

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده: شیمی

گروه شیمی فیزیک

پیش‌بینی دانسیته‌ی مخلوط ترکیبات آلی با استفاده از روش‌های

غیر خطی

نگارنده: بنت‌الهدی گودرزی

استاد راهنما:

دکتر زهرا کلانتر

استاد مشاور:

دکتر ناصر گودرزی

پایان‌نامه جهت اخذ درجه کارشناسی ارشد

بهمن ماه ۱۳۹۵



مدیریت تحصیلات تکمیلی

باسمه تعالی

شماره: ۴۱۱۳. س. ک.

تاریخ: ۹۵/۱۲/۱۶

ویرایش:

فرم شماره ۷: صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای بنت‌الهدی گودرزی به شماره دانشجویی ۹۳۱۴۷۸۴ رشته شیمی گرایش شیمی فیزیک تحت عنوان پیش‌بینی دانسیته مخلوط ترکیبات آلی با استفاده از روش‌های غیر خطی که در تاریخ ۹۵/۱۱/۱۸ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می‌گردد:

قبول (با درجه: عالی - امتیاز ۱۹.۵)	دفاع مجدد	مردود
نوع تحقیق: نظری ☒ عملی ☐		

۱- عالی (۲۰ - ۱۹)

۲- بسیار خوب (۱۸ - ۱۷/۹۹)

۳- خوب (۱۶ - ۱۷/۹۹)

۴- قابل قبول (۱۵ - ۱۴)

۵- نمره کمتر از ۱۴ غیر قابل قبول

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنمای اول	دکتر زهرا کلاتر	استادیار	
۲- استاد مشاور	دکتر ناصر گودرزی	دانشیار	
۳- نماینده شورای تحصیلات تکمیلی	دکتر مهدی میرزائی	استادیار	
۴- استاد ممتحن اول	دکتر حسین نیکوفرد	دانشیار	
۵- استاد ممتحن دوم	دکتر زینب موسوی تکیه	استادیار	

نام و نام خانوادگی رئیس دانشکده: دکتر قاسم علی باقریان

تاریخ و امضاء و مهر دانشکده:

تقدیم بہ

حمایت ایزدمنان

دعا و آرزوهای پدر و مادر بزرگوارم

ہمسرہ شینا نم

خواہر و برادر عزیزم

وروشنی بخش زندگی ام امیر علی

تشکر و قدردانی

پاس بی پایان پروردگار جهانیان که توانایی عطا فرمود تا بتوانم قدمی در زندگی پیش بگذارم.

اینک که با لطف خداوند متعال تدوین این پایان نامه را به اتمام رسانده ام، لازم می دانم از استاد برجسته و کرامتدارم سرکار خانم دکتر

کلاشگر که با زحمات و راهنمایی های فراوان و درکی مادرانه مراد به پایان رسانیدن این تحقیق یاری نمودند، تشکر و قدردانی نمایم.

بابتشکر فراوان از استاد عزیز، جناب آقای دکتر کوردزی که با نظرات و راهنمایی هایشان مراد انجام این پایان نامه یاری نمودند، همچنین از

دو استاد گرامی ام سرکار خانم دکتر موسوی و جناب آقای دکتر نیکوفرو که علاوه بر داوری این کار سرکلاس های درس این اساتید

بودم و بهره بسیار بردم سپاسگذارم.

در این تلاش کوچک از پدر و مادر عزیزم که انقبای زندگی بمن آموختند و یاری ام کردند و همسرم که همیشه برایم یک سکوی پرتاب

بوده، خواهر و برادر مشوقم سپاسگذارم.

در پایان از دوستان ماندگارم، خانم ناهید رحمتی و فاطمه داوودی نیز به خاطر تمام زحمات و حمایت هایشان در تمام طول این دوره تشکر و

قدردانی می نمایم.

فهرست مطالب

فصل اول: مقدمه

۱-۱-مقدمه.....	۲
۱-۲-دانشیته.....	۲
۱-۳-روش های پیش بینی دانشیته ی مخلوط ها.....	۲
۱-۳-۱-پیش بینی دانشیته مخلوط های آلی با استفاده از معادله ی حالت MLIR.....	۳
۱-۳-۲-پیش بینی دانشیته مخلوط های آلی با استفاده از معادله حالت GCM-GMA.....	۶
۱-۴-هدف تحقیق.....	۷

فصل دوم: کمومتری کس

۲-۱-کمومتری کس.....	۱۰
۲-۲-ارتباط کمی ساختار- خاصیت و ساختار- فعالیت.....	۱۰
۲-۲-۱-رابطه کمی ساختار - فعالیت (QSAR).....	۱۰
۲-۲-۲-رابطه کمی ساختار - خاصیت (QSPR).....	۱۱
۲-۳-مراحل مدل سازی QSPR.....	۱۲
۲-۳-۱-فراهم کردن سری داده ها.....	۱۲
۲-۳-۲-رسم و بهینه سازی ساختار هندسی مولکول ها.....	۱۳
۲-۳-۳-محاسبه ی توصیف کننده های تئوری.....	۱۳
۲-۳-۴-حذف توصیف کننده های نامناسب.....	۱۵
۲-۳-۵-تقسیم بندی داده ها.....	۱۵
۲-۳-۶-انتخاب بهترین توصیف کننده ها.....	۱۶
۲-۳-۶-۱-انتخاب بهترین توصیف کننده ها توسط الگوریتم ژنتیک.....	۱۶
۲-۳-۶-۱-الف-مقدمه.....	۱۶
۲-۳-۶-۱-ب-الگوریتم ژنتیک.....	۱۶
۲-۳-۶-۱-ج-نحوه ی انتخاب توصیف کننده ها و شرایط توقف الگوریتم ژنتیک.....	۱۸
۲-۳-۷-انتخاب توصیف کننده ها بر مبنای روش سهم گروه (GCM).....	۱۹
۲-۳-۸-ساختن مدل.....	۲۱

۲۲.....	۸-۳-۲- شبکه عصبی.....
۲۳.....	۸-۳-۲- الف- ساختار شبکه عصبی مصنوعی.....
۲۵.....	۸-۳-۲- ب- انواع شبکه‌ها از نظر ارتباطات بین نرونی.....
۲۵.....	۸-۳-۲- ج- روش‌های آموزش.....
۲۷.....	۹-۳-۲- ارزیابی قدرت پیش‌بینی مدل.....
۲۷.....	۹-۳-۲- الف- استفاده از پارامترهای آماری.....
۲۹.....	۹-۳-۲- ب- استفاده از نمودار برگشتی.....
۳۰.....	۹-۳-۲- ج- استفاده از نمودار خطای باقیمانده.....
۳۰.....	۹-۳-۲- د- استفاده از سری تست خارجی.....

فصل سوم: محاسبات

۳۲.....	۱-۳- انتخاب سری داده‌ها.....
۳۸.....	۲-۳- محاسبه‌ی توصیف‌کننده‌ها.....
۳۹.....	۳-۳- دسته‌بندی داده‌ها.....
۴۱.....	۴-۳- مدل‌سازی شبکه‌عصبی.....
۴۱.....	۱-۴-۳- مدل‌سازی شبکه‌ی عصبی مصنوعی با استفاده از توصیف‌کننده‌های سهم‌گروه.....
۴۱.....	۱-۱-۴-۳- محاسبه‌ی توصیف‌کننده‌های مخلوط‌ها با استفاده از روش سهم‌گروه.....
۴۴.....	۲-۱-۴-۳- مدل‌سازی شبکه‌ی عصبی با توصیف‌کننده‌های حاصل از روش سهم‌گروه GCM-ANN.....
۵۰.....	۳-۱-۴-۳- ارزیابی شبکه‌ی بهینه‌شده‌ی GCM-ANN.....
۵۰.....	۴-۱-۴-۳- اثر افزایش توصیف‌کننده‌ی جرم‌مولی‌کل برای مدل‌سازی GCM-ANN دانسیته‌ی مخلوط‌ها.....
۶۲.....	۲-۴-۳- مدل‌سازی شبکه‌ی عصبی مصنوعی با استفاده از توصیف‌کننده‌های تئوری.....
۶۲.....	۱-۲-۴-۳- محاسبه‌ی توصیف‌کننده‌های تئوری برای مخلوط‌ها.....
۶۳.....	۲-۲-۴-۳- انتخاب بهترین توصیف‌کننده‌ها به روش الگوریتم ژنتیک.....
۶۷.....	۳-۲-۴-۳- بهینه‌سازی شبکه‌ی عصبی مصنوعی با توصیف‌کننده‌های الگوریتم ژنتیک.....
۷۰.....	۴-۲-۴-۳- ارزیابی شبکه‌ی بهینه‌شده‌ی GA-ANN.....
۷۱.....	۳-۵- ارزیابی قدرت پیش‌بینی مدل‌های غیر خطی بهینه‌شده.....
۷۲.....	۶-۳- استفاده از پارامترهای آماری.....
۷۲.....	۷-۳- بررسی ارتباط توصیف‌کننده‌های وارد شده در مدل با دانسیته‌ی مخلوط‌ها.....

۷۴	۳-۷-۱- توصیف کننده های گروه MATS1m
۷۴	۳-۷-۲- توصیف کننده های Ds
۷۹	۳-۷-۳- توصیف کننده های گروه R1m-A
۸۲	۳-۷-۴- توصیف کننده های H047
۸۴	۳-۸- بررسی میزان مشارکت توصیف کننده های منتخب شبکه ی عصبی بهینه
۸۶	۳-۹- نتیجه گیری
۸۷	۳-۱۰- آینده نگری
۸۸	منابع

فهرست اشکال

شکل (۱-۲) - ساختار یک نرون زیستی.....	۲۳
شکل (۲-۲) - ساختار یک شبکه عصبی مصنوعی.....	۲۴
شکل (۱-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۵-۳).....	۴۵
شکل (۲-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۶-۳).....	۴۶
شکل (۳-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۷-۳).....	۴۷
شکل (۴-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۸-۳).....	۴۸
شکل (۵-۳) - نمودار میانگین مربعات خطا بر حسب μ حول نقطه بهینه برای سری ارزیابی.....	۴۹
شکل (۶-۳) - نمودار مقادیر پیش‌بینی شده‌ی دانسیته‌ی مخلوط‌ها بر حسب مقادیر تجربی برای (الف)	
سری ارزیابی (ب) سری تست.....	۵۱
شکل (۷-۳) - نمودار خطای باقیمانده بر حسب مقادیر تجربی دانسیته‌ی مخلوط‌ها (الف) سری ارزیابی	
(ب) سری.....	۵۱
شکل (۸-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۱۱-۳).....	۵۳
شکل (۹-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۱۲-۳).....	۵۴
شکل (۱۰-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول	
(۱۳-۳).....	۵۵
شکل (۱۱-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول	
(۱۴-۳).....	۵۶
شکل (۱۲-۳) - نمودار میانگین مربعات خطا بر حسب μ حول نقطه بهینه برای سری ارزیابی شبکه‌ی	
GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل.....	۵۸
شکل (۱۳-۳) - نمودار برگشتی مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری	
ارزیابی توسط مدل GCM-ANN بهینه (الف) بدون جرم‌مولی کل و (ب) در حضور جرم‌مولی کل.....	۵۹
شکل (۱۴-۳) - نمودار برگشتی مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست	
توسط مدل GCM-ANN بهینه (الف) بدون جرم‌مولی کل و (ب) در حضور جرم‌مولی کل.....	۵۹
شکل (۱۵-۳) - نمودار مقادیر خطای باقیمانده دانسیته بر حسب مقادیر تجربی برای سری ارزیابی مدل	
GCM-ANN بهینه شده (الف) بدون جرم‌مولی کل و (ب) در حضور جرم‌مولی کل.....	۶۰
شکل (۱۶-۳) - نمودار مقادیر خطای باقیمانده دانسیته بر حسب مقادیر تجربی برای سری تست توسط	
مدل GCM-ANN بهینه شده (الف) بدون جرم‌مولی کل و (ب) در حضور جرم‌مولی کل.....	۶۰

- شکل (۳-۱۷) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ ۶۹
- شکل (۳-۱۸) - نمودار برگشتی مقادیر پیش‌بینی شده‌ی دانسیته‌ی مخلوط‌ها بر حسب مقادیر تجربی (الف) سری ارزیابی (ب) سری تست ۷۰
- شکل (۳-۱۹) - نمودار خطای باقیمانده دانسیته‌ی مخلوط‌ها بر حسب مقادیر تجربی (الف) سری ارزیابی (ب) سری تست ۷۰
- شکل (۳-۲۰) - درصد مشارکت توصیف‌کننده‌ها در مدل GA-ANN ۸۴

فهرست جدول‌ها

جدول (۱-۲)- انواع توصیف‌کننده‌های محاسبه شده توسط نرم افزار Dragon.....	۱۴
جدول (۱-۳)- نام مخلوط‌ها ، محدوده‌ی دما و فشار و تعداد نقاط مورد استفاده برای هر مخلوط به همراه مرجع مورد استفاده برای هر یک از آن‌ها.....	۳۳
جدول (۲-۳)- گروه‌های عاملی موجود در اجزای مخلوط‌های مورد مطالعه در این پایان‌نامه طبق روش سهم‌گروه.....	۴۱
جدول (۳-۳) نحوه‌ی آرایش توصیف‌کننده‌های مخلوط 0.2 Hexane + 0.8 Cyclohexanone در دما و فشار محیط بر اساس روش سهم‌گروه.....	۴۲
جدول (۴-۳)- مقادیر همبستگی توصیف‌کننده‌های روش سهم‌گروه با دانسیته.....	۴۳
جدول (۵-۳)- 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.....	۴۵
جدول (۶-۳)- مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۴۶
جدول (۷-۳)- 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.....	۴۷
جدول (۸-۳)- مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی لونبرگ- مارکوارت و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۴۸
جدول (۹-۳)- توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN.....	۴۹
جدول (۱۰-۳)- توابع و مقادیر پارامترهای بهینه شده‌ی شبکه‌ی GCM-ANN برای پیش‌بینی دانسیته‌ی مخلوط‌ها برحسب دما و فشار.....	۵۰
جدول (۱۱-۳)- 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.....	۵۳
جدول (۱۲-۳)- مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی لونبرگ مارکوارت و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۵۴
جدول (۱۳-۳)- 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.....	۵۵
جدول (۱۴-۳)- مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی لونبرگ مارکوارت و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت.....	۵۶

جدول (۳-۱۵) - توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل مخلوط.....	۵۸
جدول (۳-۱۶) - توابع و مقادیر پارامترهای مدل بهینه شده‌ی GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل.....	۵۹
جدول (۳-۱۷) - تعداد دفعات تکرار توصیف‌کننده‌ها در روش انتخابی الگوریتم ژنتیک پس از ۱۶ بار اجرا.....	۵۳
جدول (۳-۱۸) - طبقه و نام کامل توصیف‌کننده‌های تئوری انتخابی توسط الگوریتم ژنتیک.....	۶۴
جدول (۳-۱۹) - همبستگی بین توصیف‌کننده‌های انتخابی الگوریتم ژنتیک.....	۶۵
جدول (۳-۲۰) - مقادیر همبستگی توصیف‌کننده‌های انتخابی با روش GA با دانسیته.....	۶۷
جدول (۳-۲۱) - توابع و پارامترهای بهینه‌ی شبکه GA-ANN.....	۶۸
جدول (۳-۲۲) - توابع و پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های تئوری انتخابی GA.....	۶۹
جدول (۳-۲۳) - پارامترهای آماری مدل‌های بهینه شده‌ی GA-ANN و GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل برای دو سری ارزیابی و تست.....	۷۳
جدول (۳-۲۴) - مثال‌هایی از مقدار توصیف‌کننده MATS1m در تعدادی از کسرمولی‌ها.....	۷۶
جدول (۳-۲۵) - مثال‌هایی از مقدار توصیف‌کننده Ds در تعدادی از کسرمولی‌ها.....	۷۸
جدول (۳-۲۶) - مثال‌هایی از مقدار توصیف‌کننده R1m-A در تعدادی از کسرمولی‌ها.....	۸۱
جدول (۳-۲۷) - مثال‌هایی از مقدار توصیف‌کننده H047 در تعدادی از کسرمولی‌ها.....	۸۳

تعهد نامه

اینجانب بنت الهدی گودرزی دانشجوی دوره کارشناسی ارشد رشته شیمی فیزیک دانشکده شیمی دانشگاه شاهرود نویسنده پایان نامه پیش‌بینی دانسیته‌ی مخلوط ترکیبات آلی با استفاده از روش‌های غیرخطی تحت راهنمایی دکتر زهرا کلانتر و مشاوره دکتر ناصر گودرزی متعهد می‌شوم :

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

امضای دانشجو

تاریخ

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده

در این تحقیق، روش ارتباط کمی ساختار- خاصیت (QSPR)، برای پیش‌بینی دانسیته‌ی مخلوط مایعات آلی در محدوده‌ی وسیعی از دما و فشار به کار گرفته شد. از دو نوع توصیف‌کننده برای اجزای مخلوط جهت ایجاد توصیف‌کننده‌های مخلوط استفاده شد: توصیف‌کننده بر مبنای روش سهم‌گروه‌ها و توصیف‌کننده‌های تئوری. در روش اول، توصیف‌کننده‌های اجزای مخلوط بر اساس روش سهم‌گروه (GCM) انتخاب شدند. بعد از آنالیز تمام اجزای مخلوط‌های مورد مطالعه در این تحقیق، ۱۵ گروه عاملی مشخص شدند. در روش دوم، ۱۴۸۱ توصیف‌کننده با استفاده از نرم‌افزار دراگون برای هر یک از اجزای مخلوط به طور جداگانه محاسبه شد. بعد از محاسبه‌ی توصیف‌کننده‌ها برای اجزای مخلوط توسط دو روش توصیف‌کننده‌های مخلوط به کمک مجموع وزن‌دار شده‌ی مولی با استفاده از مقدار توصیف‌کننده و کسر مولی هر جزء خالص در مخلوط، محاسبه شدند. به این ترتیب ۱۵ توصیف‌کننده برای مخلوط‌ها توسط روش سهم‌گروه محاسبه شد. همچنین، بهترین توصیف‌کننده‌ها توسط روش الگوریتم ژنتیک بر اساس آنالیز حداقل مربعات (GA-PLS) از تعداد زیادی از توصیف‌کننده‌های تئوری استخراج شده از مخلوط‌ها انتخاب شدند. ۲۲ توصیف‌کننده تئوری توسط روش GA-PLS استخراج شد. توصیف‌کننده‌های انتخاب شده توسط این روش‌ها و دو متغیر تجربی (دما و فشار) به همراه جرم‌مولی کل مخلوط به عنوان ورودی شبکه‌ی عصبی مصنوعی مورد استفاده قرار گرفتند. ۴ نوع شبکه‌ی عصبی مصنوعی با ترکیب الگوریتم‌های آموزشی لوبز-مارکوارت و بایزین با توابع انتقال لگاریتم سیگموئید و تانژانت سیگموئید طراحی گردید. پس از آموزش و بهینه‌سازی 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 بایزین و تابع انتقال تانژانت سیگموئید و توصیف‌کننده‌های حاصل از روش سهم‌گروه می‌تواند به درستی رابطه‌ی بین توصیف‌کننده‌های ساختاری و دانسیته‌ی مخلوط‌های موردنظر را شبیه‌سازی کند.

میانگین مربعات خطا (MSE) و ضریب تعیین (R^2) برای سری تست توسط این روش به ترتیب برابر ۱۱/۰۵۱ و ۰/۹۹۹۶ می‌باشد.

کلمات کلیدی: ارتباط کمی ساختار-خاصیت (QSPR)، شبکه عصبی مصنوعی (ANN)، دانسیته مخلوط، الگوریتم ژنتیک (GA)، روش سهم‌گروه (GCM).

نتایج حاصل از این پایان نامه در دو مقاله تحت عناوین

Density prediction of *n*-alkane and cycloalkane mixtures using calculated molecular descriptors and Levenberg–Marquardt artificial neural network

Prediction of density for primary alcohols mixtures using linear and nonlinear methods

در نوزدهمین سمینار بین المللی شیمی فیزیک ایران در شهریور ماه ۱۳۹۵ ارائه گردید.

فصل اول

مقدمه

۱-۱- مقدمه

یکی از خواص ترمودینامیکی مهم که برای توصیف یک مایع به کار می‌رود دانسیته است. دانستن دانسیته‌ی سیالات خالص و مخلوط‌های آن‌ها در شرایط مختلف دما و فشار برای طراحی بهینه‌ی واحدهای عملیاتی صنایع شیمیایی مختلف که در آن‌ها انتقال جرم و جریان سیال وجود دارد ضروری است. با مخلوط شدن دو یا چند مایع امتزاج‌پذیر دانسیته‌ی مخلوط نسبت به ترکیب آن به صورت غیرخطی تغییر می‌کند که این انحراف از رفتار خطی به دلیل برهم‌کنش بین مولکول‌ها در مخلوط است. با وجود اهمیت دانسیته‌ی مخلوط مایعات در صنایع مختلف شیمیایی، اندازه‌گیری تجربی دانسیته‌ی مخلوط‌ها در شرایط دما و فشار گوناگون مشکل بوده و به صرف هزینه و وقت زیادی نیاز دارد زیرا حتی در دما و فشار ثابت و معین، بی‌نهایت ترکیب مختلف برای هر مخلوط با اجزای معین وجود دارد. در این شرایط، استفاده از روش‌های پیش‌بینی از اهمیت ویژه‌ای برخوردار است.

۱-۲- 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 دیگر مثل g cm^{-3} یا g ml^{-1} نیز استفاده می‌شوند. اگر شرایط اندازه‌گیری دانسیته، شرایط ترمودینامیکی نرمال ($p=101325 \text{ kPa}$) یا شرایط استاندارد ($p=101325 \text{ kPa}$) باشد، دانسیته‌ی اندازه‌گیری شده به ترتیب، دانسیته‌ی نرمال و دانسیته‌ی استاندارد نامیده می‌شود [۱].

۱-۳- روش‌های پیش‌بینی دانسیته‌ی مخلوط‌ها

همان‌طور که در مقدمه شرح داده شد خواص ترمودینامیکی و تعادل فازی سیالات خالص و مخلوط‌های آن‌ها، دو ابزار مهم برای طراحی مهندسی و آنالیز فرآیندهای شیمیایی هستند. اگرچه داده‌های تجربی دقیق هستند اما اندازه‌گیری آن‌ها در یک دامنه‌ی وسیع دما، فشار و کسرمولی کاری

وقت گیر و هزینه بر است. لذا باید از روش های پیش بینی استفاده نمود. یکی از مهم ترین روش های پیش بینی دانسیته ی مخلوط ها استفاده از معادلات حالت است. معادله ی حالت یک رابطه ی ریاضی است که فشار را به صورت تابعی از دما و دانسیته بیان می کند. با استفاده از معادله ی حالت و یک مشتق مناسب از آن می توان سایر خواص ترمودینامیکی را نیز به دست آورد. دو معادله ی حالتی که در سال های اخیر برای پیش بینی خواص ترمودینامیکی مخلوط ترکیبات آلی برحسب دما و فشار معرفی شده اند و از کارایی بالایی برخوردارند معادلات حالت ¹MLIR و ²GCM-GMA هستند. در ادامه بطور مختصر راجع به هر کدام از آن ها توضیح داده خواهد شد.

