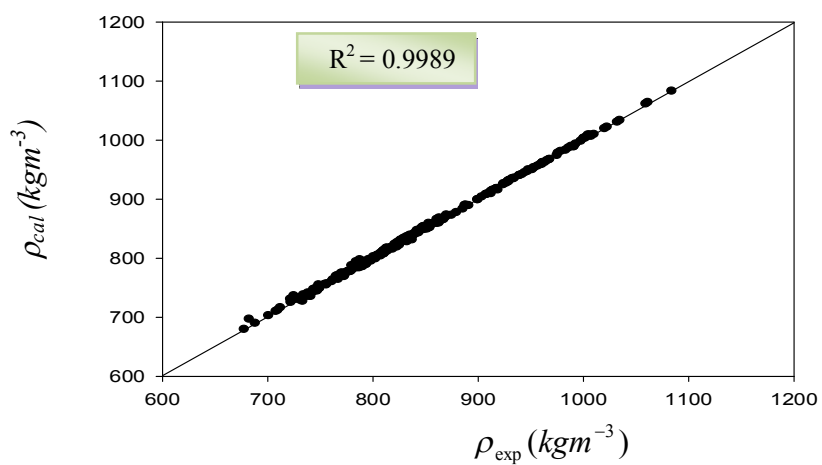
در نمودار برگشتی مقادیر پیش‌بینی شده بر حسب مقادیر تجربی رسم می‌گردد و با توجه به مقدار ضریب تعیین (R^2) به دست آمده از نمودار، پراکندگی نقاط در اطراف خط برگشت تعیین می‌شود. نتایج حاصل از نمودارهای برگشتی مربوط به سه روش انجام شده در این کار برای سری‌های ارزیابی و تست در شکل‌های (۳-۱۹) و (۳-۲۰) آورده شده است.

۳-۲-۸-۳- با استفاده از نمودار خطای باقی‌مانده

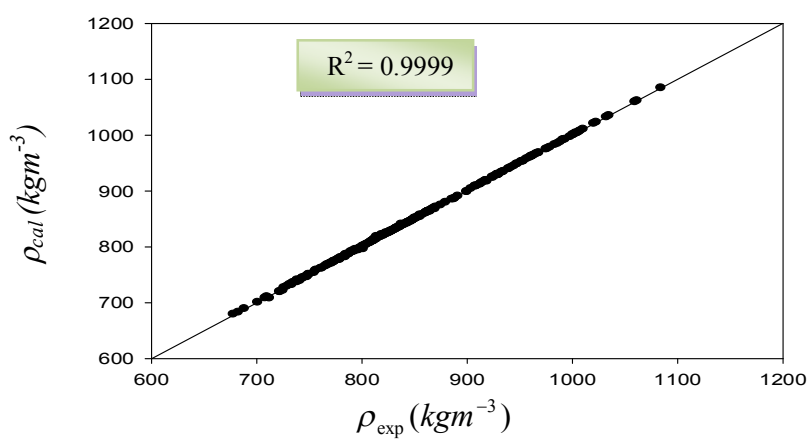
خطای باقی‌مانده به تفاوت مقادیر پیش‌بینی شده و مقادیر تجربی گفته می‌شود. اگر پراکندگی مقادیر در حول محور صفر یکنواخت باشد، این امر نشان‌دهنده‌ی تصادفی بودن خطاها است. شکل‌های (۳-۲۱) تا (۳-۲۲) نمودار خطای باقی‌مانده‌های مربوط به سه روش انجام شده‌ی SR-ANN، GA-ANN، GCM-ANN را برای سری‌های ارزیابی و تست نشان می‌دهند. پراکندگی متقارن داده‌ها در این شکل‌ها در حول خطای صفر نشان از عدم وجود خطای سیستماتیک در داده‌ها است.



(الف)

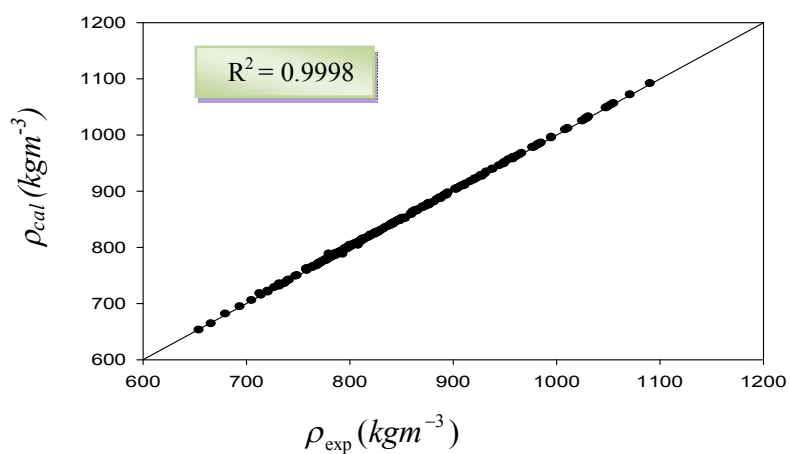


(ب)

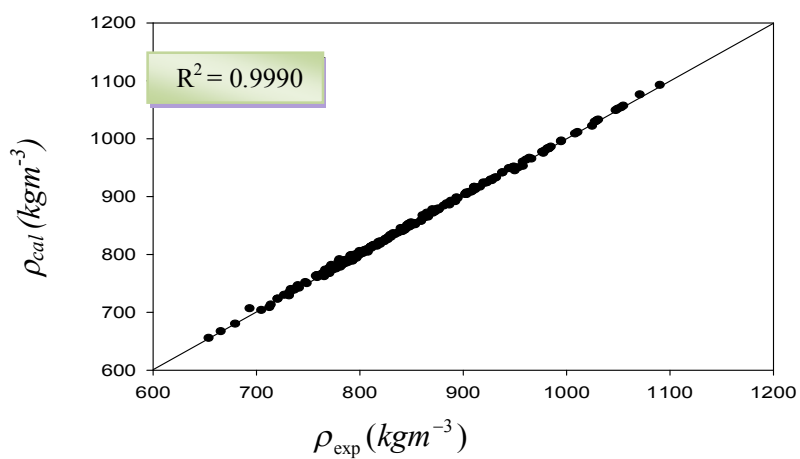


(ج)

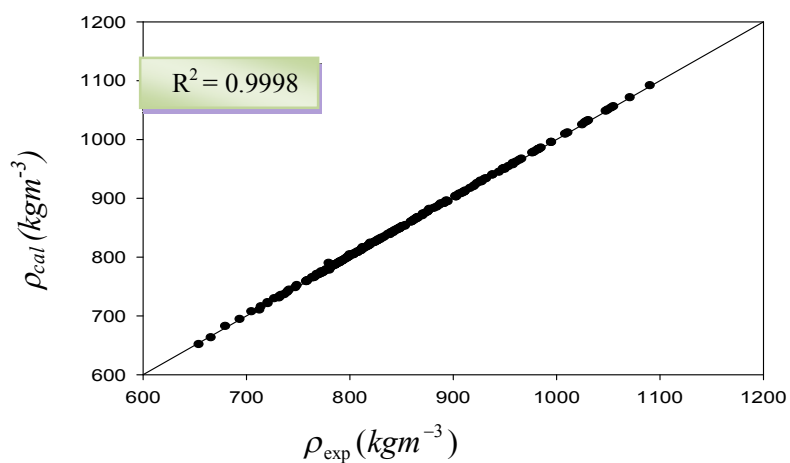
شکل (۳-۱۹) - نمودار مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری ارزیابی (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN



(الف)

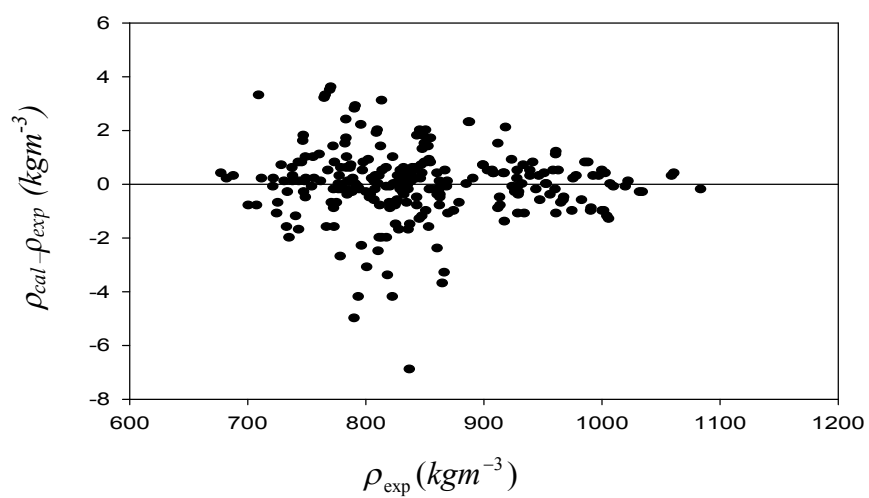


(ب)

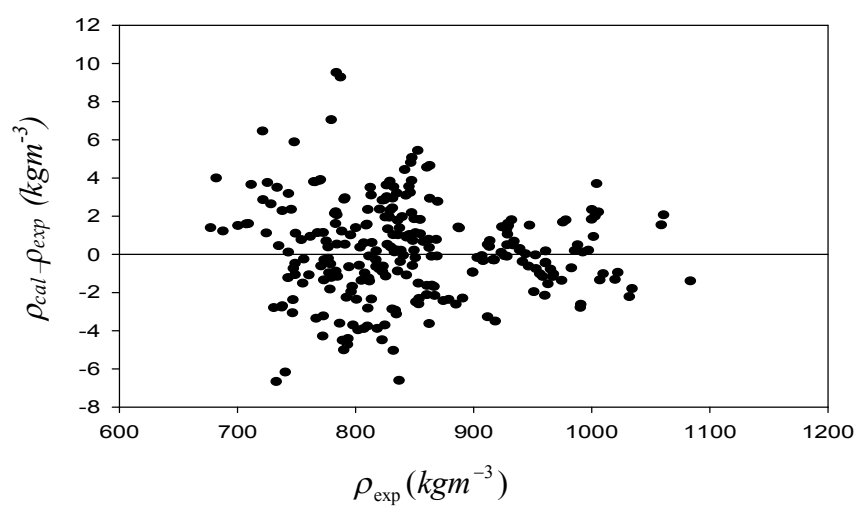


(ج)