۱-۳-۱- پیش بینی دانسیته مخلوط های آلی با استفاده از معادله ی حالت MLIR

قاعده ی همدمای خطی ^۳(LIR) قاعده ای است که در سال ۱۹۹۳ برای سیالات چگال زیر بحرانی و فوق بحرانی براساس مفهوم متوسط پتانسیل جفت مؤثر ^۴(AEPP) به دست آمده و دارای مبنای نظری است. این قاعده بیان می کند که، کمیت $(Z-1)v^2$ برحسب ρ^2 برای هر همدمای سیال چگال خطی است، که $\rho = \frac{1}{v}$ دانسیته ی مولی سیال است.

$$(Z-1)v^2 = A + B\rho^2 \quad (1-1)$$

در این رابطه پارامترهای A و B عبارتند از:

$$A = A_1 - \frac{A_2}{RT} \quad (2-1)$$

$$B = B_1 + \frac{B_2}{RT} \quad (3-1)$$

۱-Modified Linear Isotherm Regularity (MLIR)

۲-Group Contribution Method- Goharshadi-Morsali-Abbaspour (GCM-GMA)

۳-Linear Isotherm Regularity (LIR)

۴-Average Effective Pair Potential

قاعده‌ی همدمای خطی، LIR، رفتار خطی فوق را برای سیالات چگال قطبی و غیرقطبی با مولکولهای کوچک در محدوده‌های $\rho > \rho_B$ و $T < 2T_B$ نشان می‌دهد که ρ_B و T_B به ترتیب دانسیته و دمای بویل سیال می‌باشند [۲]. این معادله در سال ۱۹۹۴، با استفاده از تقریب تک سیال وان‌دروالس^۱ به مخلوط سیالات چگال نیز تعمیم داده شده است [۳]. بر اساس این تقریب که یکی از معروف‌ترین قواعد اختلاط^۲ است، هر معادله‌ی حالتی که برای سیال خالص قابل استفاده باشد عیناً برای مخلوط سیالات نیز کارایی دارد با این تفاوت که پارامترهای معادله به ترکیب سیستم نیز بستگی خواهد داشت. با استفاده از تقریب توزیع تصادفی و همچنین مدلی که برای استخراج خود معادله‌ی حالت به کار رفته بود وابستگی پارامترهای LIR به ترکیب سیستم به صورت زیر ارائه شده است:

$$B_{\text{mix}} = \sum_i \sum_j x_i x_j B_{ij} \quad (۴-۱)$$

$$\left(\frac{A}{B}\right)_{\text{mix}} = \sum_i \sum_j x_i x_j \left(\frac{A_{ij}}{B_{ij}}\right) \quad (۵-۱)$$

که در آن x_i کسر مولی گونه‌ی i در مخلوط بوده و B_{ij} و (A_{ij}/B_{ij}) مقدار پارامترها برای مخلوطی است که تمام برهم کنش‌های آن از نوع ij باشند. سپس با استفاده از قاعده‌ی ترکیبی لورنتز-برتولت پارامترهای ناجور، B_{ij} و (A_{ij}/B_{ij}) برحسب پارامترهای جور^۳ به صورت زیر به دست آمده‌اند:

$$B_{ij} = \sqrt{B_{ii} B_{jj}} \quad (۶-۱)$$

$$\left(\frac{A_{ij}}{B_{ij}}\right) = \sqrt{\left(\frac{A_{ii}}{B_{ii}}\right) \left(\frac{A_{jj}}{B_{jj}}\right)} \quad (۷-۱)$$

که معادلات فوق، تقریب متوسط هندسی^۴ نامیده می‌شوند. لازم به تذکر است که در یک مخلوط دوتایی پارامتر A یک تابع پیچیده‌ی درجه‌ی چهارم از ترکیب سیستم است. از این رو، نسبت $\frac{A}{B}$ که

۱-Vander Waals one-fluid approximation

۲-Combining rule

۳-like

۴-mean geometric approximation

می‌تواند به صورت یک تابع درجه‌ی دوم از ترکیب سیستم بیان شود به جای پارامتر A در معادله‌ی (۵-۱) آمده است [۳].

در سال ۲۰۰۳ پارسا و کلانتر، با استفاده از روش سهم‌گروه، توانستند معادله‌ی حالت LIR را برای پیش‌بینی خواص ترمودینامیکی ترکیبات آلی بلند زنجیر و مخلوط‌های آن‌ها اصلاح کنند [۴]. آن‌ها برای ارائه‌ی این معادله، هر ترکیب آلی را به‌صورت یک مخلوط فرضی از گروه‌های تشکیل دهنده‌اش در نظر گرفتند که تصادفی در مخلوط توزیع شده‌اند. سپس طبق تقریب تک سیال وان‌دروالس، معادله‌ی حالت LIR را عیناً برای چنین مخلوط فرضی به کار گرفتند اما در این حالت پارامترهای معادله‌ی حالت LIR علاوه بر دما به ترکیب گروه‌ها در سیستم (در اینجا طول زنجیر) نیز وابسته شد. طبق روش آن‌ها، اگر در دمای T و فشار p ، دانسیته‌ی مولی سیال ρ باشد، دانسیته‌ی کل گروه‌ها در مخلوط $n\rho$ می‌شود که n تعداد گروه‌های عاملی در مولکول است. در این صورت معادله‌ی LIR به شکل زیر در می‌آید:

$$\left(\frac{p}{n\rho RT} - 1\right) / n^2 \rho^2 = A_m + B_m n^2 \rho^2 \Rightarrow \left(\frac{Z}{n} - 1\right) / \rho^2 = n^2 A_m + n^4 B_m \rho^2 \quad (۸-۱)$$

که در آن، A_m و B_m پارامترهای LIR به ازای یک گروه در سیال موردنظر هستند. سپس آن‌ها با تعریف دو پارامتر A' و B' به صورت معادله‌های (۹-۱) و (۱۰-۱)،

$$A' = A_m n^2 \quad (۹-۱)$$

$$B' = B_m n^4 \quad (۱۰-۱)$$

معادله‌ی (۸-۱) را به صورت کلی زیر تعریف نمودند و آن را قاعده‌ی همدمای خطی اصلاح شده (MLIR) نامیدند.

$$\left(\frac{Z}{n} - 1\right) v^2 = A' + B' \rho^2 \quad (۱۱-۱)$$

طبق این معادله نمودار $\left(\frac{Z}{n} - 1\right) v^2$ برحسب ρ^2 برای هر همدمای ترکیب آلی چگال، خطی است.

برای تمامی ترکیبات آلی مورد مطالعه شامل آلکان‌های خطی، الکل‌های نوع اول، دوم و سوم، کتون‌ها، اسیدهای کربوکسیلیک، آمین‌ها و استرها [۴-۷] مشخص شده است که رفتار خطی معادله‌ی MLIR

نسبت به معادله‌ی LIR بسیار بهتر بوده و حتی این رفتار خطی با افزایش طول زنجیر حفظ می‌شود. این در حالیست که برای معادله‌ی LIR چنین وضعی مشاهده نمی‌شود. سپس این معادله با استفاده از روش سهم‌گروه به مخلوط ترکیبات آلی تعمیم داده شد و مشخص شد متوسط درصد خطای نسبی در پیش‌بینی دانسیته با استفاده از معادله‌ی حالت MLIR برای مخلوط ترکیبات آلی مختلف در محدوده‌ی وسیعی از دما و فشار کمتر از ۱/۱٪ بوده است.

۲-۳-۱- پیش‌بینی دانسیته مخلوط‌های آلی با استفاده از معادله حالت GCM-GMA

در سال ۲۰۰۵ گوه‌رشادی و همکارانش بر اساس مفهوم متوسط انرژی پتانسیل جفت موثر، توانستند معادله حالت عمومی GMA را برای پیش‌بینی خواص ترمودینامیکی سیالات چگال بدست آورند [۸]. این معادله حالت بیان می‌کند که، نمودار $(2Z-1)v_m^2$ برحسب ρ برای هر همدمای سیال چگال خطی است که $\rho = 1/v_m$ دانسیته‌ی مولی سیال است.

$$(2Z-1)v_m^2 = A + B\rho \quad (1-12)$$

در این رابطه Z و v_m به ترتیب فاکتور تراکم‌پذیری و حجم مولی بوده و A و B به ترتیب عرض از مبدا و شیب نمودار هستند که از طریق روابط زیر به دما وابسته اند:

$$A = A_0 - \frac{2A_1}{RT} + \frac{2A_2 \ln T}{R} \quad (1-13)$$

$$B = B_0 - \frac{2B_1}{RT} + \frac{2B_2 \ln T}{R} \quad (1-14)$$

در روابط (۱-۱۳) و (۱-۱۴)، کمیت‌های A_0 تا A_2 و B_0 تا B_2 ثابت هستند. این معادله‌ی حالت می‌تواند در محدوده‌ی $\rho > \rho_c$ و $T < T_c$ دانسیته‌ی طیف وسیعی از سیالات قطبی، غیرقطبی و سیالات دارای پیوند هیدروژنی را پیش‌بینی کند که، ρ_c و T_c به ترتیب دانسیته و دمای بحرانی سیال هستند [۹-۱۲]. همچنین این معادله حالت با استفاده از تقریب تک سیال واندروالس به مخلوط سیالات چگال نیز تعمیم داده شد [۱۴ و ۱۵].

موسوی در سال ۲۰۱۱ توانست از ترکیب روش سهم گروه با معادله حالت GMA که توسط گوهرشادی و همکارانش ارائه شده بود، دانسیته‌ی ترکیبات آلی مختلف و مخلوط‌های آن‌ها را پیش‌بینی کند [۱۶].

او نیز در این کار، هریک از این ترکیبات را به عنوان یک مخلوط فرضی از گروه‌های متیل، متیلن و یک گروه عاملی تشکیل‌دهنده‌ی آن‌ها در نظر گرفت و سپس معادله‌ی حالت GMA را به صورت زیر به این مخلوط فرضی تعمیم داد.

$$\left(\frac{2Z}{n} - 1\right)v_m^3 = A_m n^3 + B_m n^4 \rho \quad (15-1)$$

که در این معادله، n تعداد گروه‌های کربنی و A_m و B_m پارامترهای معادله‌ی حالت به ازای یک گروه کربنی هستند. او توانست با استفاده از این معادله‌ی جدید، دانسیته‌ی ترکیبات آلی خالص و مخلوط‌های آن‌ها را به ترتیب با متوسط درصد خطای نسبی کمتر از ۱/۲٪ و ۱/۷٪ پیش‌بینی کند. این معادله‌ی جدید GCM-GMA نامیده شد.

۴-۱- هدف تحقیق

در این پایان‌نامه هدف پیش‌بینی دانسیته‌ی ۹۰ مخلوط مختلف از ترکیبات آلی، بر حسب دما و فشار و در کسر مولی‌های متفاوت با استفاده از روش مدل‌سازی ارتباط کمی ساختار- خاصیت است.

فصل دوم

کمومتریکس

۱-۲- کمومتریکس

نام کمومتریکس اولین بار توسط دانشمند سوئدی جوان، اسوانت ولد^۱ در اوایل سال ۱۹۷۰ معرفی شد. همکاری او با کووالسکی^۲ که در آن زمان روی روش‌های تشخیص الگو در شیمی کار می‌کرد منتهی به پایه‌گذاری انجمن بین‌المللی کمومتریکس^۳ (ICS) در سال ۱۹۷۴ شد. چندین تعریف برای کمومتریکس بیان شده‌است که یکی از جامع‌ترین تعاریف به صورت زیر است:

" کمومتریکس یک زمینه از شیمی است که از ریاضی و اصول آمار و منطق استفاده می‌کند برای اینکه (الف) فرایندهای بهینه را طراحی و انتخاب کند، (ب) ماکزیمم اطلاعات شیمیایی مناسب را به وسیله آنالیز اطلاعات شیمیایی فراهم کند و (ج) اطلاعات بیشتری درمورد سیستم‌های شیمیایی در اختیار قرار دهد [۱۷]."

ظهور زمینه‌ی تحقیقاتی کمومتریکس در ایران به سال ۱۹۹۷ برمی‌گردد. اولین تحقیق درباره کمومتریکس در ایران در سال ۱۹۹۸ منتشر شد [۱۸].

۲-۲- ارتباط کمی ساختار- خاصیت و ساختار- فعالیت

۱-۲-۲- رابطه کمی ساختار- فعالیت^۴ (QSAR)

یکی از کاربردهای خاص روش‌های کمومتریکس در رابطه با توسعه‌ی ارتباط کمی ساختار-فعالیت (QSAR) ترکیبات دارویی و ارزیابی داده‌های تجزیه‌ای است. QSAR شامل تمام روش‌های آماری است که فعالیت‌های بیولوژیکی را به ویژگی‌های ساختاری ترکیب مربوط می‌کند. به عبارت دیگر QSAR سعی در پیدا کردن رابطه‌ای هماهنگ میان فعالیت بیولوژیکی و ویژگی‌های مولکولی دارد تا بتواند از این قواعد برای ارزیابی فعالیت ترکیبات جدید استفاده کند.

۱-Svante Wold

۲-Kowalski

۳-International Chemometrics Society

۴-Quantitative Structure-Activity Relationship

به طور کلی QSAR در پی پاسخ دادن به دو سوال زیر است:

(الف) چه جنبه‌هایی از مولکول روی فعالیت آن تاثیر می‌گذارد؟

(ب) برای افزایش آن فعالیت چه عواملی باید بهینه شوند؟

دلیل استفاده از کلمه کمی^۱ مدل ریاضی است که برای بررسی فعالیت مورد استفاده قرار می‌گیرد [۱۹].

۲-۲-۲- رابطه کمی ساختار - خاصیت^۲ (QSPR)

یکی دیگر از زمینه‌های مهم کاربرد کمومتریکس، مطالعاتی است که خواص فیزیکی و شیمیایی مولکول‌ها را به ویژگی‌های ساختاری آن‌ها نسبت می‌دهد. از نظر شیمیدانان خاصیت یک مولکول ناشی از ویژگی‌های ساختاری آن‌هاست. این نوع از مطالعات به بررسی کمی ارتباط ساختار - خاصیت (QSPR) معروف می‌باشد. هدف از این مطالعات پیدا کردن رابطه‌ای است که بین رفتار فیزیکوشیمیایی یک مولکول با پارامترهای ساختاری آن وجود دارد. مانند یافتن رابطه‌ای بین ساختار مولکول‌ها با خواصی نظیر نقطه‌ی ذوب و جوش، فشار بخار و غیره. نتایج این مطالعات علاوه بر شفاف‌سازی نحوه‌ی ارتباط خواص مولکول‌ها با ویژگی‌های ساختمانی آن‌ها، به پژوهشگران در پیش‌بینی رفتار مولکول‌های جدید بر اساس رفتار مولکول‌های مشابه کمک می‌کند.

از روش‌هایی که به منظور مطالعه‌ی ارتباط خطی ساختار - خاصیت مورد استفاده قرار می‌گیرند می‌توان به روش‌های خطی مثل حداقل مربعات متداول^۳ (OLS)، تحلیل مؤلفه‌های اساسی^۴ (PCA)، رگرسیون خطی چندگانه^۵ (MLR)، رگرسیون اجزای اصلی^۶ (PCR) و حداقل مربعات

۱-Quantitative

۲-Quantitative Structure-Property Relationship

۳-Ordinary Least Square

۴-Principle Component Analysis

۵-Multiple Linear Regression

۶-Principal Componet Regression

جزئی^۱ (PLS)، اشاره نمود. روش‌های دیگری مانند شبکه‌ی عصبی مصنوعی^۲ (ANN) و ماشین بردار پشتیبان^۳ (SVM)، ارتباط غیرخطی میان ساختار و خواص ترکیبات را مورد مطالعه قرار می‌دهند. مراحل کلی یک مدل‌سازی QSPR یا QSAR به صورت زیر است [۲۰].

۱- فراهم کردن سری داده‌ها

۲- انتخاب و محاسبه توصیف‌کننده‌ها

۳- ارزیابی، تجزیه و تحلیل آماری توصیف‌کننده‌ها

۴- تقسیم‌بندی سری داده‌ها به سری‌های آموزش، ارزیابی و تست

۵- آنالیز آماری مدل‌ها و انتخاب مدل مناسب

۶- ارزیابی اعتبار مدل‌های انتخاب شده

۷- پیش‌بینی خصوصیت موردنظر

در ادامه راجع به هر یک از مراحل یک مدل‌سازی QSPR توضیح داده خواهد شد.

۳-۲- مراحل مدل‌سازی QSPR

۱-۳-۲- فراهم کردن سری داده‌ها

یک مطالعه‌ی QSPR با جمع‌آوری داده‌ها برای خاصیت موردنظر شروع می‌شود. هرچه سری داده‌ها وسیع‌تر و متنوع‌تر باشد مدل بدست آمده مدل کلی‌تر و معتبرتری بوده و محدود به یک گروه خاص از ترکیبات نخواهد بود. داده‌های جمع‌آوری شده می‌توانند در آزمایشگاه تولید و یا از طریق مقادیر منتشر شده در مقالات استخراج شوند. البته ترکیب‌های موجود در سری داده‌ها باید مشابهت ساختاری داشته و کمیت موردنظر برای آن‌ها در شرایط یکسان تعیین شده باشند.

۱-Partial Least Squares

۲-Artificial Neural Network

۳-Support Vector Machine

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

در این مرحله، ساختار ترکیبات مربوط به سری داده‌ها با استفاده از نرم‌افزارهای تجاری در دسترس رسم شده و سپس توسط روش‌های کوانتوم - مکانیکی مختلف نظیر PM3, AM1, MNDO و..... بهینه می‌شود [۲۱]. نرم افزاری که برای این منظور معمولاً مورد استفاده قرار می‌گیرد نرم افزار Hyper Chem است.

۲-۳-۳- محاسبه‌ی توصیف‌کننده‌های تئوری

توصیف‌کننده‌ها مقادیر عددی هستند که ویژگی‌های مختلف ساختاری و الکترونی مولکول‌ها را به طور کمی نشان می‌دهند. به عبارت دیگر هر توصیف‌کننده بیانگر خصوصیت ویژه‌ای از مولکول است که ممکن است بر خاصیت موردنظر موثر باشد [۲۲].

برخی از ویژگی‌های یک توصیف‌کننده عبارت است از:

- ساده بودن
- توانایی تفسیر ساختار مولکول
- عدم همبستگی با سایر توصیف‌کننده‌ها
- قابلیت تمایز بین ایزومرهای مختلف مولکول
- قابل کاربرد برای دامنه وسیعی از ساختارهای مولکول

توصیف‌کننده‌های مولکولی را به روش‌های مختلفی طبقه‌بندی می‌کنند. ساده‌ترین طبقه‌بندی براساس ماهیت (تجربی یا تئوری) است. توصیف‌کننده‌های تجربی براساس اندازه‌گیری‌های تجربی به دست می‌آیند و توصیف‌کننده‌های تئوری با استفاده از ساختمان مولکول و بدون نیاز به داده‌های تجربی محاسبه می‌شوند که باعث صرفه جویی در وقت، هزینه، مصرف مواد و تجهیزات خواهند شد و برای هر مولکول واقعی و یا فرضی که هنوز سنتز نشده‌اند نیز در دسترس می‌باشند [۲۳]. در این

تحقیق ۱۴۸۱ توصیف‌کننده‌ی ساختاری توسط نرم‌افزار دراگون^۱ محاسبه شد که این توصیف‌کننده‌ها از نظر بیان چگونگی خصوصیات مولکول به ۱۸ دسته‌ی متفاوت تقسیم می‌شوند، که در جدول (۱-۲) نام این توصیف‌کننده‌ها آورده شده است [۱].

جدول (۱-۲) - انواع توصیف‌کننده‌های محاسبه شده توسط نرم افزار Dragon

۱	توصیف‌کننده‌های زیرساختاری ^۲	۲	توصیف‌کننده‌های توپولوژی ^۳
۳	شمارنده‌ی مولکولی ^۴	۴	توصیف‌کننده‌های BCUT ^۵
۵	شاخص بار توپولوژیکی ^۶ GALVEZ	۶	خود ارتباطی‌های دو بعدی ^۷
۷	توصیف‌کننده‌ای بار ^۸	۸	شاخص‌های آروماتیک ^۹
۹	پروفایل‌های مولکولی راندیک ^{۱۰}	۱۰	توصیف‌کننده‌های هندسی ^{۱۱}
۱۱	توصیف‌کننده‌های RDF ^{۱۲}	۱۲	توصیف‌کننده‌های سه بعدی ^{۱۳}
۱۳	توصیف‌کننده‌های WHIM ^{۱۴}	۱۴	توصیف‌کننده‌های GATAWAY ^{۱۵}
۱۵	گروه‌های عاملی ^{۱۶}	۱۶	اجزای اتم مرکزی ^{۱۷}
۱۷	توصیف‌کننده‌های تجربی ^{۱۸}	۱۸	خواص مولکولی ^{۱۹}

-
- ۱-Dragon
 - ۲-Constitutional Descriptors
 - ۳-Topological Descriptors
 - ۴-Molecular Walk Counts
 - ۵-BCUT Descriptors
 - ۶-Galvez topol Charge Indices
 - ۷-2D Autocorrelations
 - ۸-Charge Descriptors
 - ۹-Aromatic Indices
 - ۱۰-Randic Molecular Profiles
 - ۱۱-Geometrical Descriptors
 - ۱۲-Radial Distribution Function Descriptors
 - ۱۳-3D-MORSE Descriptors
 - ۱۴-Weighted Holistic Invariant Molecular Descriptors
 - ۱۵-Geometry, Topology and Atom Weighted Assembly
 - ۱۶-Functional Groups
 - ۱۷-Atom- Central Fragments
 - ۱۸-Empirical Descriptors
 - ۱۹-Molecular Properties

۴-۳-۲- حذف توصیف‌کننده‌های نامناسب

پس از آن که توصیف‌کننده‌ها محاسبه شدند، باید توصیف‌کننده‌هایی که اطلاعات کمتری در مورد تغییرات خواص شیمیایی و یا بیولوژیکی موردنظر دارند را حذف نمود. از این رو توصیف‌کننده‌هایی که برای تمام مولکول‌ها، مقادیر ثابت، تقریباً ثابت و یا صفر داشته باشند، حذف می‌شوند. همچنین توصیف‌کننده‌هایی که کمتر از ۱۰٪ مقادیر غیر صفر و یا بیش از ۹۰٪ مقادیر یکسان دارند، مفید نبوده و کنار گذاشته می‌شود. برای حذف بقیه توصیف‌کننده‌های غیرمفید، ماتریس همبستگی از طریق نرم‌افزار Matlab بین توصیف‌کننده‌ها تشکیل می‌شود. همبستگی بیش از ۹۰٪ بین توصیف‌کننده‌ها بیانگر رابطه‌ی خطی بین آنهاست. در این صورت توصیف‌کننده‌ای که ارتباطش با متغیر وابسته مشکل‌تر بوده و ضریب همبستگی‌اش کمتر باشد، حذف می‌شود.