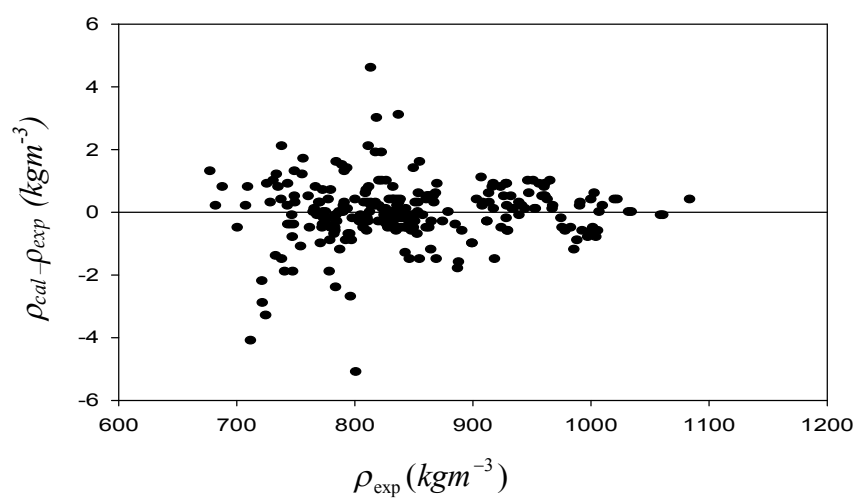
شکل (۳-۲۰) - نمودار مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN



(الف)

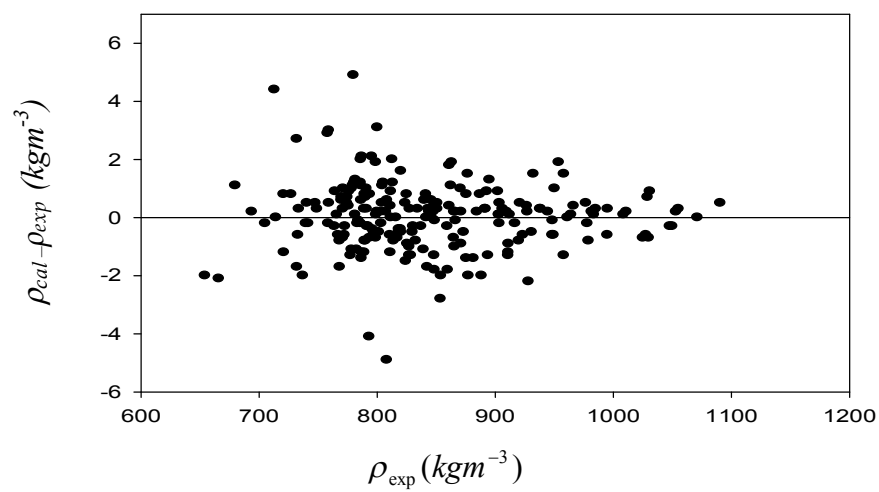


(ب)

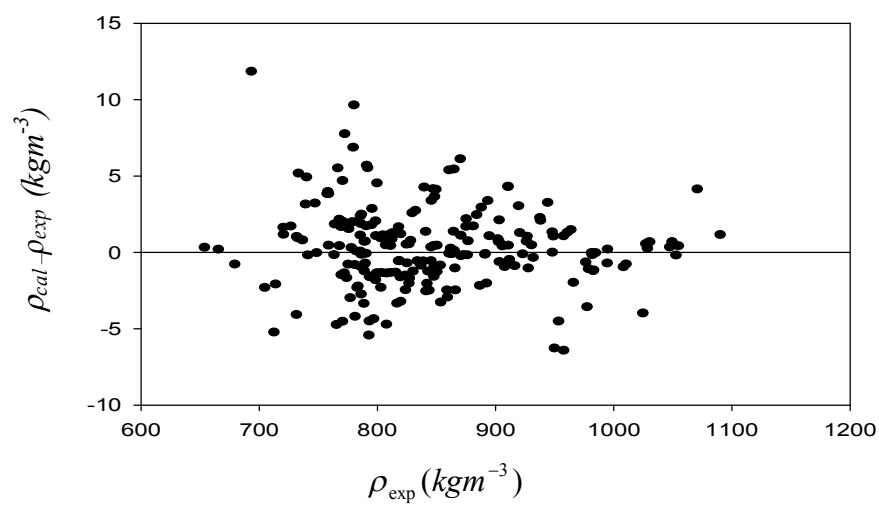


(ج)

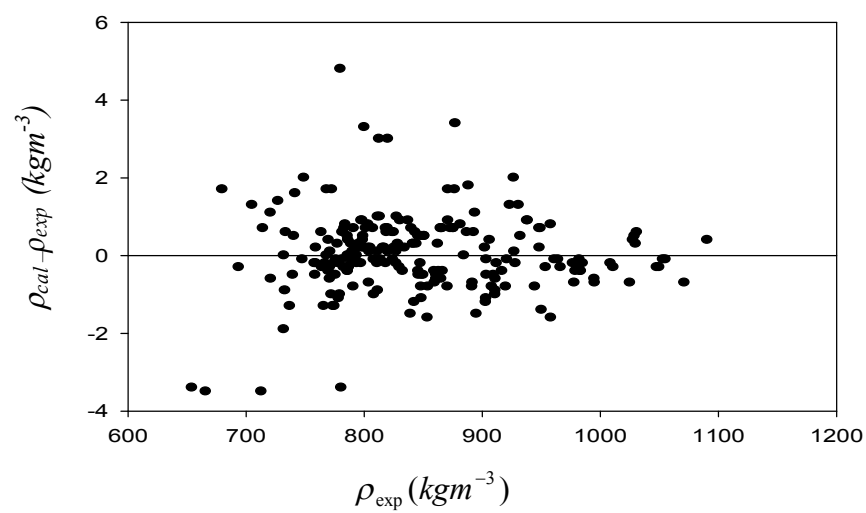
شکل (۳-۲۱)- نمودار خطای باقیمانده برحسب مقدار تجربی دانسیته برای سری ارزیابی (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN



(الف)



(ب)



(ج)

شکل (۳-۲۲) - نمودار خطای باقیمانده برحسب مقدار تجربی دانسیته برای سری تست (الف) مدل SR-ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN

۳-۲-۸-۴- ارزیابی با استفاده از آزمون y تصادفی

در این روش مقادیر متغیر وابسته؛ یعنی دانسیته، در محدوده‌ی داده‌های تجربی به صورت تصادفی تغییر داده شدند و شبکه‌های عصبی بهینه شده با استفاده از متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر وابسته ارزیابی و بررسی شد. سپس همبستگی متغیرهای مستقل با متغیرهای وابسته توسط شاخص R^2 مورد بررسی قرار گرفت. در جدول (۳-۲۳) مقادیر ضریب رگرسیون برای سری ارزیابی و تست آورده شده است. مقادیر پایین R^2 نشان‌دهنده‌ی عدم وجود همبستگی تصادفی در مدل اصلی است.

جدول (۳-۲۳)- مقادیر R^2 به دست آمده برای سری ارزیابی و سری تست از آزمون y - تصادفی در مدل SR-