۵-۳-۲- تقسیم‌بندی داده‌ها

قبل از ساختن مدل، سری داده‌ها به سه زیر مجموعه تقسیم می‌شوند. اولین زیر مجموعه سری آموزش^۱ می‌باشد که اکثریت داده‌ها را در بر گرفته و از آن برای ساخت مدل‌های خطی و غیرخطی استفاده می‌شود. دومین زیر مجموعه، سری ارزیابی^۲ است که مدل‌های به دست آمده از سری آموزش توسط آن ارزیابی شده و به این ترتیب توصیف‌کننده‌های بهینه برای ساخت مدل خطی و غیرخطی مشخص می‌شوند. زیر مجموعه‌ی سوم سری تست^۳ یا آزمایش است که در طول فرایند مدل‌سازی هیچ‌گونه دخالتی ندارند. برای تقسیم داده‌ها به این سه مجموعه می‌توان از روش تقسیم تصادفی و یا روش تحلیل مولفه‌های اساسی (PCA) استفاده نمود.

۱-train

۲-validation

۳-test

۶-۳-۲- 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. هرچه در مدل نهایی تعداد توصیف‌کننده‌ها کمتر باشند، مدل ساده‌تر خواهد شد. بنابراین باید به ترتیبی، بهترین توصیف‌کننده‌ها را از میان توصیف‌کننده‌های باقی‌مانده، انتخاب نمود. دو روش مرسوم برای انتخاب توصیف‌کننده‌ها، روش خطی رگرسیون مرحله‌ای (MLR) و روش غیرخطی الگوریتم ژنتیک^۱ است. چون در این پایان‌نامه از روش GA برای انتخاب بهترین توصیف‌کننده‌ها استفاده شده در ادامه به اختصار راجع به این روش توضیحاتی ارائه می‌گردد.

۱-۶-۳-۲- انتخاب بهترین توصیف‌کننده‌ها توسط الگوریتم ژنتیک

۱-۶-۳-۲-الف- مقدمه

الگوریتم ژنتیک یک روش بهینه‌سازی الهام گرفته از طبیعت جاندار است که می‌توان در طبقه‌بندی‌ها، آن را به عنوان یک روش عددی، جستجوی مستقیم و تصادفی معرفی کرد. این الگوریتم، الگوریتمی مبتنی بر تکرار بوده و اصول اولیه‌ی آن از علم ژنتیک اقتباس شده است و با تقلید از تعدادی از فرایندهای مشاهده شده در تکامل طبیعی ارائه شده است. این روش به طور موثری از طبیعت موجود در یک جمعیت استفاده می‌کند، تا حل‌های جدید و بهبود یافته را ایجاد کند. این الگوریتم در مسایل متنوعی نظیر بهینه‌سازی، شناسایی و کنترل سیستم، پردازش تصویر و مسایل ترکیبی، تعیین توپولوژی و آموزش شبکه‌های عصبی مصنوعی و سیستم‌های مبتنی بر تصمیم و قاعده به کار می‌رود.

۱-۶-۳-۲-ب- الگوریتم ژنتیک

الگوریتم ژنتیک (GA) یک تکنیک جستجو در علم رایانه برای یافتن راه‌حل بهینه و مسائل جستجو

^۱-Genetic Algorithm

می‌باشد. الگوریتم‌های ژنتیک یکی از انواع الگوریتم‌های تکاملی هستند که از علم زیست‌شناسی مانند وراثت، جهش، انتخاب ناگهانی، انتخاب طبیعی و ترکیب الهام گرفته‌اند. ساز و کار الگوریتم ژنتیک بر مبنای قوانین موجود در علم ژنتیک است. در ادامه با برخی مفاهیم و تعاریف الگوریتم ژنتیک آشنا خواهیم شد.

ژن : به منظور نمایش جواب یک مسئله که قرار است با الگوریتم ژنتیک پاسخ آن پیدا شود، لازم است تا یک ساختار داده‌ای^۱ مناسب انتخاب گردد. به این ساختار داده‌ای اصطلاحاً ژن یا شبه کروموزوم^۲ گفته می‌شود. هر ژن مبین یک متغیر یا یک پارامتر است که تمامی اطلاعات لازم در مورد یک جواب را در خود دارد [۲۴].

افراد^۳ یا کروموزوم^۴ : در الگوریتم ژنتیک فرض می‌شود که راه‌حل یک مسئله را می‌توان بصورت مجموعه‌ای از پارامترهای مطلوب نشان داد. از اتصال این پارامترها که آن‌ها را بنام ژن می‌شناسیم زنجیره‌ای از مقادیر به‌وجود می‌آید که تحت عنوان کروموزوم شناخته می‌شود [۲۵].

جمعیت : در الگوریتم ژنتیک به مجموعه‌ای از کروموزوم‌ها که در کنار هم قرار می‌گیرند جمعیت گفته می‌شود. وظیفه‌ی الگوریتم ژنتیک انجام پردازش بر روی این جمعیت به منظور بهبود دادن آن‌هاست. در هر بار اجرای الگوریتم ژنتیک بر روی یک جمعیت، جمعیت جدیدی به‌دست می‌آید که به آن نسل بعدی می‌گویند [۲۵].

نمایش و کدگذاری کروموزوم‌ها : تاکنون در حل مسائل بهینه‌سازی روش‌های مختلفی برای نمایش پارامترها و اطلاعات مسأله (کروموزوم‌ها) به کار برده شده است که انتخاب هر کدام از این روش‌ها باید با توجه به نوع مسأله و فضای جستجوی مورد نیاز برای حل و بهینه‌سازی آن صورت پذیرد. از پرکاربردترین روش‌ها، کدگذاری رشته‌ای است.

۱-Data structure

۲-Chromosome like

۳-Individuals

۴-Chromosome

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

تعیین جمعیت اولیه: بعد از تصمیم‌گیری در مورد شیوه‌ی کدگذاری کروموزوم‌ها، جمعیت اولیه باید اتخاذ گردد. این مرحله معمولاً با انتخاب تصادفی مقادیر در محدوده‌ی مجاز صورت می‌گیرد.

روش‌های انتخاب کروموزوم‌ها: روش‌های مختلفی برای انتخاب کروموزوم‌ها وجود دارند که در زیر به معمولترین آن‌ها اشاره می‌شود.

- انتخاب چرخ رولت^۲: در این روش هر عنصری که عدد برازش (تناسب) بیشتری داشته باشد، شانس انتخاب بیشتری دارد.
- انتخاب نخبه‌گرا یا ممتازگرا^۳: در این روش مناسب‌ترین عضو هر اجتماع انتخاب می‌شود.
- انتخاب مقیاس^۴: در این روش به موازات افزایش متوسط عدد برازش جامعه، سنگینی انتخاب هم بیشتر می‌شود. این روش وقتی به کار می‌رود که مجموعه دارای عناصری با عدد برازش بزرگ باشد و فقط تفاوت‌های کوچکی آن‌ها را از هم تفکیک کند.
- انتخاب مسابقه‌ای^۵: روشی است که در آن یک زیرمجموعه از صفات یک جامعه، انتخاب می‌شوند و اعضای آن مجموعه با هم رقابت می‌کنند تا سرانجام فقط یک صفت از هر زیر گروه برای ادغام، انتخاب شوند [۲۶].

۲-۳-۶-۱-ج- نحوه‌ی انتخاب توصیف‌کننده‌ها و شرایط توقف الگوریتم ژنتیک

الگوریتم‌های ژنتیک ذاتاً موازی هستند و به دلیل موازی بودن می‌توانند چندین رشته را در یک لحظه مورد ارزیابی قرار دهند. الگوریتم‌هایی که موازی نیستند فقط می‌توانند فضای مسأله مور نظر

۱-Fitness function

۲-Roulette wheel

۳-Elitist

۴-Scalin

۵-Tournament

را در یک جهت در یک لحظه جستجو کنند و اگر راه حل پیدا شده یک جواب بهینه‌ی محلی باشد و یا زیر مجموعه‌ای از جواب اصلی باشد باید تمام کارهایی که تا به حال انجام شده را کنار گذاشت و دوباره از اول شروع کرد. از آنجایی که الگوریتم ژنتیک چندین نقطه‌ی شروع موازی دارد، در یک لحظه می‌تواند فضای مسأله را از چند جهت مختلف جستجو کند. اگر یکی به نتیجه نرسید سایر راه‌ها ادامه می‌یابند و منابع بیشتری را در اختیار قرار می‌دهند. الگوریتم ژنتیک تغییرات تصادفی در راه‌حل‌های کاندید شده را نشان می‌دهد سپس از تابع برازش برای سنجش این که آیا آن تغییرات تصادفی پیشرفتی در انتخاب ژن مناسب ایجاد کرده‌اند یا نه استفاده می‌کند.

از شرایط زیر می‌توان برای توقف روند تکرار الگوریتم ژنتیک استفاده نمود:

- تعداد نسل‌ها: الگوریتم زمانی متوقف می‌شود که تعداد نسل‌ها به مقدار معینی برسد.
- محدودیت زمانی: در جعبه ابزار GA می‌توان این محدودیت را برحسب واحد ثانیه تعیین نمود.
- محدودیت شایستگی: الگوریتم زمانی متوقف می‌شود که بهترین مقدار شایستگی در جمعیت حاضر کوچکتر یا مساوی یک مقدار معین گردد.
- رکود نسلی: اگر در نسل‌ها به تعداد معینی هیچ‌گونه پیشرفتی حاصل نشود، الگوریتم متوقف می‌گردد.
- رکود زمانی: در صورتی که در تابع هدف برای زمان معینی هیچ‌گونه پیشرفتی حاصل نشود، الگوریتم متوقف می‌گردد.

الگوریتم در صورتی که هر یک از شرایط فوق برآورده گردد متوقف می‌شود.

۷-۳-۲- انتخاب توصیف‌کننده‌ها بر مبنای روش سهم گروه^۱ (GCM)

روش دیگر انتخاب توصیف‌کننده‌ها، روش سهم گروه است. 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

^۱-Group Contribution Method

می‌رود. این روش در واقع ارتباطی بین خاصیت یک ماده با ساختار شیمیایی آن برقرار می‌کند. مبنای به وجود آمدن روش سهم‌گروه‌ها این است که اگرچه 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 سهم‌گروه‌ها فرض می‌کند که خاصیت فیزیکی یا ترمودینامیکی هر ماده را می‌توان از سهم‌گروه‌های 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، سهم‌گروه‌های 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 سهم‌گروه‌ها 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 [۲۷].

در مطالعات QSPR که به طور گسترده‌ای برای مدل‌سازی خواص فیزیکی و شیمیایی مواد به کار می‌رود، خاصیت موردنظر 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 سهم‌گروه‌ها 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 سهم‌گروه‌ها برای پیش‌بینی خواص مختلفی مثل نقطه‌ی اشتعال آلکان‌ها [۲۸]، دانسیته‌ی مایعات یونی [۲۹]، خصوصیات نقطه‌ی بحرانی و فاکتور شکل هیدروکربن‌ها [۳۰]، هدایت یونی و ویسکوزیته‌ی مایعات یونی [۳۱]، استفاده شده است.

۸-۳-۲- ساختن مدل

پس از انتخاب بهترین توصیف‌کننده‌ها، نوبت به ساخت مدل می‌رسد. مدل، یک رابطه‌ی ریاضی است که بیان‌کننده‌ی ارتباط بین خاصیت موردنظر به عنوان متغیر وابسته و توصیف‌کننده‌ها به عنوان متغیر مستقل است. با کمک مدل ساخته شده می‌توان با داشتن مقادیر متغیرهای مستقل، متغیر وابسته را برای ترکیبات جدید پیش‌بینی کرد.

برای مدل‌سازی روش‌های خطی و غیرخطی مختلفی وجود دارد که رگرسیون خطی چندگانه جزء روش‌های خطی بوده و روش‌های شبکه‌ی عصبی مصنوعی (ANN) و ماشین بردار پشتیبان (SVM) از جمله روش‌های غیرخطی هستند. در این پایان‌نامه از روش غیرخطی ANN برای مدل‌سازی دانسیته‌ی مخلوط‌های ترکیبات آلی استفاده شده است. بنابراین این روش در ادامه به اختصار شرح داده خواهد شد.

۱-۸-۳-۲- شبکه‌ی عصبی

تحقیقات درمورد شبکه‌های عصبی مصنوعی از زمانی شروع شد که نگرش نوینی درمورد نحوه‌ی عملکرد مغز انسان به وجود آمد. در این نگرش، مغز به عنوان سیستم پردازش اطلاعات با ساختار موازی از تعداد زیادی واحد محاسباتی به نام نرون^۱ تشکیل شده که هر کدام به هزاران نرون دیگر متصل هستند. اگرچه همه‌ی نرون‌ها کارکرد یکسانی دارند اما اندازه و شکل آن‌ها به محل استقرارشان در سیستم عصبی بستگی دارد. اما با وجود این تنوع ساختار، نرون‌ها از سه بخش اساسی تشکیل شده‌اند.

۱-بدنه‌ی سلول یا جسم سلولی

۲-دندريت^۲

۳-آکسون^۳

دندريت و آکسون، عناصر ارتباطی نرون را تشکیل می‌دهند. دندريت‌ها به عنوان مناطق دریافت سیگنال‌های الکتریکی، شبکه‌هایی متشکل از فیبرهای سلولی هستند که دارای سطحی نامنظم بوده و به همین علت به آن‌ها شبکه‌های دریافتی درختی نیز گفته می‌شود. دندريت‌ها گیرنده‌های نرون بوده و وظیفه‌ی آن‌ها انتقال سیگنال‌های الکتریکی به هسته‌ی سلول است. بدنه‌ی سلول انرژی لازم برای فعالیت نرون را فراهم آورده و بر روی سیگنال‌های دریافتی عمل می‌کند. عملیاتی که در این‌جا صورت می‌گیرد در واقع انجام یک عمل جمع ساده و مقایسه‌ی آن با یک سطح آستانه است. آکسون برخلاف دندريت از سطح صاف‌تری برخوردار بوده و شاخه‌های کمتری دارد. وظیفه‌ی آکسون دریافت سیگنال الکتروشیمیایی از هسته‌ی سلول و انتقال آن به نرون‌های دیگر است. محل تلاقی یک آکسون از یک سلول و دندريت‌های سلول دیگر سیناپس^۴ نامیده می‌شود. سیناپس‌ها واحدهای ساختاری کوچکی هستند که ارتباط بین نرون‌ها را برقرار می‌کنند. پالس‌هایی که بین نرون‌ها انتقال می‌یابند از نوع

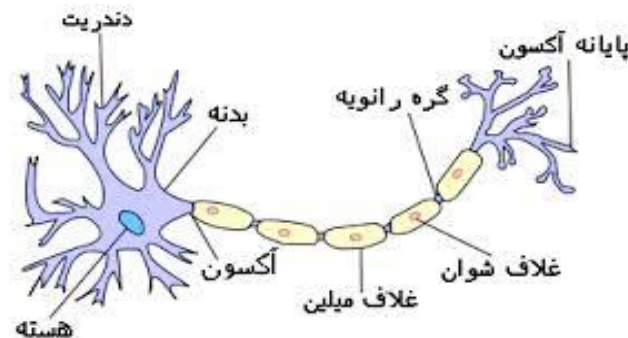
۱-Neuron

۲-Dendrite

۳-Axon

۴-Synapse

الکتروشیمیایی هستند. لازم به ذکر است که جهت حرکت پیام عصبی در طول تار عصبی یک طرفه بوده و فقط از دندریت به آکسون است [۳۲]. شکل (۱-۲) ساختار کلی یک نرون زیستی را نشان می‌دهد.



شکل (۱-۲) - ساختار یک نرون زیستی

شبکه‌ی عصبی مصنوعی یک برنامه نرم‌افزاری است که با روشی مشابه با مغز، اطلاعات را پردازش می‌کند. 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 [۳۳].

۱-۸-۳-۲-الف - ساختار شبکه عصبی مصنوعی

مهم‌ترین جزء تشکیل دهنده‌ی هر شبکه‌ی عصبی مصنوعی نرون است. یک نرون محاسباتی کوچک‌ترین واحد پردازش اطلاعات است که در اصل مدل ساده شده‌ای از نرون طبیعی است. عموماً یک نرون بیش از یک ورودی دارد. کار هر نرون این است که ورودی‌ها را از نرون‌های مجاور خود بگیرد آن‌ها را وزن‌دهی کرده و سپس مجموع آن‌ها را محاسبه کند تا سیگنال تحریک بدست بیاید. بعد با اعمال تابع محرک روی سیگنال تحریک، خروجی را محاسبه کرده و به نرون‌هایی که با آن‌ها

^۱-Intelligent

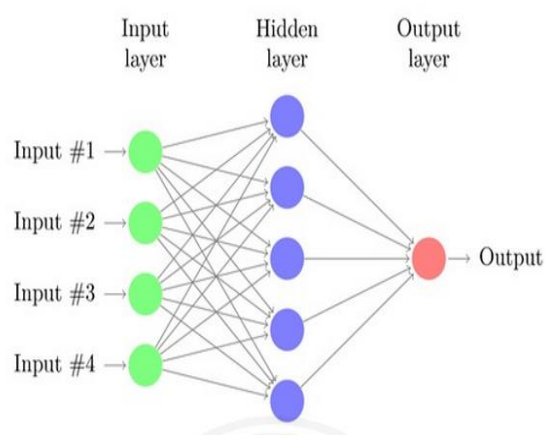
ارتباط دارد بفرستد. تابع محرک می‌تواند خطی و یا غیرخطی باشد، در واقع طراح با توجه به نوع مساله، تابع محرک را انتخاب می‌کند. تعداد توابع محرکی زیادی در ساختار شبکه‌ی عصبی می‌توانند مورد استفاده قرار گیرند که تابع محرک خطی، تابع محرک پله‌ای و تابع محرک سیگموئیدی از معروف‌ترین این توابع می‌باشند.

در سیستم‌های پردازش موازی سه نوع واحد پردازشگر قابل تفکیک وجود دارد که عبارتند از: واحدهای ورودی، واحدهای خروجی و واحدهای میانی یا پنهان.

واحدهای ورودی، ورودی‌ها را از منبع داده‌های ورودی دریافت می‌کنند و در اختیار سایر نرون‌ها قرار می‌دهند.

واحدهای خروجی، مقادیر محاسبه شده را به خارج از سیستم می‌فرستند. البته ممکن است برای گام‌های بعدی پردازش این خروجی‌ها به سیستم پس‌خور شوند.

واحدهای میانی یا پنهان، به واحدهایی اطلاق می‌شوند که با خارج از سیستم ارتباط مستقیمی ندارند. به عبارت دیگر ورودی و خروجی‌های آن‌ها در داخل سیستم قرار دارند [۳۲]. واحدهای یک شبکه‌ی عصبی مصنوعی در شکل (۲-۲) نشان داده شده است.



شکل (۲-۲) - ساختار یک شبکه عصبی مصنوعی

۱-۸-۳-۲-ب- 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 برنمی‌گردد. 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 شبکه می‌رسند. در نتیجه جریان اطلاعات در این نوع از شبکه‌ها همیشه در یک مسیر یک‌طرفه 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 [۲۷]. 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.

۱-۸-۳-۲-ج- روش‌های آموزش

همان‌طور که قبلاً اشاره شد شبکه‌های عصبی مصنوعی می‌توانند همانند مغز انسان آموزش ببینند و یاد بگیرند. به طور کلی دو نوع یادگیری برای شبکه‌ها وجود دارد: آموزش باناظر^۱ و آموزش بدون ناظر^۲.

۱-Feedforward

۲-Feedback

در یادگیری با ناظر، در هر مرحله از تکرار الگوریتم یادگیری، جواب واقعی سیستم یادگیرنده وجود دارد. لذا الگوریتم یادگیری به خطای یادگیری؛ یعنی تفاوت بین مقدار واقعی و مقدار پیش‌بینی شده، دسترسی دارد. در اکثر شبکه‌های عصبی از این نوع یادگیری استفاده می‌شود.

نوع دیگری از یادگیری به نام یادگیری بدون ناظر یا خودسامان‌ده می‌باشد که در آن جواب واقعی برای سیستم یادگیرنده موجود نیست و شبکه می‌آموزد که الگوهای ورودی را به تعداد مساوی از گروه‌ها تقسیم‌بندی کرده، ارتباطات موجود بین آن‌ها را پیدا و در خروجی شبکه کد نماید. باید توجه داشت که در این حالت، فرد کاربر است که هدف نهایی را مشخص می‌کند [۲۷].

برای حداقل شدن خطای بین مقادیر تجربی و پاسخ شبکه باید عملکرد شبکه را بهینه نمود. برای بهینه‌سازی عملکرد شبکه، ابتدا باید یک شاخص عملکرد معرفی شود. شاخص عملکرد، معیاری برای بیان عملکرد شبکه است و با عملکرد شبکه رابطه‌ی عکس دارد. یعنی هر چه عملکرد شبکه بهتر باشد، مقدار شاخص عملکرد کوچک‌تر خواهد بود و بالعکس. شاخص عملکرد در بیشتر الگوریتم‌ها، متوسط مربعات خطا (MSE)^۳ است. شاخص عملکرد را اصطلاحاً تابع هدف یا تابع خطا نیز می‌گویند، یعنی تابعی که کمینه کردن آن موردنظر است. پس از تعیین شاخص عملکرد باید پارامترهای شبکه برای کاهش مقدار این شاخص عملکرد تنظیم گردد [۳۴].

از رایج‌ترین الگوریتم‌های آموزشی، الگوریتم‌های آموزشی لونیگ - مارکوات و تنظیم بایزین است. این الگوریتم‌ها روش مناسبی برای آموزش شبکه‌هایی هستند که شاخص عملکرد آن‌ها حداقل متوسط مربعات خطا است. این روش‌ها به عنوان سریع‌ترین و مؤثرترین روش‌های آموزش شبکه‌های عصبی شناخته شده‌اند. سرعت همگرایی این روش صدها برابر بیشتر از شیوه‌های استاندارد مشابه است. در این تحقیق نیز از این دو الگوریتم برای بهینه‌سازی عملکرد شبکه استفاده شده است [۳۵].