ANN، GA-ANN و GCM-ANN

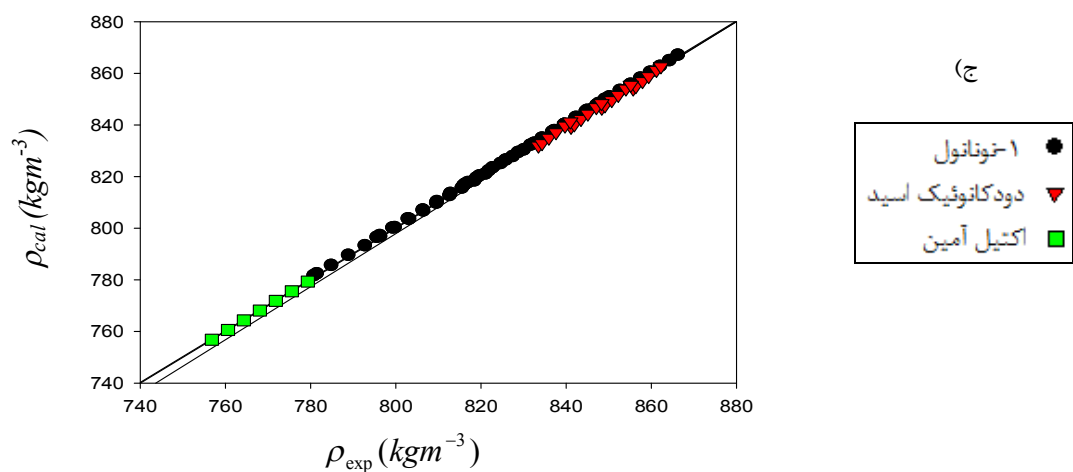
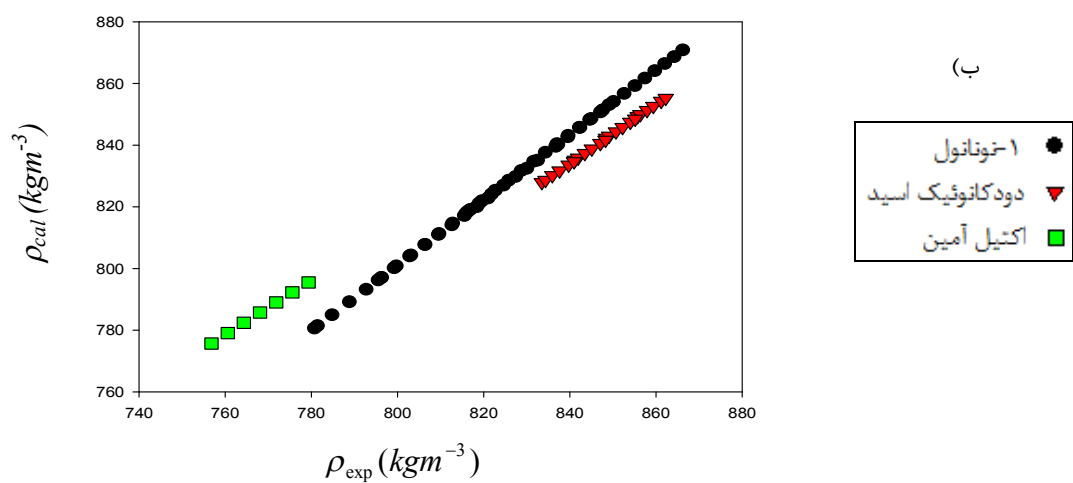
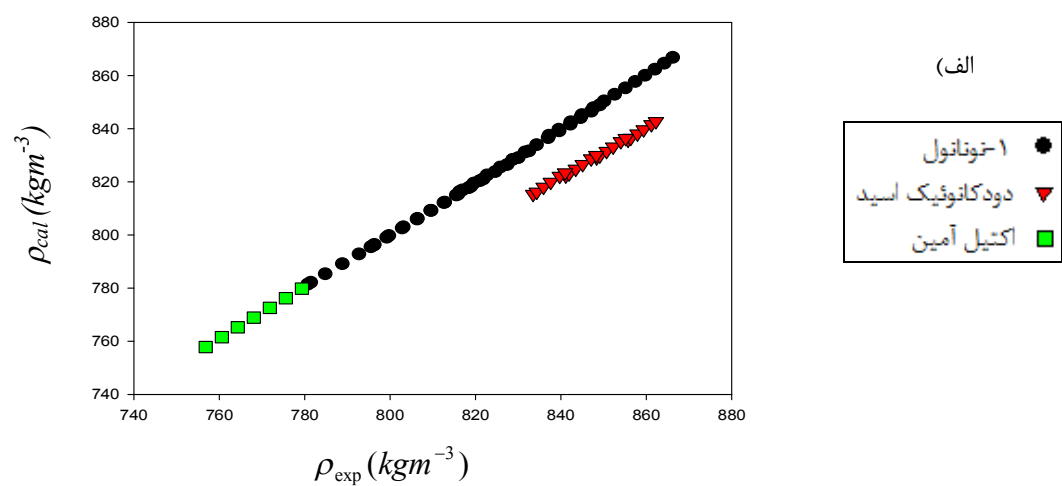
شماره تکرار آزمون	SR-ANN		GA-ANN		GCM-ANN	
	R^2 تست	R^2 ارزیابی	R^2 تست	R^2 ارزیابی	R^2 تست	R^2 ارزیابی
۱	۰/۲۸۴	۰/۱۶۰	۰/۲۴۸	۰/۰۰۱	۰/۰۱۳	۰/۰۳۳
۲	۰/۰۲۵	۰/۰۴۷	۰/۱۲۵	۰/۰۱۷	۰/۳۰۲	۰/۰۱۴
۳	۰/۰۰۰	۰/۰۰۰	۰/۱۶۴	۰/۰۹۱	۰/۰۷۱	۰/۰۸۲
۴	۰/۲۷۵	۰/۰۹۵	۰/۰۰۴	۰/۰۰۷	۰/۴۰۲	۰/۰۶۰
۵	۰/۰۰۳	۰/۰۰۴	۰/۰۵۳	۰/۰۰۰	۰/۰۷۱	۰/۰۸۲
۶	۰/۰۷۹	۰/۲۵۱	۰/۱۳۴	۰/۰۲۳	۰/۳۴۷	۰/۳۱۱
۷	۰/۰۵۹	۰/۰۱۷	۰/۰۶۵	۰/۰۶۸	۰/۴۳۹	۰/۰۳۴
۸	۰/۰۰۶	۰/۱۷۳	۰/۰۴۵	۰/۰۰۴	۰/۰۰۵	۰/۲۶۱
۹	۰/۱۴۳	۰/۱۵۷	۰/۰۰۱	۰/۰۰۶	۰/۰۱۲	۰/۰۰۶
۱۰	۰/۰۴۸	۰/۲۲۴	۰/۱۷۹	۰/۱۰۲	۰/۰۰۰	۰/۰۵۲

۳-۲-۸-۵- ارزیابی با استفاده از اعتبارسنجی بیرونی

نمودارهای مقادیر پیش‌بینی شده بر حسب مقادیر تجربی برای سری تست بیرونی که به صورت مولکولی جدا شده‌اند، در شکل‌های (۳-۲۳) رسم شده است. همچنین پارامترهای آماری مربوط به هر یک از شبکه‌های بهینه شده برای این دو سری در جدول (۳-۲۴) گردآوری شده است.

جدول (۳-۲۴)- مقادیر پارامترهای آماری برای سری تست بیرونی

SR-ANN	GA-ANN	GCM-ANN	
۹۹/۲۳۹۰	۳۹/۶۵۴۸	۱/۱۲۲۵	<i>MSE</i>
۰/۹۱۴۵	۰/۹۵۸۷	۰/۹۹۹۱	<i>R²</i>
۲/۳۳۵۳	۲/۴۹۰۲	۰/۴۶۳۱	<i>D_{max}</i>
۰/۶۷۲۸	۰/۵۵۰۲	۰/۰۸۸۷	<i>AAD</i>
-۰/۶۵۳۳	۰/۱۳۸۱	۰/۰۳۱۶	<i>bias</i>
۸۵۳۴/۵۵	۳۴۴۹/۹۶	۹۷/۶۶۲۶	<i>PRESS</i>
۵/۶۸۸۶	۴/۵۰۰۸	۰/۷۲۰۵	<i>MAE</i>



شکل (۳-۲۳)- نمودار مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست بیرونی (الف) مدل SR-
 ANN، (ب) مدل GA-ANN، (ج) مدل GCM-ANN

۳-۲-۹- بررسی ارتباط توصیف‌کننده‌های وارد شده در مدل با دانسیته‌ی

ترکیبات

در این قسمت یک بررسی اجمالی روی اثرات مختلف توصیف‌کننده‌های تئوری انتخاب شده با روش رگرسیون مرحله‌ای و الگوریتم ژنتیک بر دانسیته‌ی سیالات تجمعی مورد مطالعه صورت گرفته است. این توصیف‌کننده‌ها شامل $X0Av$ ، $ATS2e$ ، RBF ، $RDF035e$ ، $GATS4e$ و $R3m$ می‌باشند.

۳-۲-۹-۱- توصیف‌کننده‌ی $X0Av$

$X0Av$ توصیف‌کننده‌ای از گروه اندیس‌های ارتباط مولکولی است. این اندیس‌ها دسته‌ای از توصیف‌کننده‌های توپولوژیکی هستند که اطلاعاتی در مورد ساختمان، اندازه و میزان شاخه‌ای شدن مولکول، نحوه‌ی اتصال اتم‌ها و نوع اتم‌های موجود در یک مولکول را در اختیار قرار می‌دهند. جهت محاسبه‌ی این توصیف‌کننده‌ها ابتدا باید ساختار مولکول را با استفاده از تئوری گراف شیمیایی رسم نمود.

این اندیس‌ها اولین بار به‌وسیله راندیک^۱ در سال ۱۹۷۵ ارائه شدند و سپس در سال ۱۹۷۶ توسط هال^۲ اصلاح گردیدند. این توصیف‌کننده‌ها با رابطه‌ی (۳-۲) قابل محاسبه هستند [۶۱ و ۶۲].

$${}^m\chi^v = \sum_{i=1}^{N_s} \prod_{k=1}^{m+1} \left(\frac{1}{\delta_k^v} \right)^{\frac{1}{2}} \quad (2-3)$$

که در این رابطه δ_k^v ، درجه رأس ظرفیت^۳ اتم k ام در گراف مولکولی است و به شکل رابطه‌ی زیر تعریف می‌شود:

$$\delta_k^v = \frac{(Z_K^v - h_k)}{(Z - Z_k^v - 1)} \quad (3-3)$$

1- Randic

2- Hall

3- Valence vertex degree

که در آن Z عدد اتمی اتم k ، Z_k^v تعداد الکترون‌های لایه‌ی ظرفیت و h_k تعداد اتم‌های هیدروژن متصل به اتم k است. X0Av وارد شده در مدل دارای اثر متوسط منفی و برابر ۱۷/۹۲۲- بوده و بنابراین با افزایش آن دانسیته کاهش می‌یابد.

۳-۲-۹-۲- توصیف‌کننده‌های گروه 2D-Autocorrelation

توصیف‌کننده‌های دو بعدی خود همبستگی با اعمال گراف مولکولی و از طریق محاسبه‌ی مجموع اوزان اتم‌های انتهایی کل مسیرها با طول مسیر مورد نظر (lag)، به دست می‌آیند. نوعی از این توصیف‌کننده‌ها گروه خودهمبستگی گری^۱ در سال ۱۹۵۴ توسط گری^۲ پیشنهاد شدند که ضریب گری نام دارند و طبق معادله‌ی (۳-۴) محاسبه می‌شوند.

$$C(d) = \frac{\frac{1}{2A} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (\omega_i - \omega_j)^2}{\frac{1}{(A-1)} \cdot \sum_{i=1}^A (\omega_i - \bar{\omega})^2} \quad (۳-۴)$$

که ω یک ویژگی اتم، $\bar{\omega}$ میانگین مقدار آن خاصیت روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است و δ_{ij} ، دلتای کرونکر است (اگر $d_i = d_j$ باشد $\delta_{ij} = 1$ است و در غیر این صورت صفر است). Δ هم مجموع دلتاهای کرونکر است. مقدار این توصیف‌کننده‌ها از جنس فاصله است بنابراین از یک تا بی‌نهایت متغیر است. معنی عبارت ۴e در توصیف‌کننده‌ی GATS4e بیانگر فاصله‌ی توپولوژیکی ۴ و وزن دار شدن این توصیف‌کننده توسط الکترونگاتیوی اتمی ساندerson^۳ است. با توجه به این که مقدار اثر این توصیف‌کننده (۴/۶۶۲-) می‌باشد، با دانسیته رابطه‌ی عکس دارد.

نوع دیگر از این توصیف‌کننده‌ها، توصیف‌کننده‌های Moreau-Boroto autocorrelation هستند که توسط رابطه‌ی (۳-۵) محاسبه می‌شوند.