۱-Supervised learning

۲-Unsupervised learning

۳-Mean Square Error

۹-۳-۲- ارزیابی قدرت پیش‌بینی مدل

۹-۳-۲-الف- استفاده از پارامترهای آماری

برای اطمینان از این که مدل به دست آمده توانایی پیش‌بینی نمونه‌های مختلفی از یک جمعیت را داراست، باید مدل را ارزیابی کرد. این ارزیابی با محاسبه‌ی پارامترهای آماری صورت می‌گیرد. روابط ریاضی پارامترهای آماری استفاده شده در این پایان‌نامه، در ادامه توضیح داده خواهد شد.

ضریب تعیین (R^2): به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به کار می‌رود و نشان دهنده‌ی نسبت تغییرات متغیر وابسته نسبت به متغیر مستقل است. به عنوان مثال R^2 برابر ۰/۹۶۶۰ نشان می‌دهد ۹۶/۶۰ درصد تغییرات در متغیر وابسته می‌تواند توسط متغیر مستقل توضیح داده شود. رابطه‌ی ریاضی مربوط به ضریب تعیین به صورت زیر است.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (۲-۱)$$

که در آن SSR طبق رابطه‌ی (۲-۲) بیانگر مجموع مربعات انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است:

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (۲-۲)$$

و SST طبق رابطه‌ی (۳-۲) نشانگر مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (۲-۳)$$

۱-Determination Coefficient

۲-Sum Square Regression

۳-Sum Square Total

که در این روابط $\hat{y}_i$ مقدار پیش‌بینی شده‌ی متغیر وابسته، y_i مقدار واقعی متغیر وابسته و $\bar{y}$ در دو رابطه، میانگین مقادیر متغیر وابسته است. همچنین SSE^1 مبین مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است و طبق رابطه‌ی (۲-۴) تعریف می‌شود.

$$SSE = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (2-4)$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2-5)$$

طبق رابطه‌ی (۲-۵) 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.

سایر پارامترهای آماری مورد استفاده و روابط مربوط به آن‌ها به صورت زیر است.

میانگین مربع خطاها (MSE):

(۲-۶)

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

مجموع مربع باقیمانده‌ها^۲ ($PRESS$):

$$PRESS = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (2-7)$$

^۱-Sum Square Error

^۲-Predictive Residual Sum of Squares

میانگین خطای مطلق^۱ (MAE) :

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}|}{n} \quad (2-8)$$

متوسط مطلق درصد خطای نسبی^۲ (MAPE) :

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right| \quad (2-9)$$

متوسط درصد خطای نسبی^۳ (MPE) :

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{y_i - \hat{y}}{y_i} \quad (2-10)$$

ماکزیمم مطلق درصد خطای نسبی^۴ (D_{Max}) :

$$D_{Max} = \text{Max} \left(100 \left| \frac{y_i - \hat{y}}{y_i} \right| \right) \quad (2-11)$$

در روابط (۲-۲) تا (۲-۱۱) نیز y_i و $\hat{y}_i$ به ترتیب مقدار واقعی و مقدار پیش‌بینی شده‌ی متغیر وابسته و n تعداد داده‌های وارد شده به مدل است.

۹-۳-۲-ب- استفاده از نمودار برگشتی

در نمودار برگشتی مقادیر پیش‌بینی شده‌ی کمیت مورد نظر برحسب مقادیر تجربی رسم می‌شود و به کمک ضریب تعیین (R^2) به دست آمده از نمودار، پراکندگی نقاط اطراف خط برگشت تعیین می‌شود. محدوده‌ی تغییرات ضریب تعیین به دست آمده بین ۰ و ۱ است. هرچه قدر ضریب تعیین به

۱-Mean Absolute Error

۲-Mean Absolute Percentage Error

۳-Mean Percentage Error

۴-Maximum Absolute Percentage Error

۱ نزدیک تر باشد، نتایج پیش‌بینی شده به مقادیر واقعی نزدیک ترند. اما اگر این مقدار برابر صفر باشد، بین داده‌ها هیچ گونه همبستگی وجود ندارد.

۹-۳-۲-ج- استفاده از نمودار خطای باقیمانده

منظور از عبارت خطای باقیمانده، اختلاف بین مقادیر پیش‌بینی شده و مقادیر تجربی است. اگر پراکندگی مقادیر در دو طرف نمودار صفر باشد، این امر نشان دهنده‌ی تصادفی بودن خطاست. ولی اگر عمده‌ی نقاط در این نمودار، در یک طرف خط صفر باشند، این بدان معناست که خطای جهت‌داری رخ داده است.

۹-۳-۲-د- استفاده از سری تست خارجی

اعتبار و اهمیت مدل‌های پیش‌بینی کننده وقتی مشخص می‌شود که بتوانند خاصیت موردنظر را برای مولکول‌های جدیدی که در دسته‌ی آموزش موجود نیستند پیش‌بینی کنند. بدین منظور از ابتدای کار تعدادی از مولکول‌ها به عنوان سری تست خارجی کنار گذاشته می‌شود و در نهایت از مدل منتخب بهینه شده، جهت پیش‌بینی خاصیت موردنظر مولکول‌های موجود در سری تست خارجی که در مدل‌سازی استفاده نشده‌اند، استفاده می‌شود.

فصل سوم

محاسبات

**پیش‌بینی دانسیته مخلوط ترکیبات آلی با
استفاده از روش‌های غیر خطی**

در دو فصل قبل مطالبی راجع به دانسیته‌ی مخلوط‌ها و اهمیت پیش‌بینی آن و همچنین مراحل و روش‌های مدل‌سازی معرفی شدند. در این تحقیق، از روش‌های سهم‌گروه و الگوریتم ژنتیک به عنوان روش‌های انتخاب توصیف‌کننده و از روش شبکه‌ی عصبی مصنوعی جهت مدل‌سازی غیرخطی برای پیش‌بینی دانسیته‌ی مخلوط‌های آلی استفاده گردید. همچنین توانایی مدل‌های به دست آمده با روش‌های مختلف مورد ارزیابی قرار گرفت. آنچه در ادامه‌ی این فصل می‌آید چگونگی مراحل طراحی و بهینه‌سازی مدل‌ها و همچنین شیوه‌ی ارزیابی آن‌هاست.

۱-۳- انتخاب سری داده‌ها

به‌منظور پیش‌بینی دانسیته‌ی مخلوط‌های آلی، ۴۰۸۸ نقطه داده‌ی مربوط به ۸۹ مخلوط دوتایی و یک مخلوط سه‌تایی از ترکیبات آلی مختلف شامل آلکان‌ها، الکل‌ها، اسیدها، استرها، اترها، آمیدها و... در محدوده‌ی وسیعی از دما و فشار جمع‌آوری گردید. نام این مخلوط‌ها به همراه محدوده‌ی دما، فشار و تعداد نقاط مورد استفاده برای هر مخلوط و مرجع مربوط به هریک از آن‌ها در جدول (۱-۳) گزارش شده است.

جدول (۳-۱) - نام مخلوط‌ها، محدوده‌ی دما و فشار و تعداد نقاط مورد استفاده برای هر مخلوط به همراه مرجع مورد استفاده برای هر یک از آن‌ها

شماره مخلوط	جزء A	جزء B	جزء C	$\Delta p/\text{MPa}$	$\Delta T/\text{K}$	تعداد نقاط	Ref
۱	Octane	1-chlorohexane	-	۰/۱ - ۳۹/۶۲	۲۹۸/۳۲۸-۱۵/۱۵	۲۸۸	۳۶
۲	Octane	Dodecane	-	۰/۱ - ۴۱۲/۶	۲۹۸/۳۴۸-۱۵/۱۳	۳۲	۳۷
۳	Octane	1-Chloropentane	-	۲/۱۳ - ۳۸/۶	۲۹۸/۳۲۸-۱۵/۱۵	۱۰۰	۳۸
۴	Octane	1-chlorobutane	-	۲/۱۳ - ۳۸/۶	۲۹۸/۳۲۸-۱۵/۱۵	۱۰۰	۳۹
۵	Dibuthylether	1-propanol	-	۰/۱ - ۱۴۰	۲۹۳/۳۵۳-۱۵/۱۵	۱۲۰	۴۰
۶	Dibuthylether	1-hexanol	-	۰/۱ - ۱۴۰	۲۹۳/۳۵۳-۱۵/۱۵	۱۲۰	۴۰
۷	Dibuthylether	2-propanol	-	۰/۱ - ۱۴۰	۲۹۳/۳۵۳-۱۵/۱۵	۱۲۰	۴۰
۸	N,N-Dimethylformamide	1-Chlorobutane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۱
۹	N,N-Dimethylformamide	1-Chlorohexane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۱
۱۰	N,N-Dimethylformamide	1-Chlorooctane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۱
۱۱	N,N-Dimethylformamide	1-Chlorobutane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۱
۱۲	N,N-Dimethylformamide	1-Chlorohexane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۱
۱۳	N,N-Dimethylformamide	1-Chlorooctane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۱
۱۴	1-methyl-2-pyrrolidone	1-Chlorobutane	-	۰/۱ - ۱۸/۱	۲۹۸/۱۵	۴۵	۴۲
۱۵	1-methyl-2-pyrrolidone	1-Chlorohexane	-	۰/۱ - ۱۸/۱	۲۹۸/۱۵	۴۵	۴۲
۱۶	1-methyl-2-pyrrolidone	1-Chlorooctane	-	۰/۱ - ۱۸/۱	۲۹۸/۱۵	۴۵	۴۲
۱۷	1-methyl-2-pyrrolidone	1,2-Dichloroethane	-	۰/۱ - ۱۸/۱	۲۹۸/۱۵	۴۵	۴۲
۱۸	1-methyl-2-pyrrolidone	1,4-Dichlorobutane	-	۰/۱ - ۱۸/۱	۲۹۸/۱۵	۴۵	۴۲
۱۹	1-methyl-2-pyrrolidone	1,6-Dichlorohexane	-	۰/۱ - ۱۸/۱	۲۹۸/۱۵	۴۵	۴۲
۲۰	N,N-Dimethylformamide	1,2-Dichloroethane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۳
۲۱	N,N-Dimethylformamide	1,4-Dichlorobutane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۳
۲۲	N,N-Dimethylformamide	1,6-Dichlorohexane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۳
۲۳	N,N-Dimethylformamide	1,2-Dichloroethane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۳
۲۴	N,N-Dimethylformamide	1,4-Dichlorobutane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۳
۲۵	N,N-Dimethylformamide	1,6-Dichlorohexane	-	۰/۱ - ۲۰/۱	۲۹۸/۱۵	۴۵	۴۳

ادامه جدول-(۳-۱)

شماره مخلوط	جزء A	جزء B	جزء C	$\Delta p/\text{MPa}$	$\Delta T/\text{K}$	تعداد نقاط	Ref
۲۶	Hexane	1-Chlorobutane	-	۰/۱ - ۳۵	۲۸۳/۳۲۳-۱۵/۱۵	۲۷۰	۴۴
۲۷	Hexane	Decane	-	۰/۱ - ۶۰	۲۷۸/۴۶۳-۱۵/۱۵	۱۵۶	۴۵
۲۸	Hexane	Hexadecane	-	۰/۱ - ۶۰	۲۷۸/۴۶۳-۱۵/۱۵	۱۳۷	۴۵
۲۹	Hexane	Dodecane	-	۰/۱ - ۴۳۶	۲۹۸/۳۴۸-۳۱/۱۳	۱۴	۳۷
۳۰	Decane	Hexadecane	-	۰/۱ - ۴۲۲/۸	۲۹۸/۳۴۸-۳۱/۱۳	۱۶	۳۷
۳۱	Ethanol	Heptane	-	۰/۱ - ۳۰	۲۹۳/۳۳۳-۱۵/۱۵	۲۴۵	۴۶
۳۲	Ethanol	Methyl cyclohexane	-	۵ - ۴۵	۳۲۳/۱۵	۶۳	۴۷
۳۳	1,4-Butanediol	1-Decanol	-	۰/۳۳ - ۳۹/۴	۲۹۸/۱۵	۹۴	۴۸
۳۴	Butan-2-one	2,2,4-Trimethylpentane	-	۰/۱	۲۷۸/۳۳۸-۱۵/۱۳	۷۹	۴۹
۳۵	Buthylamine	1-pentanol	-	۰/۱	۳۰۳/۳۲۳-۱۵/۱۵	۲۷	۵۰
۳۶	Buthylamine	1-Hexanol	-	۰/۱	۳۰۳/۳۲۳-۱۵/۱۵	۲۷	۵۰
۳۷	Buthylamine	1-Heptanol	-	۰/۱	۳۰۳/۳۲۳-۱۵/۱۵	۲۷	۵۰
۳۸	Buthylamine	1-octanol	-	۰/۱	۳۰۳/۳۲۳-۱۵/۱۵	۲۷	۵۰
۳۹	Buthylamine	1-butanol	-	۰/۱	۳۰۳/۳۲۳-۱۵/۱۵	۲۷	۵۰
۴۰	Acetic acide	Ethyl acetate	-	۰/۱	۲۹۸/۳۰۳-۱۵/۱۵	۳۳	۵۱
۴۱	Acetic acide	Methyl acetate	-	۰/۱	۲۹۸/۳۰۳-۱۵/۱۵	۳۳	۵۱
۴۲	Acetic acide	Ethanol	-	۰/۱	۲۹۸/۳۰۳-۱۵/۱۵	۳۳	۵۱
۴۳	Decane	Icosane	-	۰/۱	۲۹۳/۳۴۳-۱۵/۱۵	۲۴	۵۲
۴۴	Decane	Docosane	-	۰/۱	۳۰۳/۳۴۳-۱۵/۱۳	۲۰	۵۲
۴۵	Decane	Tetracosane	-	۰/۱	۳۰۳/۳۴۳-۱۵/۱۳	۱۶	۵۲
۴۶	Decane	Icosane	Tetracosane	۰/۱	۳۰۵/۳۴۳-۱۵/۱۵	۲۳	۵۲

ادامه جدول-(۳-۱)

شماره مخلوط	جزء A	جزء B	جزء C	$\Delta p/\text{MPa}$	$\Delta T/\text{k}$	تعداد نقاط	Ref
۴۷	Ethanol	Butyl Acetate	-	۰/۱	۲۹۳/۳۱۳-۱۵/۱۵	۱۵	۵۳
۴۸	Ethanol	Trans-2-Hexynyl Acetate	-	۰/۱	۲۹۳/۳۱۳-۱۵/۱۵	۱۵	۵۳
۴۹	Ethanol	Cis-3-Hexenyl Acetate	-	۰/۱	۲۹۳/۳۱۳-۱۵/۱۵	۱۵	۵۳
۵۰	Ethyl Acetate	Ethanol	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۵۴
۵۱	Ethyl Acetate	Propan-1-ol	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۵۴
۵۲	Ethyl Acetate	Propan-2-ol	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۵۴
۵۳	Ethyl Acetate	Butan-1-ol	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۵۴
۵۴	Ethyl Acetate	2-Methylpropan-1-ol	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۵۴
۵۵	Ethyl Acetate	2-Methyl propan-2-ol	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۵۴
۵۶	Cyclohexane	Methyl Acetate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۵۷	Cyclohexane	Ethyl Acetate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۵۸	Cyclohexane	Buthyl Acetate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۵۹	Cyclohexane	Isoamyl Acetate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۶۰	Cyclohexane	Methyl propionate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۶۱	Cyclohexane	Ethyl Propionate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۶۲	Cyclohexane	Ethyl butyrate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۶۳	Cyclohexane	Methyl Methacrylate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۶۴	Cyclohexane	Ethyl Methacrylate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۶۵	Cyclohexane	Buthyl Methacrylate	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۸	۵۵
۶۶	Cyclohexane	Nonane	-	۰/۱ - ۳۹/۶۲	۲۹۸/۳۲۸-۱۵/۱۵	۲۸۷	۵۶
۶۷	Cyclohexane	Hexadecane	-	۶/۶۲-۸۹۵/۰۵۳	۳۱۸/۳۸۸-۱۵/۱۵	۱۸۰	۵۷

ادامه جدول (۳-۱)

شماره مخلوط	جزء A	جزء B	جزء C	$\Delta p/\text{MPa}$	$\Delta T/\text{k}$	تعداد نقاط	Ref
۶۸	Cyclohexanone	Hexane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۲	۵۸
۶۹	Cyclohexanone	Heptane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۲	۵۸
۷۰	Cyclohexanone	Octane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۲	۵۸
۷۱	Cyclohexanone	Nonane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۲	۵۸
۷۲	Cyclohexanone	Decane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۲	۵۸
۷۳	Cyclohexanone	Dodecane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۲	۵۸
۷۴	Cyclohexanone	Trimethylpentane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۲	۵۸
۷۵	Methyl Methacrylate	Propan-2-ol	-	۰/۱	۲۹۳/۳۰۸-۱۵/۱۵	۱۰	۵۹
۷۶	Methyl Methacrylate	2-Methyl propan-1-ol	-	۰/۱	۲۹۳/۳۰۸-۱۵/۱۵	۱۰	۵۹
۷۷	Methyl Methacrylate	Butan-2-ol	-	۰/۱	۲۹۳/۳۰۸-۱۵/۱۵	۱۰	۵۹
۷۸	Methyl Methacrylate	2-Methyl propan-2-ol	-	۰/۱	۲۹۳/۳۰۸-۱۵/۱۵	۱۰	۶۰
۷۹	Methyl Methacrylate	Diethyl ether	-	۰/۱	۲۹۳/۳۰۸-۱۵/۱۵	۵	۶۰
۸۰	Methyl Methacrylate	Dibutyl ether	-	۰/۱	۲۹۸/۳۱۸-۱۵/۱۵	۲۵	۶۰
۸۱	Methyl Methacrylate	Cyclohexane	-	۰/۱	۲۹۸/۳۱۸-۱۵/۱۵	۲۵	۶۰
۸۲	Methyl Methacrylate	Diisopropyl ether	-	۰/۱	۲۹۸/۳۱۸-۱۵/۱۵	۲۵	۶۰
۸۳	2-Ethoxy ethanol	1,4-Dioxane	-	۰/۱	۲۹۸/۳۱۳-۱۵/۱۵	۱۳	۶۱
۸۴	2-Ethoxy ethanol	1,2-Dimethoxy ethanol	-	۰/۱	۲۹۸/۳۱۳-۱۵/۱۵	۱۸	۶۱
۸۵	Ethyl Chloroacetate	Hexane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۶۲
۸۶	Ethyl Chloroacetate	Heptane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۶۲
۸۷	Ethyl Chloroacetate	Octane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۶۲
۸۸	Ethyl Chloroacetate	Nonane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۶۲
۸۹	Ethyl Chloroacetate	Decane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۶۲
۹۰	Ethyl Chloroacetate	Dodecane	-	۰/۱	۲۹۸/۳۰۸-۱۵/۱۵	۱۵	۶۲

۳-۲- محاسبه‌ی توصیف‌کننده‌ها

یکی از مشکلات عملی و چالش برانگیز در مدل‌سازی QSPR مخلوط‌ها، چگونگی محاسبه‌ی توصیف‌کننده‌ها برای آن‌هاست. در ساده‌ترین حالت، می‌توان یک مخلوط دوتایی از دو جزء A و B را در کسرمولی‌های مختلف در دما و فشار محیط در نظر بگیرید. اکنون سوال این است که چگونه باید برای چنین مخلوطی در کسرمولی‌های مختلف، توصیف‌کننده‌ها را محاسبه نمود؟ برای این منظور در چند مطالعه‌ی انجام شده روی مدل‌سازی QSPR مخلوط‌ها در دما و فشار محیط، دو روش کلی زیر معرفی شده‌اند [۶۳-۶۵].

۱- استفاده از توصیف‌کننده‌های وزن دار نشده^۱

۲- استفاده از توصیف‌کننده‌های وزن دار شده^۲

توصیف‌کننده‌های وزن دار نشده، برای مدل‌سازی خواصی از مخلوط به کار می‌رود که مقدار خاصیت مخلوط به کسرمولی اجزای مخلوط وابسته نباشد مثل رفتار آزنوتروپ در مخلوط‌های دوتایی [۶۴]. در این موارد هر توصیف‌کننده‌ی معین مثل D_i^{mix} برای یک مخلوط دوتایی A+B می‌تواند از طریق متوسط‌گیری ساده روی توصیف‌کننده‌های اجزای مخلوط طبق رابطه‌ی (۳-۱)،

$$D_i^{mix} = \frac{D_i^A + D_i^B}{2} \quad (3-1)$$

و یا از طریق جمع و قدرمطلق اختلاف توصیف‌کننده‌های اجزای مخلوط طبق روابط زیر بدست آید.

$$D_i^{mix,+} = D_i^A + D_i^B \text{ و } D_i^{mix,-} = |D_i^A - D_i^B| \quad (3-2)$$

در روابط (۳-۱) و (۳-۲)، D_i^A و D_i^B به ترتیب، توصیف‌کننده‌ی i ام برای دو جزء A و B در یک مخلوط دوتایی هستند. بدیهی است که در روش دوم تعداد توصیف‌کننده‌ها دو برابر روش اول است.

۱-Unweighted descriptors

۲-Weighted descriptors

برخلاف حالت فوق، از توصیف‌کننده‌های وزن‌دار شده برای پیش‌بینی خواصی از مخلوط‌ها استفاده می‌شود که مقدار خاصیت با تغییر کسر مولی اجزاء تغییر می‌کند مثل دانسیته و حجم اضافی، که بیشتر خواص مخلوط‌ها در این دسته جای دارند. در مطالعه‌ی این دسته از خواص مخلوط‌ها، می‌توان هر توصیف‌کننده‌ی یک مخلوط دوتایی A+B مثل D_i^{mix} را با استفاده از کسر مولی اجزاء و از طریق مجموع وزن‌دار شده‌ی مولی^۱ توصیف‌کننده‌های اجزای مخلوط از طریق رابطه‌ی (۳-۴)،

(۳-۳)

$$D_i^{mix} = x_A D_i^A + x_B D_i^B$$

و یا از طریق مجموع و قدر مطلق اختلاف وزن‌دار شده‌ی مولی^۲ توصیف‌کننده‌های اجزای مخلوط طبق روابط زیر بدست آورد.

$$D_i^{mix,+} = x_A D_i^A + x_B D_i^B \text{ و } D_i^{mix,-} = |x_A D_i^A - x_B D_i^B| \quad (۳-۴)$$

در روابط (۳-۳) و (۴-۳) x_A و x_B به ترتیب کسر مولی اجزای A و B در مخلوط دوتایی هستند. لازم به ذکر است که در این روش، هر مخلوط دوتایی حتی در دما و فشار ثابت با توصیف‌کننده‌های مختلفی با توجه به کسر مولی بیان می‌شود زیرا در این حالت مقدار خاصیت موردنظر با تغییر کسر مولی (حتی در دما و فشار ثابت) تغییر می‌کند.

در این پایان‌نامه از دو روش سهم‌گروه و توصیف‌کننده‌های تئوری برای محاسبه‌ی توصیف‌کننده‌های اجزای مخلوط استفاده شده است، و برای محاسبه‌ی توصیف‌کننده‌های مخلوط، روش توصیف‌کننده‌های وزن‌دار شده به کار رفته که در ادامه راجع به هر کدام به طور مفصل شرح داده خواهد شد.

۱-Mole-weighted sum

۲-Mole-weighted sum and absolute difference

۳-۳- دسته‌بندی داده‌ها

پس از محاسبه‌ی توصیف‌کننده‌ها و قبل از شروع مراحل مدل‌سازی باید داده‌ها به سه دسته‌ی آموزش، ارزیابی و تست تقسیم شوند. اما در این مرحله مشکل دوم برای مدل‌سازی QSPR مخلوط‌ها بروز می‌کند و آن انتخاب درست سری تست است. واقعیت این است که در مدل‌سازی QSPR مخلوط‌ها، روش قراردادی ارزیابی تقاطعی، که در آن نقاط یا ترکیبات به صورت تصادفی در سری تست قرار می‌گیرند قابل قبول نیست چون سبب اغراق در صحت عملکرد پیش‌بینی مدل بهینه شده می‌شود، علت این اتفاق این است که در مطالعه‌ی QSPR مخلوط‌ها، مخلوط‌هایی با اجزای یکسان ولی با کسرمولی‌های متفاوت در سری داده‌ها وجود دارند که سبب می‌شود سری‌های آموزش و تست نقطه داده‌هایی مربوط به یک مخلوط را شامل شوند، در نتیجه عملکرد مدل در پیش‌بینی خواص موردنظر در مخلوط‌ها نمی‌تواند به درستی مورد تایید قرار گیرد. به همین دلیل سه استراتژی متفاوت برای انتخاب سری تست در مطالعه‌ی خواص مخلوط‌ها در مراجع ارائه شده است که در ادامه به اختصار راجع به آنها شرح داده خواهد شد

۱-رد نقطه‌ها^۱: در این روش هر بار مجموعه‌ای از نقطه داده‌ها به صورت تصادفی در سری تست قرار می‌گیرند. در این شرایط، هر مخلوط به طور همزمان چند نقطه داده در سری آموزش و چند نقطه داده در سری تست دارد. این روش، ضعیف‌ترین روش انتخاب سری تست است.

۲-رد مخلوط‌ها^۲: همه‌ی نقطه داده‌های مربوط به یک مخلوط با اجزای معین و کسرمولی‌های معین در دما و فشارهای مختلف، انتخاب و برای سری تست کنار گذاشته می‌شود، اما سایر نقاط همان مخلوط در بقیه کسرمولی‌ها در سری آموزش قرار می‌گیرند. به این ترتیب هر مخلوط با کسرمولی معین یا در سری تست هست یا نیست اما نمی‌تواند همزمان در هر دو سری باشد. این

۱-Point out

۲-Mixtures out

روش نسبت به روش قبلی بهتر است و می‌تواند توانایی پیش‌بینی مدل را برای مخلوط‌های موجود اما با ترکیب جدید بیان کند.

۳-رد ترکیبات^۱: در این روش تمام نقطه داده‌های مربوط به یک مخلوط به طور کامل در سری تست قرار می‌گیرند و دیگر اثری از این مخلوط در سری آموزش نیست. این روش انتخاب سری تست بسیار سختگیرانه است اما در صورت پاسخ‌دهی خوب، نشان دهنده‌ی اعتبار مدل بهینه شده است زیرا می‌تواند داده‌های مخلوط‌های کاملاً جدید را پیش‌بینی کند [۶۳-۶۵].

در این تحقیق از روش ردمخلوط‌ها برای سری ارزیابی و از روش ردترکیبات برای سری تست استفاده گردید. مخلوط‌های شماره ۳، ۴، ۸، ۱۶، ۳۰، ۳۷، ۴۱، ۴۵، ۵۲، ۵۷، ۷۱، ۷۸، ۸۸ از جدول (۳-۱) به عنوان سری تست به روش ردترکیبات انتخاب شده‌اند.

۱-Compounds out

۳-۴- مدل سازی شبکه عصبی

۳-۴-۱- مدل سازی شبکه ی عصبی مصنوعی با استفاده از توصیف کننده های سهم گروه

۳-۴-۱-۱- محاسبه ی توصیف کننده های مخلوط ها با استفاده از روش سهم گروه

برای محاسبه ی توصیف کننده های مخلوط ها با استفاده از روش سهم گروه، ابتدا باید گروه های ساختاری اجزای مخلوط ها را مشخص نمود. پس از آنالیز ساختار اجزاء ۹۰ مخلوط مورد مطالعه در این پایان نامه، ۱۵ گروه عاملی در این ترکیبات شناسایی شدند که نام این گروه های عاملی و فرمول شیمیایی آنها در جدول (۲-۳) گزارش شده است.

جدول (۲-۳)- گروه های عاملی موجود در اجزای مخلوط های مورد مطالعه در این پایان نامه طبق روش

سهم گروه

	گروه عاملی	فرمول شیمیایی
۱	متیل	$-\text{CH}_3$
۲	کربن نوع دوم	$>\text{CH}_2$
۳	کربن نوع سوم	$\begin{array}{c} \\ -\text{CH} \\ \end{array}$
۴	کربن نوع چهارم	$\begin{array}{c} \\ -\text{C}- \\ \end{array}$
۵	الکل نوع اول	$-\text{CH}_2\text{OH}$
۶	الکل نوع دوم	$>\text{CHOH}$
۷	الکل نوع سوم	$\begin{array}{c} \\ -\text{COH} \\ \end{array}$
۸	کربن آلدئیدی	$\text{H}-\text{C}=\text{O}$
۹	کربوکسیلیک اسید	$-\text{COOH}$
۱۰	استر	$-\text{COO}-$
۱۱	کتون	$>\text{C}=\text{O}$
۱۲	اتر	$-\text{O}-$
۱۳	آمین نوع اول	$-\text{NH}_2$
۱۴	آمین نوع سوم	$\begin{array}{c} \\ -\text{N}- \end{array}$
۱۵	کلر	$-\text{Cl}$

پس از تعیین گروههای عاملی اجزای سازنده‌ی مخلوط‌ها، توصیف‌کننده‌های مخلوط‌ها با استفاده از کسرمولی آنها و رابطه‌ی (۳-۴) محاسبه گردید. در هر مخلوط تعداد گروههای عاملی موجود در مولکول جزء موردنظر به عنوان توصیف‌کننده‌های حاصل از روش سهم‌گروه برای آن جزء انتخاب شدند و در صورتی که هر گروه عاملی در مولکول جزء موردنظر وجود نداشت، عدد صفر برای آن گروه در نظر گرفته شد. به عنوان مثال برای مخلوط دوتایی هگزان (۱) و سیکلوهگزانون (۲) با $x_1 = 0.2$ در دما و فشار محیط می‌توان نوشت:

$$D_{-CH_3}^{mix} = x_1 \times 2 + x_2 \times 0 = 0.2 \times 2 = 0.4$$

$$D_{-CH_2}^{mix} = x_1 \times 4 + x_2 \times 5 = 0.2 \times 4 + 0.8 \times 5 = 4.8$$

$$D_{>C=O}^{mix} = x_1 \times 0 + x_2 \times 1 = 0.8 \times 1 = 0.8$$

مقدار سایر گروههای عاملی که در این مخلوط وجود ندارد صفر در نظر گرفته می‌شود. به این ترتیب نحوه‌ی آرایش توصیف‌کننده‌های این مخلوط در ماتریس ورودی اولیه مطابق جدول (۳-۳) خواهد بود.

جدول (۳-۳) - نحوه‌ی آرایش توصیف‌کننده‌های مخلوط 0.2 Hexane + 0.8 Cyclohexanone در دما و فشار محیط بر اساس روش سهم‌گروه

گروه عاملی	-CH ₃	>CH ₂	-CH-	-C-	-CH ₂ OH	>CHOH	-COH	H-C=O	-COOH	-COO-	>C=O	-O-	-NH ₂	-N-	-Cl
مقدار توصیف‌کننده	۰/۴	۴/۸	.	.	.	.	.	.	.	.	۰/۸	.	.	.	.

پس از محاسبه‌ی توصیف‌کننده‌های سهم‌گروه برای مخلوط‌های مورد مطالعه به ازای کسرمولی‌های مختلف برای هرکدام از آنها، دو متغیر تجربی دما و فشار نیز به این توصیف‌کننده‌ها اضافه شدند.

سپس داده‌های مربوط به توصیف‌کننده‌های سهم‌گروه پس از دسته‌بندی و مرتب‌سازی CR^۱ برای مدل‌سازی وارد شبکه‌ی عصبی مصنوعی گردیدند. در مرتب‌سازی CR همبستگی توصیف‌کننده‌ها با خاصیت موردنظر یعنی دانسیته توسط نرم‌افزار SPSS محاسبه شد و بعد به ترتیب نزول قدرمطلق مقادیر در ماتریس ورودی مرتب شدند. توصیف‌کننده‌ای که بیشترین مقدار همبستگی با متغیر وابسته را دارد زودتر از سایر توصیف‌کننده‌ها به شبکه وارد گردید و سایر توصیف‌کننده‌ها پس از آن در ستون‌های بعدی ماتریس ورودی شبکه مرتب شدند. لازم به ذکر است که به دلیل وابستگی دانسیته به دما و فشار این دو متغیر به ترتیب در ستون‌های اول و دوم ماتریس ورودی قرار می‌گیرند و سایر توصیف‌کننده‌های سهم‌گروه به ترتیب قدرمطلق مقادیر کاهشی در ستون‌های بعدی ماتریس قرار می‌گیرند. مقادیر همبستگی توصیف‌کننده‌های حاصل از روش سهم‌گروه با دانسیته در جدول (۳-۴) ارائه شده است.

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

توصیف کننده	-CH ₃	>CH ₂	$\begin{array}{c} \\ -CH \\ \end{array}$	$\begin{array}{c} \\ -C- \\ \end{array}$	-CH ₂ OH	>CHOH	$\begin{array}{c} \\ -COH \\ \end{array}$	$\begin{array}{c} \\ H-C=O \end{array}$
مقدار دانسیته همبستگی با	-۰/۳۹۶	-۰/۴۰۴	-۰/۰۸۹	-۰/۰۲۹	۰/۰۶۴	-۰/۰۲۹	-۰/۰۰۷	۰/۳۰۳
توصیف کننده	-COO-	>C=O	-O-	-NH ₂	$\begin{array}{c} \\ -N- \end{array}$	-Cl	-COOH	
مقدار دانسیته همبستگی با	۰/۱۶۲	۰/۴۱۹	-۰/۰۰۴	-۰/۰۸۹	۰/۶۰۷	۰/۶۴۲	۰/۱۸۳	

^۱-Correlation Ranking

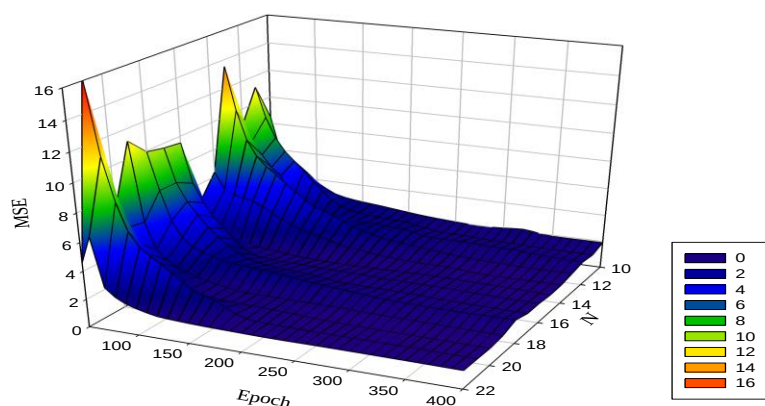
۲-۱-۴-۳- مدلسازی شبکه‌ی عصبی با توصیف‌کننده‌های حاصل از روش سهم‌گروه-GCM

ANN

همان‌طور که در قسمت (۳-۴-۱-۱) گفته شد، ۱۵ توصیف‌کننده بر مبنای روش سهم‌گروه به همراه دو متغیر تجربی دما و فشار به عنوان توصیف‌کننده‌های ورودی مدل انتخاب شدند. لایه‌ی خروجی نیز شامل یک نرون است که دانسیته‌ی متناظر با ورودی‌ها را نشان می‌دهد. برای آموزش شبکه دو الگوریتم آموزشی تنظیم بایزین و لونبرگ-مارکوات و برای انتخاب تابع انتقال مناسب، دو تابع انتقال تانژانت سیگموئیدی و لگاریتم سیگموئیدی 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 تعداد نرون‌های لایه‌ی مخفی و تعداد دور آموزش جهت آموزش شبکه استفاده گردید و به‌ازای همه‌ی ترکیب‌های ممکن از این دو پارامتر شبکه‌های طراحی شده بهینه‌سازی شدند. در تمامی این شبکه‌ها تعداد توصیف‌کننده‌ها ثابت و برابر ۱۷ است اما تعداد نرون‌های لایه‌ی مخفی از ۱۰ تا ۲۳ با گام یک و تعداد دور آموزشی از ۵۰ تا ۴۰۰ با گام ۱۰ در نظر گرفته شدند. به‌ازای همه‌ی ترکیبات ممکن از این پارامترها، شبکه‌های طراحی شده به‌طور جداگانه آموزش داده شدند. ترکیبی از دو پارامتر تعداد گره لایه‌ی و تعداد دور آموزشی که بتواند کمترین مقدار مربعات خطا (MSE) را برای سری ارزیابی ایجاد کند به‌عنوان پارامترهای بهینه انتخاب می‌شود. در این قسمت به دلیل تعداد زیاد داده‌های حداقل مربعات خطا به دست‌آمده، بخشی از روند تغییرات پارامترهای شبکه در بهینه‌سازی هم‌زمان پارامترها به همراه مقادیر MSE بدست آمده برای سری ارزیابی در جدول‌های (۳-۵) تا (۳-۸) گردآوری شده است و نتایج حاصل از کل محاسبات در نمودارهای سه بعدی (۳-۱) تا (۳-۴) نشان داده است.

جدول (۵-۳)- 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

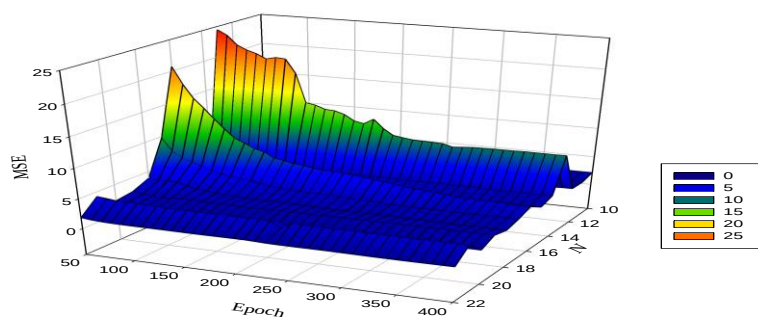
تعداد گره لایه پنهان											
دور آموزشی		۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
	۲۰۰	۲/۴۳۱	۱/۸۵۵	۱/۳۹۹	۱/۶۶۱	۱/۸۷۵	۲/۲۵۰	۱/۵۵۳	۱/۵۲۹	۱/۴۲۹	۱/۳۶۰
	۲۱۰	۲/۲۷۵	۱/۷۵۵	۱/۴۰۰	۱/۶۵۱	۱/۷۸۷	۲/۱۸۰	۱/۵۴۷	۱/۴۹۸	۱/۳۹۵	۱/۳۴۵
	۲۲۰	۲/۱۵۲	۱/۶۷۲	۱/۴۰۷	۱/۶۴۱	۱/۶۶۸	۲/۱۲۰	۱/۵۴۴	۱/۴۶۲	۱/۳۷۲	۱/۳۳۲
	۲۳۰	۲/۰۵۶	۱/۶۱۱	۱/۴۱۷	۱/۶۲۶	۱/۶۵۳	۲/۰۵۶	۱/۵۴۰	۱/۴۲۲	۱/۳۵۴	۱/۳۲۵
	۲۴۰	۱/۹۸۲	۱/۵۷۰	۱/۴۲۹	۱/۶۰۹	۱/۶۳۹	۲/۹۸۶	۱/۵۳۵	۱/۳۸۳	۱/۳۴۲	۱/۳۲۵
	۲۵۰	۱/۹۱۸	۱/۵۴۴	۱/۴۴۲	۱/۵۹۱	۱/۶۲۵	۱/۹۱۶	۱/۵۲۷	۱/۳۵۰	۱/۳۳۵	۱/۳۲۸
	۲۶۰	۱/۸۴۷	۱/۵۲۹	۱/۴۵۶	۱/۵۷۸	۱/۶۱۳	۱/۸۴۸	۱/۵۱۲	۱/۳۲۴	۱/۳۲۹	۱/۳۳۱
	۲۷۰	۱/۷۷۳	۱/۴۹۳	۱/۴۶۹	۱/۵۶۹	۱/۶۰۳	۱/۷۹۵	۱/۴۹۱	۱/۳۰۱	۱/۳۲۴	۱/۳۳۲
	۲۸۰	۱/۷۰۸	۱/۴۹۱	۱/۵۰۲	۱/۵۶۴	۱/۵۹۴	۱/۷۵۸	۱/۴۶۸	۱/۲۸۲	۱/۳۲۳	۱/۳۳۲
	۲۹۰	۱/۶۵۲	۱/۴۹۹	۱/۴۶۶	۱/۵۵۷	۱/۵۸۳	۱/۷۳۰	۱/۴۴۹	۱/۲۷۴	۱/۳۲۳	۱/۳۳۱
	۳۰۰	۱/۶۰۷	۱/۴۴۹	۱/۴۵۶	۱/۵۴۷	۱/۵۷۱	۱/۷۰۳	۱/۴۳۵	۱/۲۷۳	۱/۳۲۲	۱/۳۲۸
	۳۱۰	۱/۵۷۴	۱/۳۷۹	۱/۴۵۶	۱/۵۳۲	۱/۵۲۵	۱/۶۷۸	۱/۴۲۳	۱/۲۶۹	۱/۳۱۹	۱/۳۲۵
	۳۲۰	۱/۵۵۲	۱/۳۴۳	۱/۴۴۶	۱/۵۱۱	۱/۵۰۹	۱/۶۵۹	۱/۴۱۱	۱/۲۶۳	۱/۳۱۷	۱/۳۲۱
	۳۳۰	۱/۵۳۵	۱/۳۳۰	۱/۴۳۴	۱/۴۸۹	۱/۴۹۲	۱/۶۵۰	۱/۳۹۹	۱/۲۵۶	۱/۳۱۴	۱/۳۱۷
	۳۴۰	۱/۵۲۰	۱/۳۲۶	۱/۴۲۶	۱/۴۶۹	۱/۴۷۴	۱/۶۴۷	۱/۳۸۸	۱/۲۵۰	۱/۳۰۹	۱/۳۱۳
	۳۵۰	۱/۵۰۸	۱/۳۲۸	۱/۴۲۰	۱/۴۵۰	۱/۴۵۷	۱/۶۴۷	۱/۳۷۷	۱/۲۴۵	۱/۳۰۴	۱/۳۰۸
	۳۶۰	۱/۵۰۰	۱/۳۳۰	۱/۴۰۸	۱/۴۳۲	۱/۴۴۲	۱/۶۵۰	۱/۳۶۷	۱/۲۳۹	۱/۲۹۷	۱/۳۰۲
	۳۷۰	۱/۴۹۹	۱/۳۳۳	۱/۳۹۳	۱/۴۱۷	۱/۴۲۸	۱/۶۵۵	۱/۳۵۹	۱/۲۳۳	۱/۲۸۹	۱/۲۷۹
	۳۸۰	۱/۵۰۴	۱/۳۳۸	۱/۳۷۴	۱/۴۰۶	۱/۴۱۶	۱/۶۶۳	۱/۳۵۱	۱/۲۲۶	۱/۲۷۹	۱/۲۹۱
	۳۹۰	۱/۵۱۲	۱/۳۴۵	۱/۳۳۷	۱/۴۰۲	۱/۴۰۷	۱/۶۷۱	۱/۳۴۵	۱/۲۱۴	۱/۲۶۶	۱/۲۸۴
	۴۰۰	۱/۵۱۹	۱/۳۴۷	۱/۳۰۸	۱/۴۰۲	۱/۴۰۰	۱/۶۸۱	۱/۳۳۹	۱/۲۲۰	۱/۲۵۴	۱/۲۷۶



شکل (۳-۱)- نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۵-۳)