1- Ceary Autocorrelation

2- Ceary

3 -Atomic Sanderson electronegativities

$$ATS_d = \sum_{i=1}^N \sum_{j=1}^N \delta_{ij} . (\omega_i . \omega_j) = W^T . {}^m B . W \quad (5-3)$$

در این جا نیز پارامترهای ω ، δ و A مانند رابطه‌ی قبل تعریف می‌شوند. ${}^m B$ یک ماتریس دودویی خلوت^۱ از مرتبه‌ی m ام است و W یک بردار A بعدی از ویژگی‌های اتمی موردنظر می‌باشد. منظور از ویژگی اتمی موردنظر، جرم اتمی، قطبش‌پذیری، بار و الکترونگاتیویته می‌باشد. توصیف‌کننده‌ی ATS2e در این گروه از توصیف‌کننده‌ها جای دارد. این توصیف‌کننده نیز با الکترونگاتیوی ساندerson وزن‌دار شده و فاصله‌ی توپولوژیکی آن ۲ می‌باشد [۶۳]. با توجه به رابطه‌ی (۵-۳) ویژگی مورد نظر؛ یعنی الکترونگاتیوی، با ATS2e رابطه‌ی مستقیم دارد، بنابراین با مقایسه‌ی یک الکل و آمین هم کربن که در آن‌ها الکترونگاتیوی اتم اکسیژن نسبت به نیتروژن بیشتر است باید مقدار این توصیف‌کننده در الکل بیشتر باشد. به عنوان مثال ۱-بوتانول که مقدار توصیف‌کننده‌ی ATS2e آن ۰/۹۷۶ و برای n-بوتیل آمین ۰/۹۵۴ می‌باشد. که این ادعای ما را اثبات می‌کند. و از آن جایی که اثر متوسط این توصیف‌کننده در مدل SR-ANN به دست آمده مقداری مثبت است. بنابراین با افزایش این توصیف‌کننده دانسیته نیز افزایش می‌یابد.

۳-۲-۹-۳- توصیف‌کننده‌های Constitutional

این گروه از توصیف‌کننده‌ها ساده‌ترین و معمول‌ترین توصیف‌کننده‌های استفاده شده می‌باشند، که خصوصیات مولکول را بدون داشتن اطلاعات زیادی در مورد آن منعکس می‌کنند. تعداد اتم‌ها، تعداد پیوندها، تعداد مطلق و نسبی یک نوع اتم خاص در مولکول، تعداد پیوندهای ساده، دوگانه، سه‌گانه، و آروماتیک، تعداد حلقه‌ها، تعداد حلقه‌ها بر اساس تعداد اتم‌های آن‌ها و وزن مولکولی متداول‌ترین توصیف‌کننده‌های ساختاری می‌باشد.

1- Binary sparse matrix

یکی از این توصیف‌کننده‌ها، توصیف‌کننده‌های RBF است و بیانگر کسری از پیوندهای قابل چرخش می‌باشد. همان‌طور که از جدول (۷-۳) مشاهده می‌شود. با افزایش جرم اتمی، تعداد پیوندها نیز افزایش می‌یابد و مقدار این توصیف‌کننده‌ها نیز افزایش می‌یابد. از طرف دیگر دانسیته با جرم رابطه‌ی مستقیم دارد. لذا با افزایش توصیف‌کننده‌ی RBF دانسیته نیز افزایش می‌یابد. علامت این توصیف‌کننده در هر دو روش SR و GA مثبت و بیانگر این است که این توصیف‌کننده رابطه‌ی مستقیم با دانسیته دارد [۶۴].

۳-۲-۹-۴- توصیف‌کننده‌های تابع توزیع شعاعی (RDF)

این دسته از توصیف‌کننده‌ها، بر اساس فاصله‌ی توزیع در نمایش هندسی یک مولکول، شکل گرفته‌اند. تابع توزیع شعاعی مجموعه‌ای از اتم‌های A، به‌صورت احتمال توزیع اتم در اطراف کره‌ای به شعاع R تعریف می‌شوند. انواع این توصیف‌کننده به‌صورت RDF_{sw} مشخص می‌شوند که مقادیر s ، $10 \leq s \leq 155$ با گام ۵ و $w \in (u, m, v, e, p)$ می‌باشد و به‌صورت زیر نمایش داده می‌شود:

$$RDF_{sw} = f \cdot \sum_{i=1}^{A-1} \sum_{j=i+1}^A w_i \cdot w_j \cdot e^{-\beta(R-r_{ij})^2} \quad (۶-۳)$$

در این رابطه، f فاکتور اندازه‌گیری، w خواص اتمی اتم‌های i و j ، r_{ij} فاصله‌ی بین اتمی اتم‌های i و j و پارامتر تنظیم‌پذیر β به عنوان فاکتور دما که به جنبش اتم‌ها بستگی دارد، تعریف می‌شود.

RDF035e یکی از این نوع توصیف‌کننده‌هاست که توسط روش GA به‌دست آمده است و با الکترونگاتیوی اتمی ساندerson^۱ وزن دار شده است.

در این رابطه توصیف‌کننده‌ی RDF با الکترونگاتیوی رابطه‌ی مستقیم و با عامل دما β رابطه‌ی عکس دارد. بنابراین انتظار می‌رود مقدار این توصیف‌کننده در الکل‌ها نسبت به آمین‌های با تعداد کربن یکسان بیشتر باشد که مقادیر در جدول (۷-۳) این موضوع را نشان می‌دهد. مقدار مثبت اثر متوسط این

1- Sanderson

توصیف‌کننده (۱۷/۹۵۲) این را نشان می‌دهد که توصیف‌کننده‌ی RDF با دانسیته رابطه‌ی مستقیم دارد [۶۵].

۳-۲-۹-۵- توصیف‌کننده‌های GETAWAY^۱

این توصیف‌کننده‌ها بیان‌کننده‌ی خاصیت‌های توپولوژیکی و هندسی مولکول می‌باشند که توسط ماتریس تأثیر مولکولی (MIM)^۲ محاسبه می‌شوند. این ماتریس به آسانی از مختصات فضایی اتم‌های مولکول در یک صورتبندی مشخص به دست می‌آید که با رابطه‌ی زیر بیان می‌گردد:

$$H = M(M^T M) \cdot M^T \quad (۷-۳)$$

در رابطه‌ی فوق M ماتریس مختصات اتمی، T به معنای ماتریس ترانپوز و H ماتریس تأثیر مولکولی است. عناصر قطری (h_{ij}) ماتریس H لوریج^۳ نام دارند که هر یک بیانگر اثر یک اتم در ایجاد شکل کلی مولکول است. به عنوان مثال اتم‌های دورتر نسبت به اتم‌های نزدیک‌تر به مرکز اتم، دارای (h_{ij}) بزرگتری هستند. یک سری از این نوع توصیف‌کننده‌ها Rkw می‌باشد که ترکیبی از اطلاعات ماتریس تأثیر-فاصله است که به صورت زیر تعریف می‌شود:

$$Rkw = \sum_{i=1}^{n-1} \sum_{j>1}^n R_{ij} \cdot w_i w_j \cdot \delta(k, d_{ij}) \quad (۸-۳)$$

$$R_{ij} = \left[\frac{\sqrt{h_{ii} h_{jj}}}{r_{ij}} \right]_{ij} \quad i \neq j \quad (۹-۳)$$

1- Geometrical, Topology and atom Weighted assembly
2- Molecular Influence Matrix
3- Leverage

h_{jj} و h_{ii} لوریج (قدرت نفوذ) دو اتم i و j و r_{ij} فاصله‌ی آن دو اتم است و مقدار $\delta(k; d_{ij}) = 1$ است، اگر $k = d_{ij}$ باشد و در غیر این صورت برابر صفر می‌باشد. توصیف‌کننده‌ی R3m نیز جزء این دسته می‌باشد که در آن دو اتم در فاصله‌ی توپولوژیکی ۳ قرار دارند و با وزن اتمی وزن‌دار شده، محاسبه می‌شود.

با توجه به روابط (۳-۸) و (۳-۹) این توصیف‌کننده با فاصله‌ی دو اتم رابطه‌ی عکس و با جرم اتمی رابطه‌ی مستقیم دارد. همان‌طور که از مقادیر توصیف‌کننده‌ها در جدول (۳-۷) مشاهده می‌شود، با افزایش جرم اتمی این توصیف‌کننده افزایش می‌یابد. به عنوان مثال با مقایسه‌ی مقدار توصیف‌کننده‌ی R3m اتانول و اتان دیول مشاهده می‌کنیم که مقدار آن برای اتان دیول بیشتر است. اما در مورد سیالاتی که جرم اتمی یکسان دارند و فقط جای گروه عاملی تغییر کرده، مقدار این توصیف‌کننده به فاصله‌ی اتم‌ها بستگی دارد. از طرف دیگر با افزایش فاصله‌ی بین دو اتم، حجم نیز افزایش یافته و چون دانسیته با حجم رابطه‌ی عکس دارد، بنابراین دانسیته نیز با فاصله‌ی دو اتم رابطه‌ی عکس دارد. همچنین دانسیته با جرم رابطه‌ی مستقیم دارد. از آنجایی که وابستگی دانسیته و توصیف‌کننده‌ی R3m به فاصله‌ی بین دو اتم و جرم مشابه یکدیگرند بنابراین می‌توان نتیجه گرفت که این دو با یکدیگر رابطه‌ی مستقیم دارند. این توصیف‌کننده در مدل SR-ANN دارای علامت مثبت با مقدار $13/890$ می‌باشد که خود شاهده‌ی بر این مدعا است [۶۱].

۳-۲-۱۰- بررسی میزان مشارکت توصیف‌کننده‌های منتخب شبکه عصبی

میزان مشارکت توصیف‌کننده‌های انتخاب شده به صورت زیر تعیین شد:

۱. ابتدا وزن مربوط به توصیف‌کننده‌ی مورد نظر از شبکه بهینه حذف شد.
۲. مقدار متغیر وابسته با استفاده از توصیف‌کننده‌های باقی‌مانده برای هر یک از ترکیبات سری ارزیابی پیش‌بینی گردید.

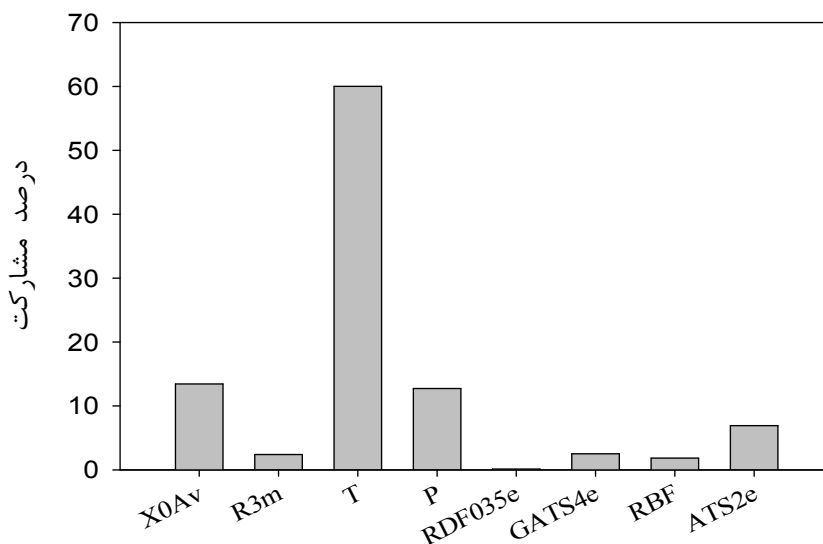
۳. میانگین خطای مطلق (MAE) حاصل از ترکیبات سری ارزیابی محاسبه شد.