جدول (۳-۶) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت

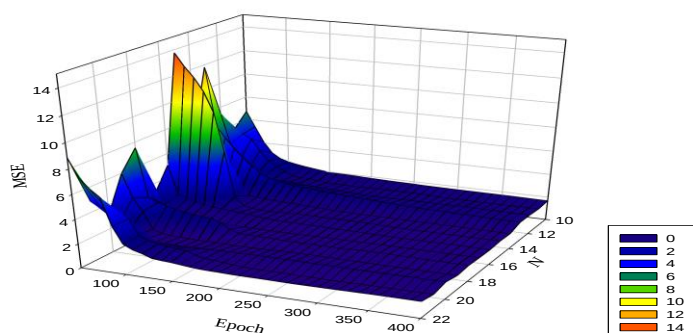
تعداد گره لایه پنهان											
دور آموزشی		۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
	۲۰۰	۱۰/۶۴۱	۲/۷۶۳	۲/۱۲۸	۵/۶۵۲	۳/۰۰۵	۲/۱۷۹	۲/۲۲۵	۱/۳۱۱	۳/۴۹۷	۱/۴۶۱
	۲۱۰	۱۱/۵۶۰	۲/۷۶۱	۲/۱۱۴	۵/۵۹۶	۲/۹۹۱	۲/۱۳۶	۲/۲۳۶	۱/۲۶۴	۳/۵۲۹	۱/۴۶۵
	۲۲۰	۱۰/۰۹۳	۲/۷۵۹	۲/۰۹۷	۵/۵۳۰	۲/۸۷۸	۲/۱۰۰	۲/۲۴۹	۱/۲۲۹	۳/۵۱۱	۱/۴۱۷
	۲۳۰	۹/۰۸۸	۲/۷۵۸	۲/۰۵۳	۵/۴۰۹	۲/۸۷۲	۲/۰۳۳	۲/۲۴۵	۱/۲۱۰	۳/۴۹۲	۱/۳۴۴
	۲۴۰	۸/۸۳۵	۲/۷۵۶	۲/۰۳۷	۵/۱۷۶	۲/۷۷۵	۲/۰۰۷	۲/۲۵۲	۱/۲۰۵	۳/۴۳۴	۱/۳۱۵
	۲۵۰	۸/۶۱۸	۲/۷۵۵	۲/۰۲۹	۵/۰۱۴	۲/۶۸۵	۱/۹۸۵	۲/۲۶۲	۱/۲۰۲	۳/۴۴۱	۱/۳۰۹
	۲۶۰	۸/۵۶۱	۲/۷۵۲	۲/۰۳۱	۴/۸۳۱	۲/۶۶۱	۱/۹۷۰	۲/۲۷۰	۱/۲۰۶	۳/۴۴۱	۱/۳۰۹
	۲۷۰	۸/۵۲۲	۲/۷۵۰	۲/۰۳۲	۴/۵۳۴	۲/۵۴۵	۱/۹۶۱	۲/۲۸۴	۱/۲۱۳	۳/۴۵۲	۱/۳۱۸
	۲۸۰	۸/۴۶۹	۲/۷۴۸	۲/۰۲۴	۴/۲۹۹	۲/۴۶۵	۱/۹۵۳	۲/۲۸۸	۱/۲۱۸	۳/۴۳۴	۱/۳۲۱
	۲۹۰	۸/۰۲۰	۲/۷۴۲	۲/۰۰۸	۴/۱۴۱	۲/۳۹۹	۱/۹۴۲	۲/۲۹۳	۱/۲۱۷	۳/۴۳۱	۱/۳۱۶
	۳۰۰	۸/۱۶۳	۲/۷۳۷	۱/۹۸۴	۳/۹۹۷	۲/۳۹۸	۱/۹۲۷	۲/۲۹۳	۱/۲۱۷	۳/۴۲۴	۱/۳۰۲
	۳۱۰	۸/۲۹۷	۲/۷۳۲	۱/۹۴۸	۳/۸۹۹	۲/۳۸۳	۱/۹۰۷	۲/۲۸۸	۱/۲۱۷	۳/۴۰۵	۱/۲۸۵
	۳۲۰	۸/۳۱۸	۲/۷۲۸	۱/۸۹۳	۳/۷۲۵	۲/۳۷۶	۱/۸۸۹	۲/۲۸۳	۱/۲۱۳	۳/۳۸۵	۱/۲۶۸
	۳۳۰	۸/۳۶۰	۲/۷۲۵	۱/۹۰۴	۳/۴۹۵	۲/۳۶۴	۱/۸۷۱	۲/۲۷۹	۱/۲۰۸	۳/۳۷۶	۱/۲۵۶
	۳۴۰	۸/۳۶۱	۲/۷۲۲	۱/۹۱۷	۳/۳۷۷	۲/۳۴۰	۱/۸۴۸	۲/۲۷۶	۱/۲۰۲	۳/۳۷۹	۱/۲۵۱
	۳۵۰	۸/۳۵۸	۲/۷۱۹	۱/۹۲۰	۳/۴۰۳	۲/۳۳۸	۱/۸۲۶	۲/۲۷۴	۱/۱۹۷	۳/۳۷۳	۱/۲۵۰
	۳۶۰	۸/۴۰۵	۲/۶۴۴	۱/۹۱۷	۳/۳۹۰	۲/۳۲۷	۱/۸۰۹	۲/۲۶۹	۱/۱۹۳	۳/۳۷۵	۱/۲۵۲
	۳۷۰	۸/۳۸۴	۲/۶۴۵	۱/۹۱۱	۳/۳۳۷	۲/۳۲۷	۱/۷۹۳	۲/۲۶۴	۱/۱۹۱	۳/۳۵۰	۱/۲۵۷
	۳۸۰	۸/۳۸۶	۲/۶۴۳	۱/۹۰۴	۳/۲۹۰	۲/۳۲۸	۱/۷۷۸	۲/۲۵۵	۱/۱۸۹	۳/۳۲۰	۱/۲۶۳
	۳۹۰	۸/۳۷۰	۲/۶۳۷	۱/۸۹۸	۳/۲۵۹	۲/۳۳۰	۱/۷۶۲	۲/۲۴۸	۱/۱۸۵	۳/۳۳۴	۱/۲۶۸
	۴۰۰	۸/۴۰۹	۲/۶۲۹	۱/۸۹۱	۳/۲۰۱	۲/۳۳۳	۱/۷۴۸	۲/۲۴۶	۱/۱۸۷	۳/۳۹۸	۱/۲۷۲



شکل (۳-۲) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۶)

جدول (۷-۳) - 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

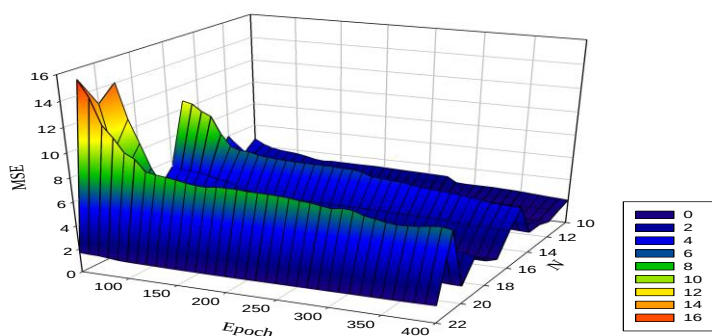
	تعداد گره لایه پنهان										
		۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
دور آموزشی	۲۰۰	۲/۱۳۴	۱/۴۱۳	۱/۴۰۴	۱/۳۵۲	۱/۲۰۴	۱/۳۴۷	۱/۱۷۲	۱/۴۱۰	۱/۲۱۸	۱/۳۲۴
	۲۱۰	۲/۰۸۸	۱/۳۸۱	۱/۴۰۶	۱/۳۴۲	۱/۲۰۹	۱/۳۲۵	۱/۱۸۸	۱/۴۱۷	۱/۱۷۳	۱/۳۰۲
	۲۲۰	۲/۰۴۶	۱/۳۶۶	۱/۴۰۸	۱/۳۲۹	۱/۲۱۶	۱/۳۱۹	۱/۲۰۶	۱/۴۱۹	۱/۱۳۳	۱/۲۸۹
	۲۳۰	۲/۰۰۸	۱/۳۵۷	۱/۴۰۷	۱/۳۱۶	۱/۲۲۵	۱/۳۱۳	۱/۲۲۳	۱/۴۱۸	۱/۱۲۵	۱/۲۹۰
	۲۴۰	۱/۹۷۲	۱/۳۴۹	۱/۴۰۳	۱/۳۰۸	۱/۲۳۲	۱/۳۱۰	۱/۲۲۳	۱/۴۱۵	۱/۱۲۰	۱/۲۹۲
	۲۵۰	۱/۹۳۴	۱/۳۴۲	۱/۳۹۴	۱/۳۰۳	۱/۲۳۷	۱/۳۱۰	۱/۲۳۷	۱/۴۱۲	۱/۱۱۳	۱/۲۹۰
	۲۶۰	۱/۸۹۳	۱/۳۳۵	۱/۳۷۹	۱/۲۹۹	۱/۲۴۱	۱/۳۱۷	۱/۲۳۵	۱/۴۰۹	۱/۱۱۴	۱/۲۸۴
	۲۷۰	۱/۸۵۰	۱/۳۲۸	۱/۳۶۱	۱/۲۹۱	۱/۲۴۴	۱/۳۲۸	۱/۲۳۲	۱/۴۰۶	۱/۱۱۵	۱/۲۷۷
	۲۸۰	۱/۸۱۲	۱/۳۲۳	۱/۳۴۳	۱/۲۷۸	۱/۲۴۷	۱/۳۳۹	۱/۲۲۷	۱/۴۰۳	۱/۱۱۰	۱/۲۷۷
	۲۹۰	۱/۷۸۴	۱/۳۲۰	۱/۳۲۸	۱/۲۶۳	۱/۲۴۹	۱/۳۴۹	۱/۲۲۳	۱/۴۰۰	۱/۱۰۶	۱/۲۸۴
	۳۰۰	۱/۷۶۲	۱/۳۱۸	۱/۳۱۸	۱/۲۴۹	۱/۲۵۲	۱/۳۵۵	۱/۲۱۹	۱/۳۹۸	۱/۱۰۶	۱/۲۹۱
	۳۱۰	۱/۷۴۴	۱/۳۱۷	۱/۳۱۳	۱/۲۴۰	۱/۲۵۴	۱/۳۵۷	۱/۲۱۴	۱/۳۹۷	۱/۱۰۸	۱/۲۹۹
	۳۲۰	۱/۷۲۸	۱/۳۱۶	۱/۳۱۰	۱/۲۳۹	۱/۲۵۵	۱/۳۵۴	۱/۲۰۹	۱/۳۹۶	۱/۱۱۲	۱/۳۰۸
	۳۳۰	۱/۷۱۵	۱/۳۱۴	۱/۳۰۷	۱/۲۴۵	۱/۲۵۶	۱/۳۴۹	۱/۲۰۵	۱/۳۹۵	۱/۱۱۸	۱/۳۱۹
	۳۴۰	۱/۷۰۵	۱/۳۱۳	۱/۳۰۵	۱/۲۵۴	۱/۲۵۶	۱/۳۴۳	۱/۲۰۱	۱/۳۹۶	۱/۱۲۴	۱/۳۲۸
	۳۵۰	۱/۶۹۴	۱/۳۱۱	۱/۳۰۲	۱/۲۶۲	۱/۲۵۶	۱/۳۳۶	۱/۱۹۹	۱/۳۹۶	۱/۱۳۰	۱/۳۳۷
	۳۶۰	۱/۶۸۰	۱/۲۹۹	۱/۲۹۹	۱/۲۶۶	۱/۲۵۵	۱/۳۲۹	۱/۱۹۸	۱/۳۹۸	۱/۱۳۶	۱/۳۴۳
	۳۷۰	۱/۶۶۳	۱/۲۹۲	۱/۳۰۲	۱/۲۶۱	۱/۲۵۵	۱/۳۲۳	۱/۱۹۷	۱/۴۰۰	۱/۱۴۱	۱/۳۴۶
	۳۸۰	۱/۶۴۸	۱/۲۹۰	۱/۳۱۰	۱/۲۴۹	۱/۲۵۵	۱/۳۱۷	۱/۱۹۷	۱/۴۰۱	۱/۱۴۵	۱/۳۴۸
	۳۹۰	۱/۶۳۶	۱/۲۸۸	۱/۳۱۶	۱/۲۴۰	۱/۲۵۶	۱/۳۱۱	۱/۱۹۶	۱/۴۰۲	۱/۱۴۷	۱/۳۴۸
	۴۰۰	۱/۶۲۸	۱/۲۸۷	۱/۳۲۰	۱/۲۳۹	۱/۲۵۹	۱/۳۰۶	۱/۱۹۵	۱/۴۰۲	۱/۱۴۹	۱/۳۴۵



شکل (۳-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۷-۳)

جدول (۳-۸) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت

تعداد گره لایه پنهان											
دور آموزشی		۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
	۲۰۰	۲/۳۵۳	۲/۰۷۴	۶/۳۳۱	۳/۷۳۲	۱/۹۲۰	۲/۲۴۱	۴/۴۶۹	۲/۵۸۸	۷/۹۰۴	۱/۳۷۰
	۲۱۰	۲/۳۳۲	۲/۰۶۰	۶/۳۰۴	۳/۷۲۱	۱/۸۷۸	۲/۲۶۶	۴/۴۲۵	۲/۴۹۰	۷/۸۳۲	۱/۳۷۱
	۲۲۰	۲/۳۰۶	۲/۰۴۳	۶/۲۷۲	۳/۷۰۶	۱/۸۵۱	۲/۲۲۱	۴/۴۲۸	۲/۴۳۰	۷/۸۲۱	۱/۳۷۴
	۲۳۰	۲/۲۷۹	۲/۰۳۰	۶/۲۶۳	۳/۷۱۶	۱/۸۲۶	۲/۰۶۵	۴/۳۸۹	۲/۲۶۰	۷/۸۱۳	۱/۳۷۷
	۲۴۰	۲/۲۵۶	۲/۰۱۰	۶/۱۲۰	۳/۷۰۳	۱/۷۸۹	۲/۰۵۳	۴/۳۵۴	۲/۲۴۱	۷/۷۷۸	۱/۳۸۱
	۲۵۰	۲/۲۳۹	۲/۰۱۱	۵/۸۰۹	۳/۷۰۱	۱/۷۴۶	۲/۰۹۶	۴/۳۵۰	۲/۲۲۷	۷/۷۲۰	۱/۳۸۴
	۲۶۰	۲/۲۲۳	۲/۰۰۸	۵/۷۳۵	۳/۶۹۰	۱/۷۰۳	۲/۱۱۲	۴/۲۷۲	۲/۲۰۴	۷/۶۰۰	۱/۳۸۶
	۲۷۰	۲/۲۰۶	۲/۰۰۶	۵/۶۸۹	۳/۶۶۰	۱/۶۶۸	۲/۱۰۹	۴/۱۱۳	۲/۱۸۸	۷/۵۳۹	۱/۳۸۷
	۲۸۰	۲/۱۸۷	۲/۰۰۴	۵/۶۲۸	۳/۶۳۶	۱/۶۴۴	۲/۱۲۰	۳/۸۹۷	۲/۱۶۱	۷/۴۰۲	۱/۳۸۸
	۲۹۰	۲/۱۶۵	۲/۰۰۲	۵/۵۰۶	۳/۶۲۵	۱/۶۳۰	۲/۱۲۹	۳/۷۲۳	۲/۱۱۷	۷/۲۸۰	۱/۳۹۰
	۳۰۰	۲/۱۳۷	۱/۹۹۹	۵/۳۹۱	۳/۶۰۹	۱/۶۲۵	۲/۱۴۱	۳/۶۳۱	۲/۰۵۹	۷/۴۱۲	۱/۳۹۲
	۳۱۰	۲/۱۰۸	۱/۹۹۷	۵/۲۹۴	۳/۵۹۵	۱/۶۲۶	۲/۱۵۱	۳/۶۰۱	۲/۰۱۰	۷/۳۳۰	۱/۳۹۵
	۳۲۰	۲/۰۴۵	۱/۹۹۶	۵/۲۱۹	۳/۵۷۶	۱/۶۲۸	۲/۱۵۵	۳/۵۸۱	۱/۹۶۵	۷/۰۳۹	۱/۳۹۷
	۳۳۰	۲/۰۰۵	۱/۹۹۴	۵/۱۷۵	۳/۵۴۹	۱/۶۲۸	۲/۱۵۲	۳/۵۳۵	۱/۹۰۴	۶/۸۹۳	۱/۳۹۸
	۳۴۰	۲/۰۱۲	۱/۹۹۳	۵/۱۱۰	۳/۴۷۷	۱/۶۲۶	۲/۱۴۴	۳/۴۵۵	۱/۸۴۹	۶/۷۳۰	۱/۴۰۰
	۳۵۰	۲/۰۱۵	۱/۹۸۹	۴/۹۶۱	۳/۳۹۵	۱/۶۲۵	۲/۱۳۴	۳/۴۳۲	۱/۷۹۳	۶/۶۲۶	۱/۴۰۲
	۳۶۰	۱/۹۹۲	۱/۹۷۷	۴/۸۷۳	۳/۳۵۵	۱/۶۲۰	۲/۱۲۵	۳/۴۲۴	۱/۷۱۲	۶/۶۶۶	۱/۴۰۵
	۳۷۰	۱/۹۴۳	۱/۹۲۷	۴/۷۵۵	۳/۳۳۱	۱/۵۴۹	۲/۱۱۷	۳/۴۰۹	۱/۶۱۹	۶/۷۲۵	۱/۴۱۲
	۳۸۰	۱/۸۸۵	۱/۸۸۲	۴/۶۸۷	۳/۳۱۴	۱/۵۳۱	۲/۱۱۵	۳/۴۲۹	۱/۵۵۶	۶/۸۵۸	۱/۴۲۰
	۳۹۰	۱/۸۲۷	۱/۸۳۵	۴/۶۴۲	۳/۳۰۰	۱/۵۱۶	۲/۱۲۳	۳/۴۰۹	۱/۵۱۰	۶/۹۱۹	۱/۴۳۰
	۴۰۰	۱/۷۸۴	۱/۷۹۷	۴/۵۹۰	۳/۲۸۹	۱/۵۰۸	۲/۱۲۳	۳/۲۲۹	۱/۴۹۳	۶/۷۶۵	۱/۴۳۹

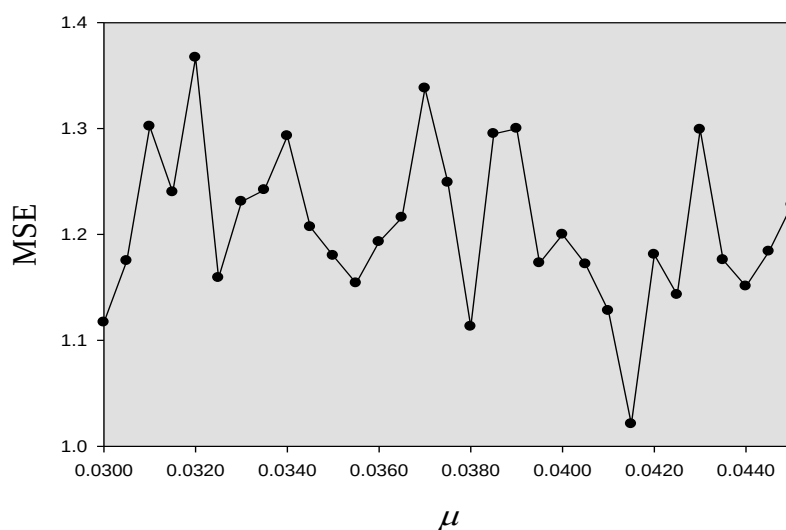


شکل (۳-۴) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۸)

جدول (۳-۹) - توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN

تابع انتقال	تابع آموزش	تعداد توصیف کننده	تعداد نرون لایه پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموئید	تنظیم بایزین	۱۷	۲۰	۳۹۰	۱/۲۱۴
لگاریتم سیگموئید	لونبرگ-مارکوارت	۱۷	۲۰	۳۹۰	۱/۱۸۵
تانژانت سیگموئید	تنظیم بایزین	۱۷	۲۱	۳۰۰	۱/۱۰۶
تانژانت سیگموئید	لونبرگ-مارکوارت	۱۷	۲۲	۲۰۰	۱/۳۷۰

با توجه به نتایج موجود در جدول (۳-۹) شبکه‌ی با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی با تعداد نرون لایه‌ی پنهان ۲۱ و دور آموزشی ۳۰۰ به عنوان شبکه‌ی بهینه انتخاب گردید. 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 حداقل مربعات خطا، MSE، مقدار بهینه‌ی μ ، ۰/۰۴۱۵ به دست آمد. نمودار مقدار میانگین مربعات خطا بر حسب μ حول نقطه‌ی بهینه نیز در شکل (۳-۵) رسم شده است. توابع و مقادیر پارامترهای بهینه شده‌ی GCM-ANN برای پیش‌بینی دانسیته‌ی مخلوط‌ها، در جدول (۳-۱۰) ارائه شده است.



شکل (۳-۵) - نمودار میانگین مربعات خطا بر حسب μ حول نقطه بهینه برای سری ارزیابی

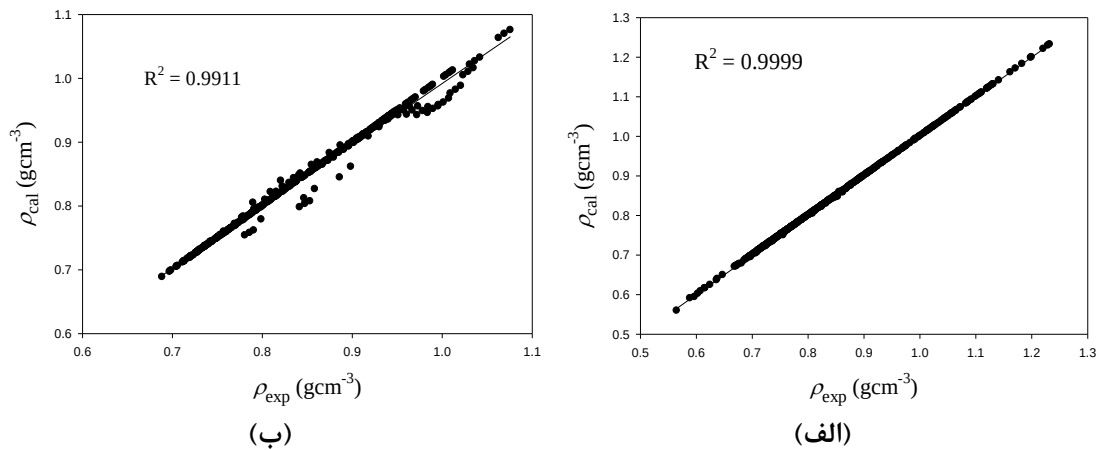
جدول (۳-۱۰) - توابع و مقادیر پارامترهای بهینه شده‌ی شبکه‌ی GCM-ANN برای پیش‌بینی دانسیته‌ی مخلوط‌ها برحسب دما و فشار

Train br	تابع آموزش
Tan sig	تابع انتقال لایه‌ی پنهان
pure line	تابع انتقال لایه‌ی خروجی
۲۱	تعداد نرون لایه‌ی پنهان
۱۷	تعداد متغیرهای ورودی
۳۰۰	تعداد دوره‌های آموزش
۰/۰۴۱۵	پارامتر μ
۱/۰۲۰	مقدار MSE سری ارزیابی

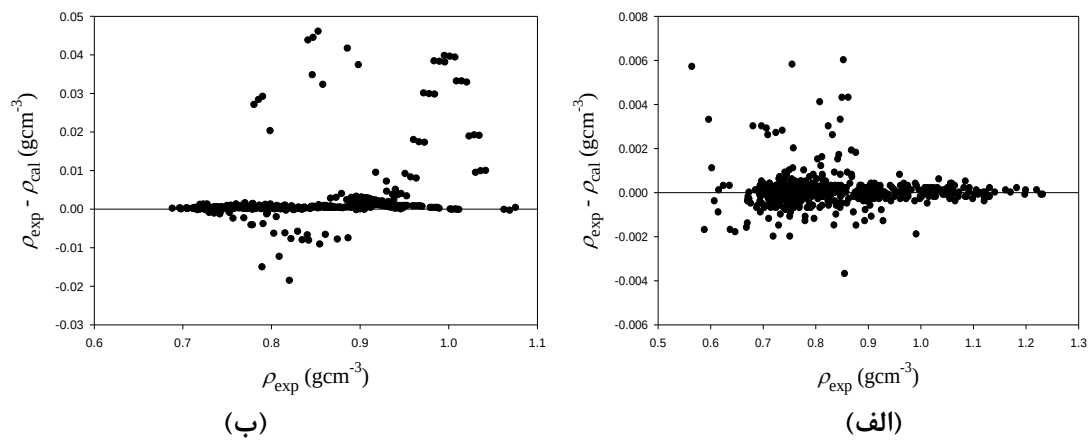
۳-۴-۱-۳- ارزیابی شبکه‌ی بهینه شده‌ی GCM-ANN

برای ارزیابی شبکه‌ی بهینه شده‌ی GCM-ANN، از دو روش نمودار برگشتی و نمودار خطای باقیمانده استفاده گردید. در نمودار برگشتی که در شکل (۳-۶) آمده است مقادیر پیش‌بینی شده توسط شبکه‌ی بهینه برای دو سری ارزیابی و تست برحسب مقادیر تجربی رسم شده است. در نمودار خطای باقیمانده نیز که در شکل (۳-۷) آمده است خطای باقیمانده که تفاوت مقادیر پیش‌بینی شده توسط شبکه با مقادیر تجربی برای دو سری ارزیابی و تست است، بر حسب مقادیر تجربی رسم شده است همان‌طور که این نمودارها نشان می‌دهند شبکه بهینه شده‌ی GCM-ANN با ۱۷ توصیف‌کننده در پیش‌بینی دانسیته‌ی تعدادی از مخلوط‌ها در سری تست خطای فاحشی دارد و دارای انحراف است در حالیکه برای سری ارزیابی چنین مشکلی وجود ندارد. این نمودارها به خوبی چالش انتخاب درست سری تست در مخلوط‌ها را نشان می‌دهد و بیان می‌کند که شبکه کارایی لازم در پیش‌بینی دانسیته‌ی مخلوط‌های جدید را ندارد. لذا برای رفع این مشکل، ایده‌ی جدیدی مورد بررسی قرار گرفت و آن افزایش توصیف‌کننده‌ی جرم‌مولی کل برای هر مخلوط بود، زیرا با تغییر کسر مولی هر مخلوط با اجزای

معین حتی در دما و فشار ثابت، جرم‌مولی کل مخلوط تغییر می‌کند. آنچه در ادامه می‌آید بررسی افزایش توصیف‌کننده‌ی جرم‌مولی کل در بهبود یا عدم بهبود نتایج است.



شکل (۳-۶)-نمودار مقادیر پیش‌بینی شده‌ی دانسیته‌ی مخلوط‌ها برحسب مقادیر تجربی برای (الف) سری ارزیابی و (ب) سری تست



شکل (۳-۷)-نمودار خطای باقیمانده برحسب مقادیر تجربی دانسیته‌ی مخلوط‌ها (الف) سری ارزیابی و (ب) سری تست

۴-۱-۳- اثر افزایش توصیف‌کننده‌ی جرم‌مولی کل برای مدل‌سازی GCM-ANN

دانشیته‌ی مخلوط‌ها

برای محاسبه‌ی توصیف‌کننده‌ی جرم‌مولی کل، M_{mix} ، برای هر مخلوط نیز طبق روش ارائه شده در بخش (۳-۲) از مجموع وزن‌دار شده‌ی مولی جرم اجزای مخلوط طبق رابطه‌ی زیر استفاده گردید:

$$M_{total} = \sum_{i=1}^n x_i m_i \quad (۷-۳)$$

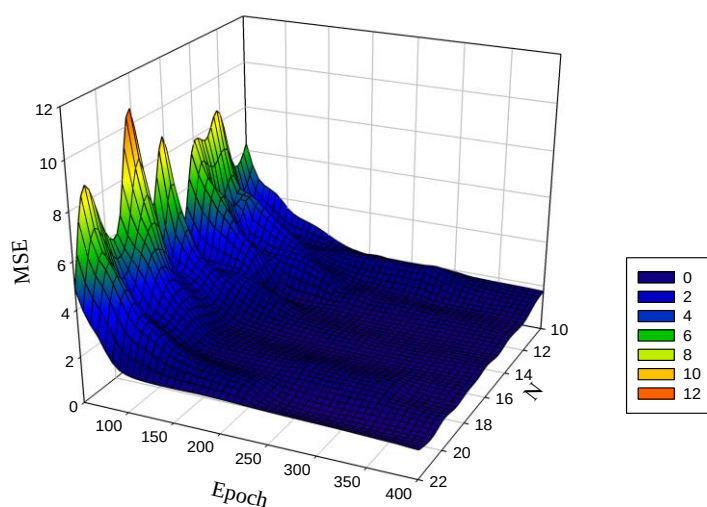
که در آن x_i و m_i به ترتیب کسر مولی و جرم جزء i ام در مخلوط است و n تعداد کل اجزای مخلوط را نشان می‌دهد. به عنوان مثال جرم‌مولی کل مخلوط دوتایی 0.8008 n-Decane + 0.1992 n-Hexane می‌شود:

$$M_{total} = (0.1992 \times 86.18) + (0.8008 \times 142.29) = 131.112$$

پس از افزودن توصیف‌کننده‌ی جرم‌مولی کل به ۱۷ توصیف‌کننده‌ی قبلی برای مخلوط‌ها، داده‌ها مثل بخش قبلی دسته‌بندی و پس از مرتب‌سازی CR به شبکه عصبی مصنوعی وارد شدند. در این بخش نیز شبکه‌ی عصبی مصنوعی یک شبکه‌ی سه لایه است که برای آموزش آن از دو الگوریتم آموزشی لونبرگ-مارکوارت و تنظیم بایزین استفاده شده است. به منظور به حداقل رساندن تابع، عملکرد MSE در سری ارزیابی، دوباره شبکه‌هایی با مقدار μ ثابت، تعداد ورودی ثابت ۱۸، تعداد متفاوت گره‌های لایه‌ی پنهان از ۱۰ تا ۲۳ با گام یک و تعداد دور آموزشی متغیر از ۵۰ تا ۴۰۰ با گام ۱۰ طراحی شدند. ترکیبی از این دو پارامتر که بتواند کمترین مقدار مربعات خطا (MSE) را برای سری ارزیابی ایجاد کند به عنوان پارامترهای بهینه انتخاب شدند. سپس بهینه‌سازی شبکه‌های طراحی شده به‌ازای همه ترکیبات‌های ممکن از دو پارامتر تعداد گره‌های لایه پنهان و تعداد دور آموزشی انجام گردید. نتایج حداقل مربعات خطا و تغییرات پارامترها در توابع شبکه حول نقطه‌ی بهینه در جداول (۳-۱۱) تا (۳-۱۴) و نتایج حاصل از کل محاسبات در نمودارهای سه بعدی (۳-۸) تا (۳-۱۱) گردآوری شده است.

جدول (۱۱-۳) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN در حضور جرم‌مولی کل با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت

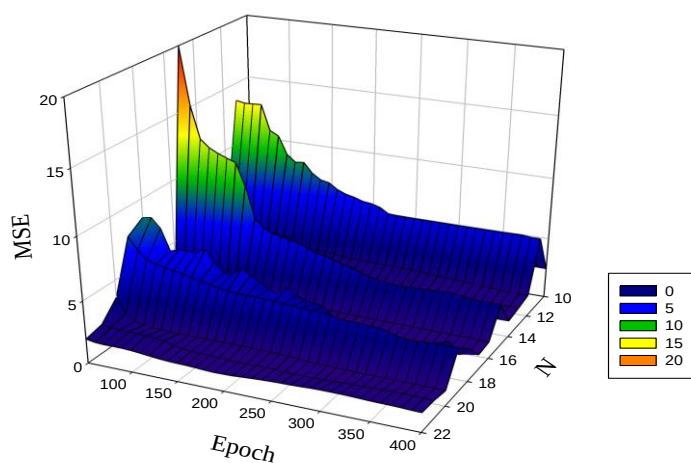
دور آموزشی	تعداد گره لایه پنهان									
		۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
	۲۰۰	۱/۴۵۸	۱/۵۰۳	۱/۵۷۴	۱/۳۸۸	۱/۵۷۱	۱/۷۱۶	۱/۶۵۳	۱/۹۴۲	۱/۵۰۲
	۲۱۰	۱/۴۹۲	۱/۴۵۱	۱/۵۴۸	۱/۳۶۰	۱/۵۴۸	۱/۷۱۱	۱/۶۳۱	۱/۸۷۸	۱/۴۸۹
	۲۲۰	۱/۴۶۴	۱/۴۱۳	۱/۵۲۱	۱/۳۴۹	۱/۵۲۹	۱/۷۰۸	۱/۶۰۵	۱/۸۱۳	۱/۴۷۳
	۲۳۰	۱/۴۲۳	۱/۳۹۰	۱/۴۹۳	۱/۳۴۲	۱/۵۲۱	۱/۷۰۶	۱/۵۷۲	۱/۵۵۹	۱/۴۵۶
	۲۴۰	۱/۳۹۲	۱/۳۷۳	۱/۴۷۰	۱/۳۳۲	۱/۵۱۷	۱/۷۶۷	۱/۵۳۷	۱/۳۳۵	۱/۴۳۸
	۲۵۰	۱/۳۹۰	۱/۳۵۷	۱/۴۵۵	۱/۳۲۲	۱/۵۱۰	۱/۷۸۹	۱/۵۰۶	۱/۲۱۶	۱/۴۲۳
	۲۶۰	۱/۳۸۷	۱/۳۴۳	۱/۴۴۵	۱/۳۱۴	۱/۴۹۷	۱/۶۹۱	۱/۴۸۳	۱/۱۶۵	۱/۴۱۱
	۲۷۰	۱/۳۸۱	۱/۳۳۱	۱/۴۳۵	۱/۳۰۶	۱/۴۷۹	۱/۶۷۵	۱/۴۶۸	۱/۱۳۷	۱/۴۰۰
	۲۸۰	۱/۳۸۱	۱/۳۲۱	۱/۳۵۵	۱/۳۰۱	۱/۴۶۲	۱/۶۶۰	۱/۴۵۸	۱/۱۰۸	۱/۳۹۲
	۲۹۰	۱/۳۸۱	۱/۳۱۲	۱/۳۴۴	۱/۲۹۷	۱/۴۴۸	۱/۶۴۷	۱/۴۵۱	۱/۰۹۳	۱/۳۸۴
	۳۰۰	۱/۳۸۱	۱/۳۰۲	۱/۳۳۸	۱/۲۹۳	۱/۴۳۷	۱/۶۳۵	۱/۴۴۵	۱/۰۸۸	۱/۳۷۷
	۳۱۰	۱/۳۸۲	۱/۲۹۲	۱/۳۳۶	۱/۲۸۹	۱/۴۳۰	۱/۶۲۳	۱/۴۴۱	۱/۰۸۴	۱/۳۶۹
	۳۲۰	۱/۳۸۲	۱/۲۸۱	۱/۳۳۹	۱/۲۸۵	۱/۴۲۴	۱/۶۱۰	۱/۴۳۷	۱/۰۷۵	۱/۳۶۰
	۳۳۰	۱/۳۸۴	۱/۲۷۴	۱/۳۴۴	۱/۲۸۱	۱/۴۲۰	۱/۵۹۷	۱/۴۳۴	۱/۰۷۴	۱/۳۴۹
	۳۴۰	۱/۳۸۶	۱/۲۷۰	۱/۳۴۹	۱/۲۷۸	۱/۳۷۳	۱/۴۷۶	۱/۴۲۳	۱/۰۷۲	۱/۳۳۵
	۳۵۰	۱/۳۹۰	۱/۲۶۸	۱/۳۵۲	۱/۲۷۶	۱/۳۴۹	۱/۲۹۳	۱/۳۷۹	۱/۰۷۳	۱/۳۲۰
	۳۶۰	۱/۳۹۰	۱/۲۶۷	۱/۳۵۴	۱/۲۷۵	۱/۳۲۷	۱/۲۲۲	۱/۳۸۵	۱/۰۷۳	۱/۳۰۶



شکل (۸-۳) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۱۱-۳)

جدول (۳-۱۲) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN در حضور جرم‌مولی کل با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید با تعداد گره و دور آموزشی متفاوت

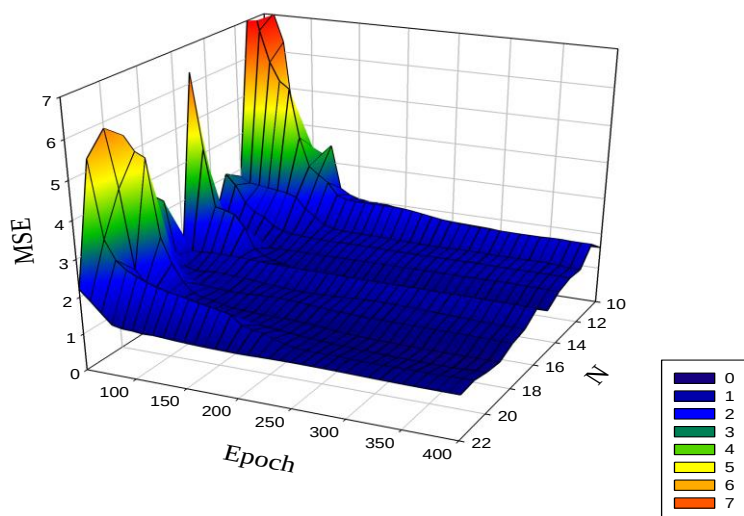
	تعداد گره لایه پنهان									
		۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
دور آموزشی	۲۰۰	۱/۷۹۱	۵/۱۷۷	۱/۶۲۴	۱/۴۵۳	۴/۱۳۸	۴/۶۵۲	۱/۵۷۴	۱/۳۵۰	۱/۳۶۹
	۲۱۰	۱/۷۰۳	۵/۰۷۱	۱/۶۰۱	۱/۴۳۲	۴/۸۰۵	۴/۶۱۸	۱/۵۶۳	۱/۳۸۲	۱/۳۸۵
	۲۲۰	۱/۶۴۵	۴/۷۹۴	۱/۵۶۷	۱/۴۱۰	۴/۳۱۵	۴/۵۵۱	۱/۵۴۴	۱/۳۸۲	۱/۴۴۲
	۲۳۰	۱/۵۹۸	۴/۵۹۴	۱/۵۴۳	۱/۳۸۷	۴/۰۸۰	۴/۵۲۸	۵/۱۸	۱/۳۷۹	۱/۴۸۶
	۲۴۰	۱/۵۴۹	۴/۴۵۸	۱/۵۳۱	۱/۳۶۳	۳/۹۶۰	۴/۴۷۳	۱/۴۹۸	۱/۳۶۹	۱/۵۳۰
	۲۵۰	۱/۵۱۳	۴/۱۵۹	۱/۵۱۸	۱/۳۳۳	۳/۸۸۱	۴/۴۳۷	۱/۴۳۳	۱/۳۴۱	۱/۵۶۱
	۲۶۰	۱/۴۸۲	۳/۹۴۷	۱/۵۰۳	۱/۳۰۶	۳/۰۱۶	۴/۴۰۷	۱/۳۶۴	۱/۳۱۸	۱/۵۸۱
	۲۷۰	۱/۴۶۳	۳/۸۰۳	۱/۴۹۰	۱/۲۹۱	۲/۹۴۶	۴/۳۲۳	۱/۳۶۴	۱/۳۰۸	۱/۵۹۸
	۲۸۰	۱/۴۴۹	۳/۶۶۵	۱/۴۷۸	۱/۲۷۹	۲/۷۷۲	۴/۱۵۳	۱/۳۳۹	۱/۳۱۴	۱/۶۲۴
	۲۹۰	۱/۴۴۳	۳/۶۴۳	۱/۴۶۷	۱/۲۵۰	۲/۵۵۰	۴/۰۹۹	۱/۳۰۵	۱/۳۳۲	۱/۶۶۱
	۳۰۰	۱/۴۳۱	۳/۶۲۰	۱/۴۵۸	۱/۲۲۳	۲/۲۷۵	۴/۰۹۳	۱/۲۹۸	۱/۳۵۷	۱/۶۸۰
	۳۱۰	۱/۳۹۷	۳/۶۵۱	۱/۴۴۹	۱/۲۱۲	۲/۰۷۷	۴/۰۶۶	۱/۲۹۴	۱/۳۸۶	۱/۶۸۱
	۳۲۰	۱/۳۸۶	۳/۶۶۲	۱/۴۳۸	۱/۲۱۱	۲/۳۳۷	۴/۰۳۶	۱/۲۹۱	۱/۴۱۴	۱/۶۷۹
	۳۳۰	۱/۳۴۷	۳/۷۱۲	۱/۴۲۶	۱/۲۰۹	۲/۳۴۱	۳/۸۳۴	۱/۲۸۴	۱/۴۴۰	۱/۶۶۸
	۳۴۰	۱/۳۴۸	۳/۷۴۹	۱/۴۱۶	۱/۲۰۹	۱/۸۰۷	۳/۶۲۵	۱/۲۷۳	۱/۴۶۷	۱/۶۴۲
	۳۵۰	۱/۳۴۵	۳/۷۷۰	۱/۴۱۰	۱/۲۰۹	۱/۴۳۲	۳/۵۲۶	۱/۲۶۴	۱/۴۸۸	۱/۶۱۶
	۳۶۰	۱/۳۴۰	۳/۶۳۵	۱/۴۰۶	۱/۲۱۰	۱/۸۶۰	۳/۵۵۶	۱/۲۵۶	۱/۴۹۹	۱/۶۰۳



شکل (۳-۹) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۲)

جدول (۳-۱۳) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN در حضور جرم‌مولی کل با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت

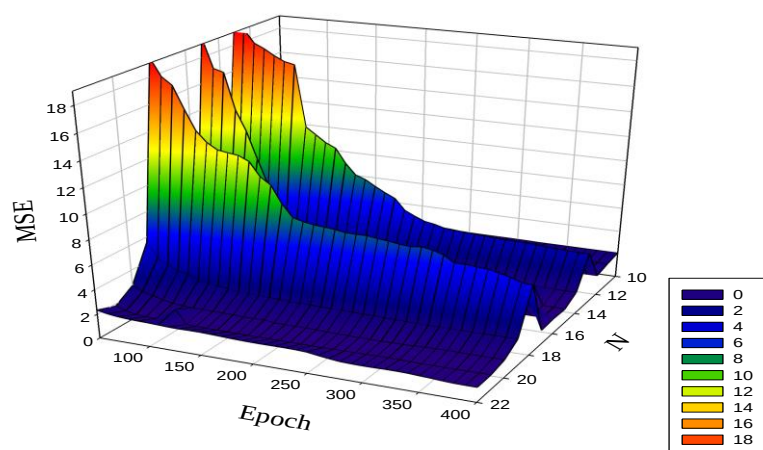
	تعداد گره لایه پنهان									
		۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
دور آموزشی	۲۰۰	۱/۴۴۷	۱/۲۰۵	۱/۶۱۴	۱/۳۳۸	۱/۳۸۶	۱/۲۳۶	۱/۲۹۹	۱/۵۳۷	۱/۲۲۲
	۲۱۰	۱/۴۵۱	۱/۱۹۹	۱/۶۱۲	۱/۳۴۱	۱/۳۶۲	۱/۲۲۵	۱/۲۷۰	۱/۴۸۹	۱/۲۲۵
	۲۲۰	۱/۴۵۱	۱/۱۹۵	۱/۶۱۰	۱/۳۴۸	۱/۳۳۷	۱/۲۰۸	۱/۲۴۳	۱/۳۹۶	۱/۲۳۲
	۲۳۰	۱/۴۵۳	۱/۱۹۳	۱/۶۰۸	۱/۳۶۵	۱/۳۱۷	۱/۱۸۶	۱/۲۱۷	۱/۳۰۳	۱/۲۳۷
	۲۴۰	۱/۴۵۴	۱/۱۹۱	۱/۶۰۶	۱/۳۸۴	۱/۳۰۲	۱/۱۶۴	۱/۱۹۲	۱/۲۵۲	۱/۲۴۱
	۲۵۰	۱/۴۵۵	۱/۱۹۱	۱/۶۰۵	۱/۳۹۸	۱/۲۹۲	۱/۱۳۶	۱/۱۷۱	۱/۲۲۹	۱/۲۴۳
	۲۶۰	۱/۴۵۶	۱/۱۹۱	۱/۶۰۴	۱/۴۰۶	۱/۲۸۵	۱/۱۱۱	۱/۱۵۳	۱/۲۲۸	۱/۲۴۲
	۲۷۰	۱/۴۵۸	۱/۱۹۳	۱/۶۰۳	۱/۴۰۷	۱/۲۷۹	۱/۰۹۶	۱/۱۳۸	۱/۲۲۷	۱/۲۴۰
	۲۸۰	۱/۴۵۹	۱/۱۸۶	۱/۶۰۲	۱/۴۰۴	۱/۲۷۴	۱/۰۸۶	۱/۱۲۶	۱/۲۲۴	۱/۲۳۷
	۲۹۰	۱/۴۶۰	۱/۱۹۱	۱/۶۰۲	۱/۳۹۷	۱/۲۷۰	۱/۰۷۸	۱/۱۱۵	۱/۲۱۶	۱/۲۳۴
	۳۰۰	۱/۴۶۱	۱/۱۹۳	۱/۶۰۱	۱/۳۸۷	۱/۲۶۶	۱/۰۷۲	۱/۱۰۶	۱/۲۰۷	۱/۲۲۹
	۳۱۰	۱/۴۶۲	۱/۱۹۴	۱/۵۹۹	۱/۳۷۶	۱/۲۶۴	۱/۰۶۷	۱/۰۹۹	۱/۲۰۴	۱/۲۲۳
	۳۲۰	۱/۴۶۳	۱/۱۹۴	۱/۵۹۸	۱/۳۶۴	۱/۲۶۱	۱/۰۶۳	۱/۰۹۱	۱/۲۰۴	۱/۲۱۸
	۳۳۰	۱/۴۶۴	۱/۱۹۴	۱/۵۹۶	۱/۳۵۲	۱/۲۶۰	۱/۰۵۹	۱/۰۸۴	۱/۲۰۵	۱/۲۱۳
	۳۴۰	۱/۴۷۱	۱/۱۹۶	۱/۵۹۴	۱/۳۴۱	۱/۲۵۷	۱/۰۵۴	۱/۰۷۷	۱/۲۰۶	۱/۲۰۹
	۳۵۰	۱/۴۷۶	۱/۱۹۷	۱/۵۹۲	۱/۳۳۰	۱/۲۵۵	۱/۰۴۸	۱/۰۷۲	۱/۲۲۳	۱/۲۰۶
	۳۶۰	۱/۴۷۲	۱/۲۰۰	۱/۵۸۹	۱/۳۲۲	۱/۲۵۱	۱/۰۴۹	۱/۰۷۲	۱/۲۰۸	۱/۲۰۴



شکل (۳-۱۰) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۳)

جدول (۳-۱۴) - مقادیر MSE سری ارزیابی برای شبکه‌ی GCM-ANN در حضور جرم‌مولی کل با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال تانژانت سیگموئید با تعداد گره و دور آموزشی متفاوت