۴. مراحل ۱ تا ۳ برای دیگر توصیف‌کننده‌ها نیز تکرار شد.

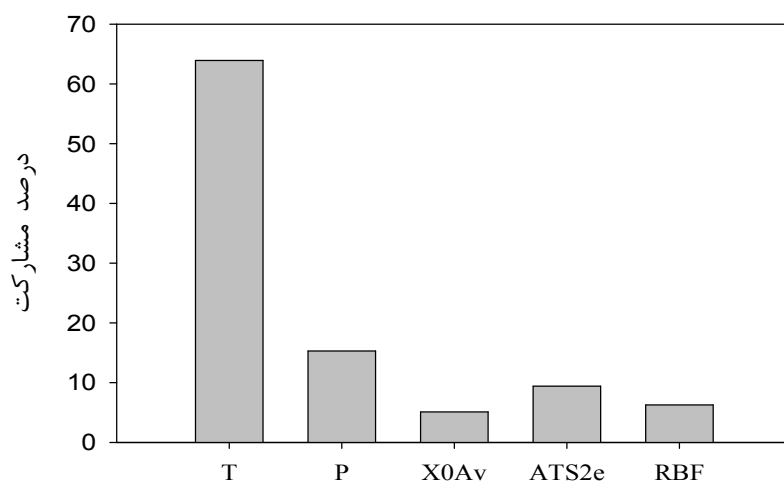
۵. در نهایت درصد مشارکت هر توصیف‌کننده توسط رابطه‌ی (۳-۱۰) محاسبه گردید:

$$C_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad (3-10)$$

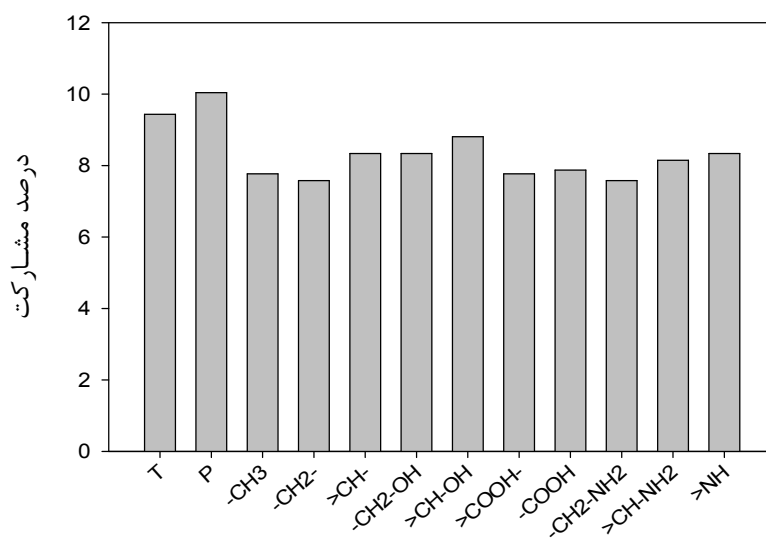
در این رابطه C_i درصد مشارکت توصیف‌کننده‌ی حذف شده i ام، N تعداد توصیف‌کننده‌ها در مدل و Δm_i میانگین خطای مطلق حاصل از سری ارزیابی در غیاب توصیف‌کننده‌ی i ام را نشان می‌دهد. بنابراین درصد مشارکت توصیف‌کننده‌های منتخب در شبکه‌های بهینه شده، با توصیف‌کننده‌های منتخب به روش SR، GA، و GCM مصابق شکل‌های (۳-۲۴) تا (۳-۲۶) می‌باشد.



شکل (۳-۲۴) - درصد مشارکت توصیف‌کننده‌ها در مدل SR-ANN



شکل (۳-۲۵) - درصد مشارکت توصیف‌کننده‌ها در مدل GA-ANN



شکل (۳-۲۶) - درصد مشارکت توصیف‌کننده‌ها در مدل GCM-ANN

بر اساس شکل‌های به‌دست آمده، در دو مدل SR-ANN و GA-ANN بیشترین درصد مشارکت

مربوط به دما سپس فشار است. این روند در داده‌های تجربی دانسیته نیز مشاهده می‌گردد، یعنی میزان

تغییرات دانسیته با تغییر دما در فشار ثابت بیشتر از میزان تغییرات دانسیته با تغییرات فشار در یک دمای ثابت می‌باشد.

روش سهم گروه در مقایسه با دو روش قبل رفتار متفاوتی دارد. به طوریکه اولاً میزان مشارکت دما و فشار در این روش تقریباً یکسان بوده به طوری که درصد مشارکت فشار کمی بیشتر از دما می‌باشد و ثانیاً درصد مشارکت این دو توصیف‌کننده‌ی تجربی نسبت به دو روش ذکر شده بسیار کمتر شده است. در روش GCM-ANN با وجود این که تعداد گروه‌های ساختاری تشکیل‌دهنده ترکیبات به عنوان توصیف‌کننده‌های ورودی به شبکه‌ی عصبی مصنوعی در نظر گرفته شدند و بنابراین در ماتریس ورودی تعداد زیادی صفر وجود دارد، اما درصد مشارکت توصیف‌کننده‌های ساختاری مقادیر قابل توجهی دارند. همان‌طور که از شکل (۳-۲۶) مشاهده می‌شود تمام گروه‌ها تقریباً درصد مشارکت یکسانی دارند.

۳-۳- نتیجه‌گیری

در این پایان‌نامه از روش مدل‌سازی شبکه عصبی مصنوعی به عنوان روشی قدرتمند و کارآمد برای پیش‌بینی دانسیته‌ی سیالات تجمعی استفاده شده است. از آنجایی که اندازه‌گیری دانسیته به صورت تجربی با صرف هزینه، زمان و پیچیدگی‌های زیادی همراه است، بنابراین استفاده از این روش‌ها از اهمیت بالایی برخوردار است. برای انتخاب توصیف‌کننده‌های معتبر و مناسب از روش‌های GA، SR و GCM استفاده گردید. مقایسه‌ی نتایج سه روش نشان می‌دهد که:

✓ با مقایسه‌ی پارامترهای آماری روش مدل‌سازی GA-ANN و SR-ANN، روش مدل‌سازی SR-ANN نتایج بهتری را نسبت به GA-ANN برای پیش‌بینی دانسیته ارائه می‌دهد. اما از آنجایی که تعداد توصیف‌کننده در روش GA کمتر از روش SR می‌باشد بنابراین از نظر پیچیدگی شبکه از SR بهتر است.

✓ روش سهم گروه در مقایسه با دو روش دیگر نتایج بهتری را ارائه می‌دهد. همان‌طور که در جدول (۳-۲۲) مشاهده می‌گردد، مدل حاصل از این روش دارای میانگین مربعات خطای کمتر و ضریب رگرسیون بالاتری است. با توجه به سادگی روش سهم گروه این نتایج از اهمیت بالایی برخوردارند. در واقع این روش پیچیدگی دو روش دیگر را نداشته و نتایج بهتری را نیز ارائه می‌نماید.

✓ بررسی نمودارهای باقی‌مانده در شکل‌های (۳-۲۱) و (۳-۲۲) نشان می‌دهد که گستره‌ی باقی‌مانده‌های سری‌های ارزیابی و تست در مدل حاصل از روش سهم گروه کمتر است، به این معنا که میزان خطا در این روش نسبت به دو روش دیگر کمتر است. بنابراین ارزیابی مدل توسط نمودارهای باقی‌مانده نیز تأیید می‌کند که روش سهم گروه نسبت به الگوریتم ژنتیک و رگرسیون مرحله‌ای عملکرد بهتری دارند.

۳-۴- آینده‌نگری:

- ✓ نتایج حاصل از پیش‌بینی توسط مدل‌سازی شبکه‌ی عصبی مصنوعی سیالات تجمعی نشان داد که می‌توان از این روش برای پیش‌بینی دانسیته‌ی سایر سیالات نیز استفاده نمود.
- ✓ برای انتخاب توصیف‌کننده‌ی مناسب و معتبر می‌توان از سایر روش‌ها مثل روش‌های تخمین تابع ژنتیک (GFA)^۱، الگوریتم جستجوی ممنوع^۲، الگوریتم اجتماع مورچگان^۳ و الگوریتم بهینه‌سازی جمعیت ذره‌ها^۴ استفاده و نتایج را با این کار مقایسه کرد.
- ✓ برای مدل‌سازی می‌توان از روش‌های تابع پایه‌ی شعاعی (RBF)^۵، ماشین برداری پشتیبان (SVM)^۶، حداقل مربعات ماشین‌های بردار پشتیبان (LS-SVM)^۶ و سیستم استنتاج فازی-عصبی سازگار (ANFIS)^۷ استفاده کرد.

1- Genetic Function Approximation

2- Tabu Search

3- Ant Colony Optimization

4- Honey-bee Matikg Optimization Algorithm

5- Radial basis function

6- Least Squares Support Vector Machines

7-Adaptive neuron-fuzzy inference system

پیوست

جدول (۱-پ) - مقایسه‌ی مقادیر تجربی و مقادیر پیش‌بینی شده‌ی دانسیته برای ترکیبات سری

تست بیرونی بر حسب kg.m^{-3}

1-nonanol	Experimental	SR-ANN	GA-ANN	GCM-ANN
1	837.40	837.26	839.97	837.75
2	839.70	839.53	842.48	840.07
3	842.50	842.30	845.50	842.88
4	845.10	844.99	848.41	845.61
5	847.80	847.62	851.21	848.25
6	850.30	850.17	853.93	850.81
7	852.80	852.67	856.55	853.31
8	855.30	855.10	859.08	855.73
9	857.60	857.49	861.54	858.09
10	859.90	859.82	863.92	860.40
11	862.20	862.11	866.23	862.64
12	864.40	864.36	868.48	864.83
13	866.40	866.57	870.67	866.96
14	816.50	816.11	818.30	817.01
15	817.10	816.69	818.94	817.61
16	819.80	819.23	821.76	820.20
17	822.90	822.31	825.15	823.34
18	826.00	825.29	828.41	826.36
19	828.90	828.18	831.55	829.29
20	831.80	830.99	834.56	832.12
21	834.50	833.71	837.47	834.87
22	837.20	836.36	840.27	837.54
23	839.80	838.94	842.97	840.14
24	842.40	841.46	845.58	842.67
25	844.80	843.91	848.10	845.15
26	847.30	846.32	850.55	847.56
27	849.30	848.67	852.92	849.93
28	795.70	795.33	796.07	796.39
29	796.40	795.99	796.81	797.07
30	799.40	798.87	800.02	800.03
31	803.00	802.36	803.88	803.58
32	806.50	805.72	807.57	806.98
33	809.80	808.95	811.11	810.25
34	813.0	812.08	814.51	813.40
35	816.10	815.09	817.77	816.44
36	819.10	818.01	820.90	819.38
37	822.00	820.84	823.92	822.23
38	824.80	823.58	826.82	824.99
39	827.60	826.25	829.61	827.69
40	830.20	828.84	832.31	830.319
41	832.70	831.37	834.91	832.88
42	780.90	781.22	780.45	781.53
43	781.60	781.95	781.26	782.28
44	785.00	785.13	784.78	785.54
45	789.00	788.96	789.01	789.44

ادامہ جدول (۱-پ)

46	792.90	792.63	793.05	793.16
47	796.50	796.16	796.91	796.72
48	799.90	799.55	800.61	800.14
49	803.30	802.81	804.16	803.43
50	806.60	805.96	807.55	806.60
51	809.70	809.00	810.81	809.65
52	812.80	811.94	813.95	812.61
53	815.70	814.78	816.96	815.48
54	818.60	817.54	819.86	818.27
55	821.20	820.21	822.66	820.98
Dodecanoic Acid	Experimental	SR-ANN	GA-ANN	GA-ANN
1	855.74	835.76	849.35	854.05
2	856.36	836.49	849.96	854.95
3	857.93	838.10	851.32	856.93
4	859.4	839.67	852.66	858.88
5	861.22	841.61	854.30	861.26
6	862.27	842.76	855.27	862.66
7	848.38	829.07	842.34	846.62
8	849.13	829.84	842.98	847.54
9	850.7	831.53	844.40	849.58
10	852.24	833.18	845.79	851.59
11	854.05	835.21	847.50	854.07
12	855.18	836.41	848.51	855.53
13	841.16	822.29	835.21	839.24
14	841.91	823.09	835.88	840.18
15	843.49	824.86	837.37	842.27
16	845.10	826.60	838.82	844.34
17	847.11	828.73	840.61	846.88
18	848.29	829.99	841.66	848.39
19	833.51	815.36	827.95	831.96
20	834.34	816.21	828.66	832.91
21	835.95	818.08	830.21	835.03
22	837.63	819.91	831.73	837.13
23	839.74	822.15	833.60	839.73
24	840.97	823.47	834.70	841.28
Octyl Amine	Experimental	SR-ANN	GA-ANN	GCM-ANN
1	779.35	779.82	795.49	781.45
2	775.6	776.20	792.27	778.02
3	771.85	772.57	789.02	774.55
4	768.10	768.92	785.73	771.06
5	764.33	765.25	782.42	767.53
6	760.61	761.57	779.06	763.95
7	756.82	757.85	775.66	760.32

جدول (۲-پ)- مقایسه‌ی مقادیر تجربی و مقادیر پیش‌بینی شده‌ی دانسیته برای
ترکیبات سری تست درونی بر حسب kg.m^{-3}

number	Experimental	SR-ANN	GA-ANN	GCM-ANN
1	846.1	846.3	846.41	846.6
2	848.6	849.0	849.01	849.1
3	851.1	851.6	851.5	851.6
4	862.7	863.8	862.9	863.0
5	818.7	818.2	818.1	819.4
6	819.3	818.8	818.7	819.9
7	798.0	797.5	796.5	798.9
8	798.7	798.2	797.2	799.6
9	801.7	801.2	800.3	802.4
10	834.5	834.8	833.9	834.7
11	783.6	783.4	781.3	784.3
12	784.3	784.2	782.0	785.1
13	861.3	863.1	861.2	860.8
14	863.4	865.3	863.3	863.0
15	811.4	811.8	812.5	811.2
16	846.1	846.7	845.5	845.7
17	789.7	788.9	790.4	789.7
18	790.4	789.6	791.1	790.5
19	778.7	777.6	778.9	778.5
20	783.0	781.9	783.0	782.7
21	787.1	785.9	786.9	786.8
22	808.4	803.5	803.6	808.3
23	824.5	823.0	822.0	824.4
25	793.6	787.5	788.1	793.8
26	825.7	824.8	825.0	825.9
27	846.4	846.4	849.7	845.9
28	849.0	848.9	852.6	848.5
29	830.1	829.6	832.6	829.8
30	833.0	832.2	835.7	832.6
31	811.1	810.5	811.9	811.0
32	814.2	813.4	815.3	814.1
33	786.4	787.5	785.4	786.4
34	790.4	791.1	789.6	790.3
35	866.3	866.5	865.2	867.0
36	871.3	871.5	871.0	872.2
37	875.8	876.6	877.9	876.5
38	884.8	885.0	887.2	884.8
39	865.7	864.7	865.7	866.4
40	873.4	872.9	873.2	874.1
41	841.8	841.8	839.2	842.1
42	844.4	844.6	841.8	844.7
43	817.3	816.6	813.9	817.5
44	820.4	819.9	817.1	820.5
45	793.9	793.2	789.4	793.9
46	797.6	797.1	793.2	797.4
48	805.4	806.6	806.5	805.6
49	774.0	774.7	775.8	772.7
51	778.7	779.7	780.6	777.6
52	779.6	780.7	781.6	778.6

ادامہ جدول (۲-پ)

number	Experimental	SR-ANN	GA-ANN	GCM-ANN
53	786.4	788.4	788.7	786.0
54	787.2	789.3	789.6	786.9
55	758.3	761.2	762.1	758.1
56	759.3	762.3	763.1	759.1
57	818.7	818.3	820.4	818.5
58	870.9	871.9	876.7	870.1
59	811.3	810.1	812.2	811.4
60	861.3	861.7	866.7	860.7
61	865.5	865.7	870.9	864.9
62	921.2	921.7	922.4	921.1
63	927.4	927.6	928.4	927.5
64	966.6	967.0	964.6	966.3
65	903.8	903.6	905.9	903.7
66	791.6	791.3	797.3	791.6
67	792.4	792.1	797.9	792.3
68	767.2	766.6	772.7	767.0
69	771.3	770.6	775.7	771.0
70	800.2	800.4	798.8	800.4
71	803.7	803.9	801.4	803.9
72	758.8	758.6	762.7	758.3
73	842.7	841.0	840.6	841.5
74	848.6	847.3	847.0	847.8
75	813.4	814.6	814.6	814.4
76	825.3	826.1	825.8	825.9
77	721.3	720.1	722.4	720.7
78	733.2	732.6	734.1	732.3
79	876.9	878.4	876.6	878.6
80	800.4	803.5	804.8	803.7
81	780.2	788.1	787.0	789.0
82	854.3	852.3	851.0	853.5
83	860.0	858.2	856.9	859.3
84	830.8	830.5	829.5	831.7
85	837.5	837.2	836.6	838.4
86	812.9	814.9	813.7	815.9
87	820.4	822.0	821.5	823.4
88	1053.4	1053.6	1053.1	1053.3
89	1055.5	1055.8	1055.8	1055.4
90	1071.4	1071.4	1075.5	1070.7
91	1025.3	1024.6	1021.3	1024.6
92	1048.1	1047.8	1048.4	1047.8
93	1050.1	1049.8	1050.7	1049.8
94	1027.9	1027.3	1028.4	1028.3
95	1030.2	1029.5	1030.8	1030.5
96	1029.2	1029.9	1029.4	1029.7
97	1031.2	1032.1	1031.8	1031.8
98	1008.9	1009.0	1007.9	1008.7
99	1011.2	1011.4	1010.4	1010.9
100	982.6	982.8	982.4	982.5
101	985.3	985.6	985.2	985.1
102	995.5	995.8	995.7	994.8
103	961.4	961.4	962.7	961.3

ادامہ جدول (۲-پ)

number	Experimental	SR-ANN	GA-ANN	GCM-ANN
104	964.5	964.6	965.9	964.4
105	827.4	826.1	827.9	827.5
106	828.8	827.5	829.5	829.1
107	877.4	875.4	878.1	880.8
108	781.9	783.2	777.7	782.5
109	865.1	864.4	866.4	865.8
110	871.4	870.5	872.5	873.1
111	887.3	888.1	885.0	887.9
112	893.0	893.9	890.9	893.6
113	859.7	859.4	857.2	859.3
114	866.7	866.6	864.2	866.3
115	843.6	843.9	842.3	844.2
116	851.1	851.4	849.8	851.6
117	888.6	886.6	891.5	890.4
118	894.1	892.8	897.4	895.2
119	875.7	874.3	877.3	876.4
120	881.8	880.4	883.5	882.6
121	841.6	842.4	842.9	841.9
122	787.0	785.6	784.2	787.0
123	799.2	798.5	797.4	799.6
124	772.6	772.0	771.2	771.6
125	804.3	804.8	802.9	803.6
126	808.5	809.1	807.1	807.5
127	795.8	797.9	797.6	796.1
128	799.2	801.1	801.2	799.4
129	995.2	994.6	994.5	994.6
130	983.3	983.5	982.1	982.9
131	983.8	983.9	982.6	983.4
132	977.1	977.6	976.4	976.9
133	981.9	982.1	981.8	981.7
134	958.3	957.0	959.3	959.1
135	978.2	978.0	974.6	977.5
136	948.7	948.1	949.9	949.4
137	949.6	949.0	950.6	950.3
138	938.3	938.6	940.5	939.2
139	938.8	939.1	940.8	939.7
140	954.0	955.9	949.4	953.7
141	958.4	959.9	951.9	956.8
142	950.5	951.5	944.2	949.1
143	920.2	919.4	923.2	919.4
144	911.2	909.9	915.5	910.2
145	911.4	910.2	915.7	910.5
146	944.9	945.1	948.1	944.1
147	917.0	916.8	916.1	916.6
148	908.4	908.6	907.4	907.6
149	909.9	910.1	909.1	909.4
150	912.4	912.5	911.9	912.2
151	903.9	904.4	903.2	903.4
152	927.0	927.4	927.7	929.0
153	931.1	930.6	931.5	932.4
154	923.6	923.0	923.5	924.9

ادامہ جدول (۲-پ)

number	Experimental	SR-ANN	GA-ANN	GCM-ANN
155	928.4	926.2	927.3	928.2
156	891.7	892.0	891.5	890.9
157	891.9	892.2	891.8	891.2
158	911.5	910.6	911.9	910.9
159	903.3	903.4	903.9	902.1
160	903.5	903.6	904.2	902.4
161	895.4	896.7	896.4	893.9
162	902.5	903.4	903.3	902.7
163	906.5	906.8	906.9	906.9
164	853.9	851.1	853.0	852.3
165	848.8	847.0	847.9	847.7
166	839.5	838.4	838.9	838.0
167	850.2	850.5	854.3	849.7
168	847.9	848.1	851.9	847.7
169	840.3	840.9	844.5	841.0
170	766.1	766.2	761.3	764.8
171	771.3	771.6	766.7	770.7
172	705.4	705.2	703.0	706.7
173	714.5	714.5	712.3	715.2
174	770.2	770.9	768.7	770.6
175	774.7	775.2	773.0	774.5
176	721.0	721.8	722.6	722.1
177	727.5	728.3	729.1	728.9
178	654.5	652.5	654.8	651.1
179	666.1	664.0	666.3	662.6
180	732.3	730.6	733.3	732.3
181	737.4	735.4	738.2	736.1
182	776.0	776.4	775.2	775.5
183	781.6	781.7	780.7	781.5
184	680.1	681.2	679.3	681.8
185	759.4	759.9	759.8	759.6
186	768.5	767.7	770.6	768.5
187	777.7	776.4	774.7	778.0
188	789.3	788.1	788.4	789.6
189	786.3	786.1	787.4	786.8
190	796.1	795.7	798.9	796.3
191	713.2	717.6	707.9	709.7
192	732.2	734.9	728.1	730.3
193	789.3	789.6	785.9	789.3
194	764.1	763.8	763.9	764.7
196	794.1	794.9	792.5	793.9
197	771.5	772.5	773.5	771.6
198	776.5	777.4	778.0	776.4
199	733.9	734.2	739.0	734.5
200	740.8	741.3	745.7	741.3
201	807.7	807.9	808.1	808.4
202	811.9	811.9	812.3	812.9
203	799.4	799.5	800.4	799.9
204	803.7	803.5	804.6	804.5
205	786.7	787.3	788.5	787.1
206	791.2	791.5	792.9	791.9

ادامہ جدول (۲-پ)

number	Experimental	SR-ANN	GA-ANN	GCM-ANN
207	764.6	765.5	766.4	764.3
208	770.0	770.6	771.6	769.6
209	811.8	812.7	810.4	810.9
210	824.0	824.5	822.5	824.6
211	827.8	828.1	826.1	828.8
212	816.2	816.2	814.8	816.4
213	820.1	819.7	818.4	820.8
214	786.2	787.4	786.2	785.8
215	791.1	792.1	790.9	790.3
216	741.9	741.7	741.7	743.5
217	749.3	749.6	749.2	751.3
218	948.8	948.7	948.8	949.0
219	979.3	978.5	978.2	978.9
220	932.6	934.1	932.2	933.1
221	1090.8	1091.3	1091.9	1091.2
222	773.0	772.7	780.7	774.7
223	780.9	782.1	790.5	777.5
224	694.1	694.3	705.9	693.8
225	747.9	748.4	751.1	747.8
226	739.9	739.7	743.0	739.4
227	768.6	766.9	768.9	770.3

فهرست منابع

- [1] H. Touba, G.A. Mansoori, "Equation of state for associating fluids (based on the analytic chain associating theory)", **Fluid Phase Equilibri.**, 119 (1996) 51-65.
- [2] L. Pauling, "The nature of the chemical Bond and the Stricture of Molecules and Crystals", 3rd Ed. Cornell University press, New York (1960).
- [3] D .M. Carey, G. M. Korenowski, "Measurement of the raman spectrum of liquid water", **J. Chem. Phys.** 108 (1998) 2669-2675.
- [4] L. A.Burke, J. O. Jensen, J. L. Jensen, P. N. Krishan, "theoretical study of water clusters. I. Pentamer" **Chem. Phys. Lett.**, 206 (1993) 293-296.
- [5] G. Franzese, K. Stokely, X.-Q. Chu, P. Kumar, M.G. Mazza, S- H. Chen, H. E. Stanley, **J. Phys.: Condens. Matter.** 20 (2008) 494210/1– 494210/10.
- [6] A. A. Rohani, Gh. Pazuki, H. Abedini Najafabadi, S. Seyfi, M. Vossoughi, "Comparison between the artificial neural network system and SAFT equation in obtaining vapor pressure and liquid density of pure alcohols", **Exp. Sys. Applicat.**, 38 (2011) 1738–1747.
- [7] J.M. Prausnitz, R. N. LichtenthalerAzevedo, " Molecular thermodynamics of fluid-phase equilibria", 2nd. Edit, Prantice-Hall, Englewood Cliffs, NJ (1999).
- [8] M. Moosavi, N. Soltani, "Prediction of hydrocarbon densities using an artificial neural network–group contribution method up to high temperatures and pressures", **Thermochim. Acta.**, 556 (2013) 89– 96.
- [9] J. Navas, "Practical Density Measurement and Hydrometry", IOP Publishing Ltd (2002).
- [10] J. Davila, R. Alcalde, M. Atilhan, S. Aparicio, "P ρ T measurements and derived properties of liquid 1-alkanols", **J. Chem. Thermodyn.**, 47 (2012) 241-259.
- [11] H. Geyer, P. Ulbig, M.Gornert, A. Susanto, "Measurement of densities and excess molar volumes for (1,2-propanediol, or 1,2-butanediol + water) at the temperatures (288.15, 298.15, and 308.15) K and at the pressures (0.1, 20, 40, and 60) MPa", **J. Chem. Thermodyn.**, 33 (2001) 987-997.
- [12] C. Boned, A. Baylaucq, J.P. Bazile, "Liquid density of 1-pentanol at pressures up to 140MPa and from 293.15 to 403.15K", **Fluid Phase Equilib.**, 270 (2008) 69–74.

- [13] R. L. Gardas, I. Johnson, D. M. D. Vaz, I. M. A. Fonseca, and A. G. M. Ferreira, "PVT Property Measurements for Some Aliphatic Esters from (298 to 393) K and up to 35 Mpa", **J. Chem. Eng. Data.**, 52 (2007) 737-751.
- [14] M. Atilhan, S. Aparicio, " $P\rho T$ measurements and derived properties of liquid 1,2-alkanediols", **J. Chem. Thermodyn.**, 57 (2013) 137–144.
- [15] D. Gonzalez-Salgado, J. Troncoso, F. Plantier, J.L. Daridon, D. Bessie`res, "Study of the volumetric properties of weakly associated alcohols by means of high-pressure speed of sound measurements", **J. Chem. Thermodyn.**, 38 (2006) 893–899.
- [16] S. Lago, P.A. Giuliano Albo, "Thermodynamic properties of acetone calculated from accurate experimental speed of sound measurements at low temperatures and high pressures", **J. Chem. Thermodyn.**, 41 (2009) 506–512.
- [17] G. Ihm, Y. Song & E. A. Mason, **J. Chem. Phys.**, 94 (1991) 3839.
- [18] A. R. Zabeti, B. Haghighi, M. M. Papari, "Prediction of density for normal alcohols based on statistical-thermodynamics prediction theory", **Indian J. Ppure Appl. Phys.**, 49 (2011) 460-466.
- [19] G.A. Parsafar, Z. Kalantar, "Extension of linear isotherm regularity to long chain primary, secondary and tertiary alcohols, ketones and 1-carboxylic acids by group contribution method", **Fluid Phase Equilib.**, 234 (2005) 11–21.
- [20] Z. Kalantar, H. Nikoofard, and F. Javadi, "Extension of LIR Equation of State to Alkylamines Using Group Contribution Method", **ISRN Physical Chemistry**, (2013).
- [21] E. Goharshadi, A. Morsali, M. Abbaspour, "New regularities and an equation of state for liquids", **Fluid Phase Equilib.**, 230 (2005) 170–175.
- [22] M. Moosavi, "Extension of GCM-GMA equation to long chain primary, secondary and tertiary alcohols, primary and secondary amines, and ketones using group contribution method", **Fluid Phase Equilib.**, 310 (2011) 63–73.
- [23] M. Moosavi, "Prediction of thermodynamic properties of long chain 1carboxylic acids and esters using a group contribution equation", **Fluid Phase Equilib.**, 316 (2012) 122–131.
- [24] D. L. Massart, B. G. Vandeginste, L. M. C. Buydens, "Handbook of Chemometrics and Qualimetrics Part A", **Elsevier**, Amesterdom, (1997).

[۲۵] محسنی ب، (۱۳۸۷)، پایان نامه ارشد، "مدل سازی تمایل اتصال بنزودیازپین ها به گیرنده گابا-A به روش QSPR با استفاده از تکنیک های الگوریتم ژنتیکی و شبکه ی عصبی مصنوعی"، دانشکده علوم، دانشگاه شهید باهنر کرمان.

- [26] L. Pauling, "The nature of the chemical Bond and the Stricture of Molecules and Crystals", 3nd Ed. Cornell University press, New York (1960).
- [27] M. Mitchell, "AN introduction to Genetic Algorithms", Cambridge, Massachusetts, (1999).
- [28] S. E. Randy, "Practical Genetic Algorithms", John Wiley & Sons, (2004).
- [29] D. A. Coley, "An Introduction to Genetic Algorithms for Scientists and Engineers", World Scientific, (1999).
- [30] M. N. Hassan and P. C. Jurs, "Advances in Quantative Structure – Property Relationships", **J. Anal. Chem.**, 62 (1990) 2318.
- [۳۱] کلانتر ز، (۱۳۸۵)، پایان نامه دکتری: "پیش بینی خواص ترمودینامیکی و انتقالی سیالات آلی چگال با استفاده از روش سهم گروه ها و شبکه ی عصبی موجک"، دانشکده ی شیمی، دانشگاه صنعتی اصفهان.
- [32] Y. Pan, J. Jiang, Z. Wang, "Quantitative structure–property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network", **J. Hazard. Mater**, 147 (2007) 424–430.
- [33] J. O. Valderrama, A. Reategui, R. E. Rojas, "Density of Ionic Liquids Using Group Contribution and Artificial Neural Networks", **Ind. Eng. Chem. Res.**, (2009).
- [34] F. Gharagheizi, A. Eslamimanesh, A. H. Mohammadi, "Determination of Critical Properties and Acentric Factors of Pure Compounds Using the Artificial Neural Network Group Contribution Algorithm", **J. Chem. Eng. Data**, 56, (2011) 2460–2476.
- [35] H. Matsuda, H. Yamamoto, K. Kurihara, K. Tochigi, "Computer-aided reverse design for ionic liquids by QSPR using descriptors of group contribution type for ionic conductivities and viscosities", **Fluid Phase Equilibr.**, 261 (2007) 434–443.
- [36] M. Moosavi, N. Soltani, "Prediction of hydrocarbon densities using an artificial neural network–group contribution method up to high temperatures and pressures", **Thermochim. Acta.**, 556 (2013) 89– 96.

[۳۷] کاوه، ع، ثروتی، ه، شبکه‌های عصبی مصنوعی در تحلیل و طراحی سازه‌ها، مرکز تحقیقات ساختمان و مسکن (۱۳۸۰).

[38] V.K. Stamations, Understanding neural networks and fuzzy logic, IEEE PRESS, (1996).

[۳۹] حسن زاده سورشجانی، ح، پیش‌بینی ویسکوزیته مخلوط چندتایی سیالات با استفاده از شبکه عصبی موجک، پایان‌نامه کارشناسی ارشد، دانشکده شیمی، دانشگاه صنعتی اصفهان، (۱۳۸۸).

[۴۰] منهای، م.ب، "هوش محاسباتی"، جلد اول، چاپ اول، مرکز نشر دانشگاه صنعتی امیرکبیر (واحد تفرش)، (۱۳۷۹).

[۴۱] مهناج، م، "مبانی شبکه‌های عصبی (هوش محاسباتی)" جلد اول، چاپ پنجم، مرکز نشر دانشگاه امیرکبیر، تهران، (۱۳۸۷).

[۴۲] کیا، م، "شبکه‌های عصبی در MATLAB"، چاپ دوم، انتشارات کیان رایانه سبز، (۱۳۸۷).

[۴۳] سعیدی، م، "عملیات شبکه‌های عصبی (قسمت دوم)"، ماهنامه شبکه، شماره ۵۲، (۱۳۷۹).

[۴۴] یزدان‌دوست، ح، پایان‌نامه ارشد، "پیش‌بینی شاخص بازداري تعدادی روغن‌های ضروری با استفاده از روش‌های QSPR خطی و غیر خطی"، دانشکده شیمی، دانشگاه صنعتی شاهرود، (۱۳۹۰).

[45] H. Kuba, Y. Tanaka, and T. Makita, "Volumetric Behavior of Pure Alcohols and Their Water Mixtures Under High Pressure", **Int. J. Thermophys.**, 8 (1987) 47-70.

[46] M. Dzida, "Study of the Effects of Temperature and Pressure on the Thermodynamic and Acoustic Properties of 2-Methyl-1-butanol at Temperatures from 293K to 318K and Pressures up to 100Mpa", **Int. J. Thermophys.**, 31 (2010) 55–69.

[47] S. Wappmann, N. Karger, and H.-D. Ludemann, "pVT Data of Liquid 1-, 2-, and 3-Pentanol from 10 to 200 MPa and from 233 to 433 K", **J. Chem. Eng. Data.**, 40 (1995) 233-236.

[48] E. Zorebski, M. Dzida, and M. Cempa, "Study of the Effects of Temperature and Pressure on the Acoustic and Thermodynamic Properties of 2-Methyl-2,4-pentanediol", **J. Chem. Eng. Data.**, 53 (2008) 1950–1955.

- [49] A. Zuniga-Moreno and L.A. Galicia-Luna, "Densities, Isothermal Compressibilities, and Isobaric Thermal Expansivities of Hexan-2-ol, Octan-1-ol, and Decan-1-ol from (313 to 363) K and Pressures up to 22 MPa", **J. Chem. Eng. Data.**, 52 (2007) 1773-1783.
- [50] A. Zuniga-Moreno, L. A. Galicia-Luna, F. F. Betancourt-Cardenas, "Densities and derived thermodynamic properties of 1-heptanol and 2-heptanol at temperatures from 313 K to 363 K and pressures up to 22 MPa", **J. Chem. Thermodyn.**, 40 (2008) 96–106.
- [51] S. Matsua and T. Makita, "Volumetric Properties of 1-Alkanols at Temperatures in the Range 298-348 K and Pressures up to 40 MPa", **Int. J. Thermophys.**, 10 (1989) 885-897.
- [52] D. Papaioannou, M. Bridakis, and C. G. Panayiotou, "Excess Dynamic Viscosity and Excess Volume of N-Butylamine + 1-Alkanol Mixtures at Moderately High Pressures", **J. Chem. Eng. Data.**, 38 (1993) 370-378.
- [53] M. Yoshimura, A. Baylaucq, J.P. Bazile, H. Ushiki, and C. Boned, "Volumetric Properties of 2-Alkylamines (2-Aminobutane and 2-Aminooctane) at Pressures up to 140 MPa and Temperatures between (293.15 and 403.15) K", **J. Chem. Eng. Data.**, 54 (2009) 1702–1709.
- [54] M. Yoshimura, C. Boned, G. Galliero, J.P. Bazile, A. Baylaucq, H. Ushiki, "Influence of the chain length on the dynamic viscosity at high pressure of some 20 Alkylamines: Measurements and comparative study of some models", **Chem. Phys.**, 369 (2010) 126-137.
- [55] Y. Miyake, A. Baylaucq, F. Plantier, D. Bessieres, H. Ushiki, C. Boned, "High-pressure (up to 140 MPa) density and derivative properties of some (pentyl-, hexyl-, and heptyl-) amines between (293.15 and 353.15) K", **J. Chem. Thermodyn.**, 40 (2008) 836–845.
- [56] E. Alvarez, F. Cerdeira, D. Gomez-Diaz, and J. M. Navaza, "Density, Speed of Sound, Isentropic Compressibility, and Excess Volume of Binary Mixtures of 1-Amino-2-propanol or 3-Amino-1-propanol with 2-Amino-2-methyl-1-propanol, Diethanolamine, or Triethanolamine from (293.15 to 323.15) K", **J. Chem. Eng. Data.**, 55 (2010) 2567–2575.

- [57] S.L. Oswal, P. Oswal, R.L. Gardas, S.G. Patel, R.G. Shinde, "Acoustic, volumetric, compressibility and refractivity properties and reduction parameters for the ERAS and Flory models of some homologous series of amines from 298.15 to 328.15 K", **Fluid Phase Equilib.**, 216 (2004) 33–45.
- [58] W. T. Vong and F. N. Tsai, "Densities, Molar Volumes, Thermal Expansion Coefficients, and Isothermal Compressibilities of Organic Acids from 293.15 K to 323.15 K and at Pressures up to 25 MPa", **J. Chem. Eng. Data.**, 42 (1997) 1116-1120.
- [59] T. S. Banipal, S. K. Gage, and J. C. Ahluwalia, "Densities of some higher alkan-1-oic acids at temperatures from 343.15 K to 373.15 K and at pressures up to 9 MPa", **J. Chem. Thermodyn.**, 24 (1992) 729-735.
- [60] A. Toubaie, H. Golmohammadi, Z. Dashtbozorgi, W. E. Acree, "QSPR studies for predicting gas to acetone and gas to acetonitrile solvation enthalpies using support vector machine", **J. Mol. Liq.**, 175 (2012) 24-32.
- [61] http://www.strandls.com/sarchitect/documents/manual_html/desctheory.html
- [62] R. Todeschini, V. Consonni, "Handbook of Molecular Descriptors", John Wiley-VCH, New York, pp 84.
- [63] R. Todeschini, V. Consonni, "Handbook of Molecular Descriptors", John Wiley-VCH, New York, pp 17-18.
- [64] R. Todeschini, V. Consonni, "Handbook of Molecular Descriptors", John Wiley-VCH, New York, pp 91.
- [65] R. Todeschini, V. Consonni, "Handbook of Molecular Descriptors", John Wiley-VCH, New York, pp 366.

Abstract

A quantitative structure-property relationship (QSPR) approach was employed to predict the density of associative fluids such as alcohols, primary and secondary amines and carboxylic acids. Two different methods were used to generate the significant descriptors. In the first method, a large number of descriptors was calculated using Dragon software and the best calculated descriptors were selected from 18 classes of Dragon descriptors by stepwise regression (SR) and genetic algorithm based on partial least square (GA-PLS) methods. 8 descriptors with SR and 5 descriptors with GA-PLS methods were selected. In the second method, the best descriptors are selected based on group contribution method. After analyzing the chemical structures of all compounds in this work, 10 functional groups ($-\text{CH}_3$, $-\text{CH}_2-$, $>\text{CH}-$, $>\text{CH}-\text{OH}$, $-\overset{|}{\text{C}}-\text{OH}$, $-\text{COOH}$, $-\text{CH}_2-\text{NH}_2$, $>\text{CH}-\text{NH}_2$, $>\text{NH}$) were found useful. The selected descriptors with three feature selection methods and two experimental descriptors (pressure and temperature) were used as input nodes for constructing different artificial networks. 4 feed forward artificial neural networks were designed by combination of Levenberg-Marquardt and Bayesian regularized (BR) algorithms with logarithm-sigmoid and tangent-sigmoid transfer functions. After training and optimization of the ANNs parameters such as, number of neurons in input and hidden layers, and number of iteration (epoch), the performance of the models was investigated by the test set. The results showed that the Levenberg-Marquardt artificial neural network with tangent-sigmoid transfer function and GCM based descriptors, may be simulated the relationship between the structural descriptors and density of the desired molecules accurately. The mean square errors (*MSEs*) and Absolute Average percent Deviation (*AADs*) for the test set are 1.2432 and 0.0867%, respectively.

Keywords: quantitative structure–property relation (QSPR), artificial neural network (ANN), Density, SR, associative fluids, group contribution method (GCM)



Shahrood University of Technology
Faculty of Chemistry
M.Sc. Thesis in Physical Chemistry

ANN modeling for density prediction of associative fluids using their molecular structure descriptors

Nazli Azmoonfar

Supervisors:

Dr. Z. Kalantar

Advisor:

Dr. M. Arab Cham jangali

February 2014