		تعداد گره لایه پنهان								
دور آموزشی		۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	۲۱	۲۲
	۲۰۰	۲/۰۴۷	۳/۷۴۰	۱/۳۲۴	۳/۱۱۹	۸/۰۱۰	۲/۳۸۴	۱/۴۵۹	۱/۱۶۹	۱/۷۶۸
	۲۱۰	۲/۰۴۴	۳/۶۹۳	۱/۳۰۷	۳/۱۲۴	۷/۹۴۷	۲/۳۷۶	۱/۴۵۷	۱/۱۶۷	۱/۷۶۵
	۲۲۰	۲/۰۵۵	۳/۵۶۴	۱/۲۹۹	۳/۱۲۳	۷/۸۶۹	۲/۳۶۳	۱/۴۵۵	۱/۱۶۵	۱/۷۷۱
	۲۳۰	۱/۹۴۰	۳/۴۵۲	۱/۲۹۵	۳/۱۲۶	۷/۷۳۴	۲/۳۵۳	۱/۴۵۲	۱/۱۶۳	۱/۷۷۳
	۲۴۰	۱/۹۰۷	۳/۳۴۸	۱/۲۹۴	۳/۱۲۱	۷/۷۱۴	۲/۳۴۴	۱/۴۱۸	۱/۱۶۱	۱/۷۶۳
	۲۵۰	۱/۸۶۲	۳/۱۸۴	۱/۲۹۵	۳/۱۱۷	۷/۷۳۳	۲/۳۲۲	۱/۳۹۰	۱/۱۶۱	۱/۷۲۴
	۲۶۰	۱/۸۳۲	۲/۹۱۸	۱/۲۹۶	۳/۱۰۸	۷/۶۱۰	۲/۲۹۷	۱/۳۸۱	۱/۱۸۰	۱/۶۰۳
	۲۷۰	۱/۸۰۳	۲/۵۰۱	۱/۲۹۶	۳/۱۱۰	۷/۵۳۳	۲/۲۹۴	۱/۳۷۲	۱/۱۹۱	۱/۴۹۵
	۲۸۰	۱/۸۰۱	۲/۱۵۳	۱/۲۹۱	۳/۱۲۲	۷/۴۱۱	۲/۲۶۵	۱/۳۶۸	۱/۱۸۹	۱/۴۰۳
	۲۹۰	۱/۷۸۵	۲/۲۲۳	۱/۲۸۷	۳/۱۳۳	۷/۳۳۸	۲/۲۵۲	۱/۳۶۷	۱/۲۰۳	۱/۳۷۴
	۳۰۰	۱/۷۷۸	۲/۱۸۵	۱/۲۸۳	۳/۱۳۳	۷/۴۴۳	۲/۲۳۵	۱/۳۷۳	۱/۲۱۳	۱/۳۶۴
	۳۱۰	۱/۷۷۷	۲/۱۳۸	۱/۲۸۰	۳/۱۲۶	۷/۳۰۸	۲/۲۱۳	۱/۳۹۱	۱/۲۱۲	۱/۳۷۰
	۳۲۰	۱/۷۸۳	۲/۱۴۱	۱/۲۷۷	۳/۱۱۶	۷/۰۴۰	۲/۲۰۷	۱/۴۲۱	۱/۲۰۵	۱/۲۸۹
	۳۳۰	۱/۷۸۲	۲/۱۳۱	۱/۲۷۵	۳/۱۱۱	۶/۵۳۵	۲/۲۱۶	۱/۴۵۰	۱/۱۹۴	۱/۳۹۴
	۳۴۰	۱/۸۰۱	۲/۱۳۳	۱/۲۷۳	۳/۱۱۶	۶/۵۴۰	۲/۲۱۶	۱/۴۷۲	۱/۱۸۶	۱/۳۷۹
	۳۵۰	۱/۸۰۹	۲/۲۳۶	۱/۲۷۱	۳/۱۲۱	۶/۴۷۵	۲/۲۱۴	۱/۴۸۷	۱/۱۸۳	۱/۳۲۹
	۳۶۰	۱/۸۱۴	۲/۳۶۳	۱/۲۶۹	۳/۱۳۲	۶/۴۱۸	۲/۲۲۰	۱/۴۷۵	۱/۱۸۶	۱/۲۹۲



شکل (۳-۱۱) - نمودار MSE بر حسب تعداد گره و دور آموزشی برای کل محاسبات جدول (۳-۱۴)

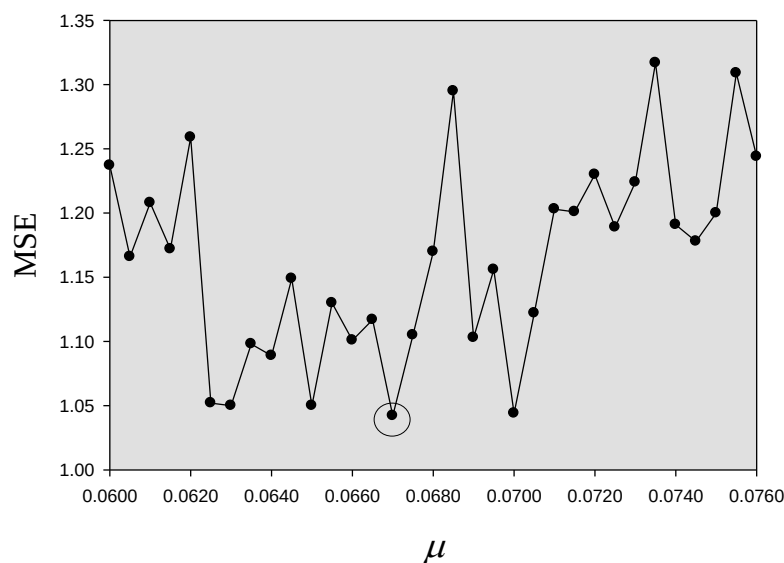
با توجه به نتایج موجود در جداول (۱۱-۳) تا (۱۴-۳) مقادیر بهینه‌ی پارامترهای هر یک از شبکه‌های GCM-ANN طراحی شده در حضور توصیف‌کننده‌ی جرم‌مولی کل مخلوط در جدول (۱۵-۳) آورده شده است.

جدول (۱۵-۳)- توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل مخلوط

تابع انتقال	تابع آموزش	تعداد توصیف کننده	تعداد نرون لایه پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموئید	تنظیم بایزین	۱۸	۲۱	۳۴۰	۱/۰۷۲
لگاریتم سیگموئید	لونبرگ-مارکوارت	۱۸	۱۷	۳۴۰	۱/۲۰۹
تانژانت سیگموئید	تنظیم بایزین	۱۸	۱۹	۳۵۰	۱/۰۴۸
تانژانت سیگموئید	لونبرگ-مارکوارت	۱۸	۲۱	۲۵۰	۱/۱۶۱

نتایج موجود در این جدول نشان می‌دهد که شبکه‌ی با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئید با ۱۹ گره در لایه‌ی پنهان و دور آموزشی ۳۵۰ مقدار MSE کمتری نسبت به سایر شبکه‌ها دارد لذا این شبکه به عنوان شبکه‌ی بهینه انتخاب گردید.

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، مقدار بهینه‌ی μ ، ۰/۰۶۷۰ به دست آمد. نمودار مقدار میانگین مربعات خطا بر حسب μ حول نقطه‌ی بهینه نیز در شکل (۱۲-۳) رسم شده است. توابع و مقادیر پارامترهای بهینه شده‌ی GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل مخلوط برای پیش‌بینی دانسیته‌ی مخلوط‌ها، در جدول (۱۶-۳) ارائه شده است.



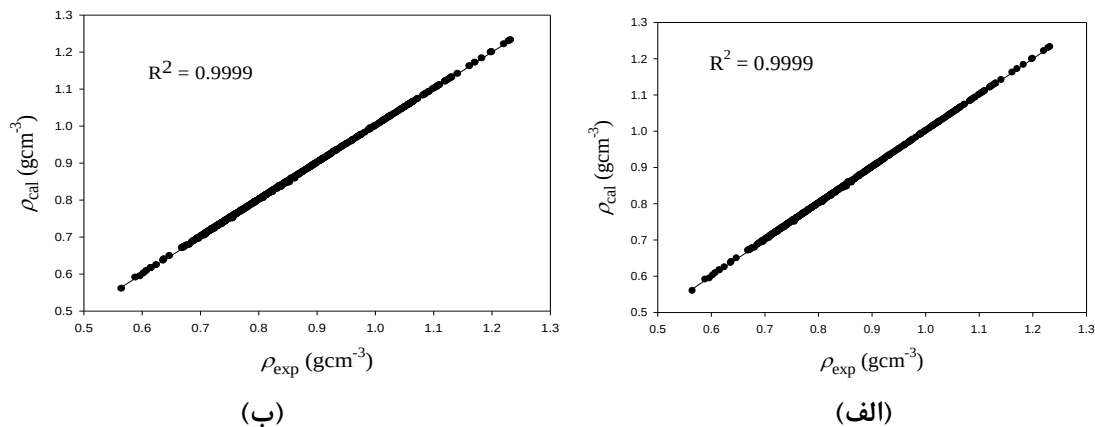
شکل (۳-۱۲) - نمودار میانگین مربعات خطا بر حسب μ حول نقطه بهینه برای سری ارزیابی شبکه‌ی GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل

جدول (۳-۱۶) - توابع و مقادیر پارامترهای مدل بهینه شده‌ی GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل

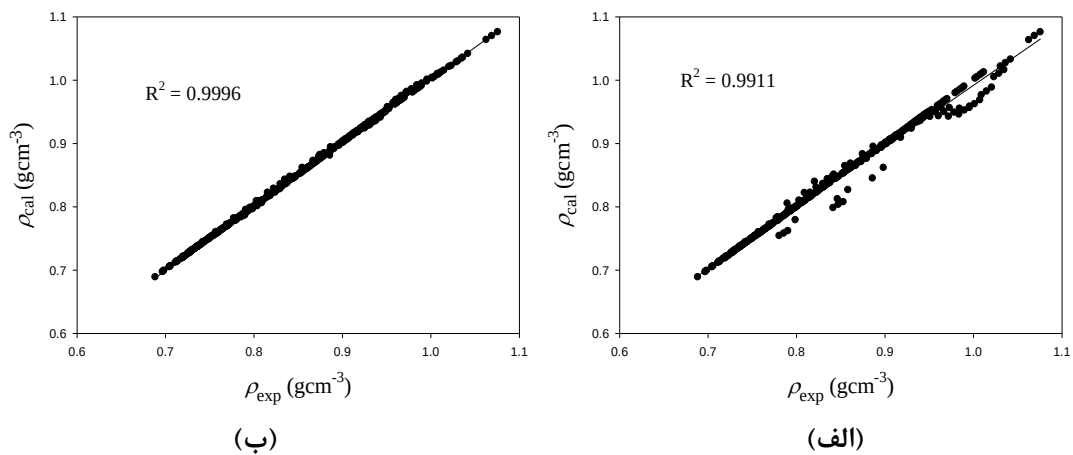
تابع آموزش	Train br
تابع انتقال لایه‌ی پنهان	Tan sig
تابع انتقال لایه‌ی خروجی	pure line
تعداد نرون لایه‌ی پنهان	۱۹
تعداد متغیرهای ورودی	۱۸
تعداد دوره‌های آموزش	۳۵۰
پارامتر μ	۰/۰۶۷
مقدار MSE سری ارزیابی	۱/۰۴۳

اکنون باید قدرت پیش‌بینی GCM-ANN بهینه شده در حضور توصیف‌کننده‌ی جرم‌مولی کل را برای سری‌های ارزیابی و تست بررسی شود. برای این منظور دوباره نمودارهای برگشتی و خطای باقیمانده برای این دو سری داده ترسیم گردید که این نمودارها در شکل‌های (۳-۱۳) تا (۳-۱۶) نشان داده

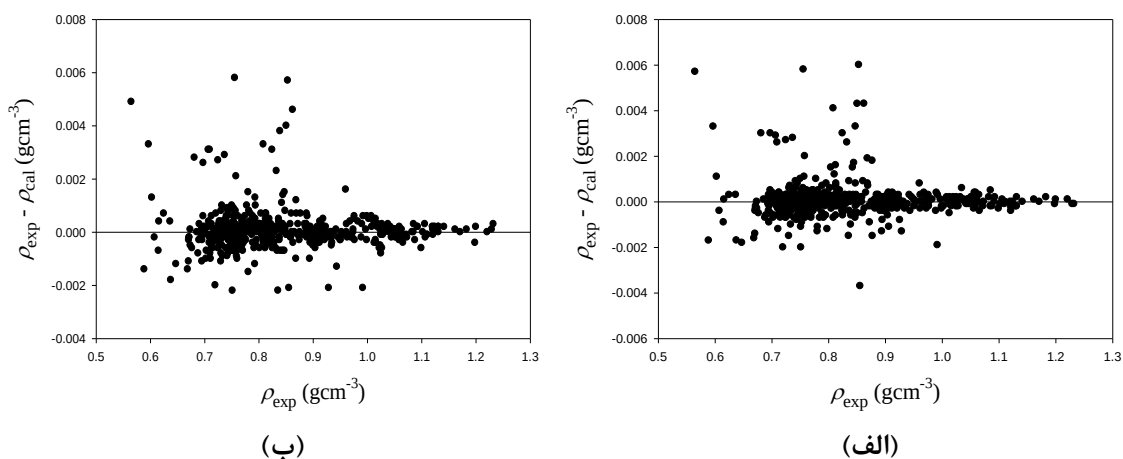
شده‌اند. برای آنکه اثر توصیف‌کننده‌ی جرم‌مولی کل مشخص شود در همه‌ی نمودارها اثر حضور و عدم حضور توصیف‌کننده جرم‌مولی کل در دو نمودار کنار هم نشان داده شده است.



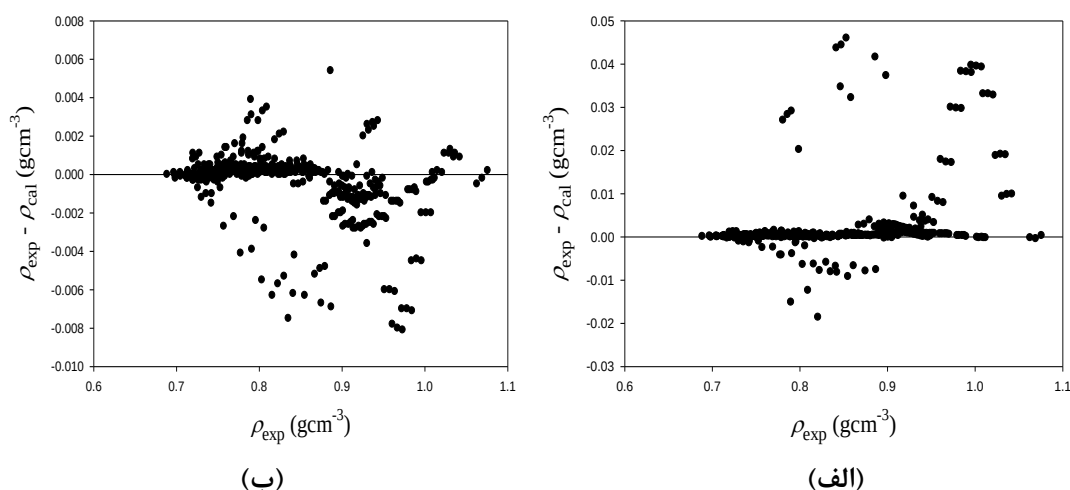
شکل (۳-۱۳) - نمودار برگشتی مقادیر پیش‌بینی شده دانسیته برحسب مقادیر تجربی برای سری ارزیابی توسط مدل GCM-ANN بهینه (الف) بدون جرم‌مولی کل و (ب) درحضور جرم‌مولی کل



شکل (۳-۱۴) - نمودار برگشتی مقادیر پیش‌بینی شده دانسیته برحسب مقادیر تجربی برای سری تست توسط مدل GCM-ANN بهینه (الف) بدون جرم‌مولی کل و (ب) درحضور جرم‌مولی کل



شکل (۳-۱۵)- نمودار مقادیر خطای باقیمانده دانسیته برحسب مقادیر تجربی برای سری ارزیابی توسط مدل GCM-ANN بهینه شده (الف) بدون جرم‌مولی کل و (ب) درحضور جرم‌مولی کل



شکل (۳-۱۶)- نمودار مقادیر خطای باقیمانده دانسیته برحسب مقادیر تجربی برای تست توسط مدل GCM-ANN بهینه شده (الف) بدون جرم‌مولی کل و (ب) درحضور جرم‌مولی کل

مقایسه نمودارهای (۳-۱۳) تا (۳-۱۶) یکبار دیگر اهمیت انتخاب سری تست به روش ردترکیبات را نشان می‌دهد. شکل (۳-۱۳) نمودار برگشت برای سری ارزیابی مدل GCM-ANN را در حضور و عدم حضور توصیف‌کننده جرم‌مولی کل نشان می‌دهد که چندان تفاوتی بین R^2 دو نمودار مشاهده نمی‌شود. لازم به ذکر است که سری ارزیابی بر اساس ردمخلوط‌ها انتخاب شده بودند، اما با نگاه

دقیق‌تر به نمودار (۳-۱۴) که نمودار برگشت برای سری تست در دو شبکه‌ی فوق است به‌خوبی می‌توان مشاهده نمود که افزایش توصیف‌کننده‌ی جرم‌مولی کل مخلوط که با تغییر کسرمولی تغییر می‌کند مقدار R^2 را از ۰/۹۹۱۱ به ۰/۹۹۹۶ افزایش داده است. به عبارت دیگر شبکه‌ی GCM-ANN بهینه شده در حضور توصیف‌کننده‌ی جرم‌مولی کل می‌تواند به خوبی دانسیته‌ی مخلوط‌های جدیدی که در سری آموزش حضور نداشتند را پیش‌بینی کند.

این نتایج به شکل دیگری در نمودارهای (۳-۱۵) و (۳-۱۶) تایید می‌شود. بازه‌ی خطای مطلق در پیش‌بینی دانسیته برای سری ارزیابی با دو شبکه‌ی GCM-ANN در حضور و عدم حضور توصیف‌کننده‌ی جرم‌مولی کل تقریباً یکسان است اما این تفاوت در بازه‌ی خطای مطلق در سری تست به‌خوبی نمایان است، بازه‌ی خطای مطلق از ± 0.05 برای شبکه‌ی GCM-ANN بدون جرم‌مولی کل به ± 0.008 برای شبکه‌ی GCM-ANN در حضور جرم‌مولی کل کاهش یافته است. این نتایج اهمیت حضور توصیف‌کننده‌ی جرم‌مولی کل مخلوط را در آموزش شبکه‌ی GCM-ANN نشان می‌دهد و به همین دلیل در بخش بعدی جرم‌مولی کل مخلوط از ابتدا به عنوان یک توصیف‌کننده‌ی خوب در نظر گرفته شد.

۲-۴-۳- مدل سازی شبکه‌ی عصبی مصنوعی با استفاده از توصیف‌کننده‌های تئوری

۱-۲-۴-۳- محاسبه‌ی توصیف‌کننده‌های تئوری برای مخلوط‌ها

همانطور که قبلاً توضیح داده شد برای محاسبه‌ی توصیف‌کننده‌های تئوری برای مخلوط‌ها ابتدا باید این توصیف‌کننده‌ها را برای اجزای مخلوط‌ها محاسبه نمود و سپس با استفاده از روش‌های گفته شده در بخش (۲-۳) مقدار این توصیف‌کننده‌ها را برای مخلوط‌ها به دست آورد. بنابراین ابتدا ساختار مولکول‌ها با نرم‌افزار هایپرکم رسم گردید، سپس با استفاده از روش نیمه تجربی AM1 ساختار سه‌بعدی ترکیبات بهینه شد. در این مرحله ۱۹ توصیف‌کننده برای هر مولکول به دست آمد. در ادامه ساختار بهینه شده‌ی مولکول‌ها به عنوان ورودی به نرم‌افزار دراگون داده شد و برای هر مولکول ۱۴۸۱ توصیف‌کننده‌ی تئوری به دست آمد. پس از محاسبه‌ی ۱۵۰۰ توصیف‌کننده‌ی تئوری برای اجزای مخلوط‌ها، مقدار توصیف‌کننده‌های تئوری برای مخلوط‌های مورد مطالعه با استفاده از کسرمولی آن‌ها و رابطه‌ی (۳-۴) به دست آمدند.

ماتریس نهایی که بعد از محاسبات بالا بدست می‌آید شامل ۱۵۰۰ توصیف‌کننده است، با توجه به این که تعداد زیاد توصیف‌کننده‌ها 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 ۲۲۳ توصیف‌کننده از مجموع ۱۵۰۰ توصیف‌کننده‌ی تئوری باقی ماندند. در این مرحله دو متغیر تجربی دما، فشار و یک توصیف‌کننده‌ی جرم‌مولی کل نیز به ماتریس ورودی اضافه شدند.

هدف بعدی انتخاب بهترین توصیف‌کننده‌ها از بین توصیف‌کننده‌های باقیمانده برای ساخت مدل است. در این پایان‌نامه از روش غیرخطی الگوریتم ژنتیک برای انتخاب توصیف‌کننده‌های مناسب استفاده شد که در ادامه در مورد آن توضیح داده خواهد شد.

۲-۲-۴-۳- انتخاب بهترین توصیف‌کننده‌ها به روش الگوریتم ژنتیک

به منظور اجرای برنامه‌ی الگوریتم ژنتیک، تمام توصیف‌کننده‌های باقیمانده به همراه سه توصیف‌کننده‌ی دما، فشار و جرم‌مولی کل به عنوان ورودی به این برنامه داده شدند. برنامه M-file استفاده شده در این تحقیق در بسته‌ی نرم افزاری MATLAB دارای مقادیر پیش فرض احتمال ترکیب ۰/۵، احتمال جهش ۰/۰۱، اندازه‌ی جمعیتی برابر ۳۰ و با تکرار ۵۰۰ می‌باشد. این برنامه ۱۶ بار تکرار گردید تا احتمال انتخاب توصیف‌کننده‌های تصادفی به حداقل برسد. توصیف‌کننده‌هایی که بیشتر از ۵۰٪ تکرار شده‌اند به همراه تعداد دفعات تکرار در جدول (۳-۱۷) نشان داده شده‌اند. طبقه و نام کامل آنها نیز در جدول (۳-۱۸) آورده شده است. همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۳-۱۹) ارائه شده که این ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول (۳-۱۷)- تعداد دفعات تکرار توصیف‌کننده‌ها در روش انتخابی الگوریتم ژنتیک پس از ۱۶ بار اجرا

توصیف کننده	X5A	H047	R2m-A	JGI8	MATS1m	G2v	Mor30u	HATS5m
تعداد تکرار	۱۰	۱۲	۱۰	۱۶	۱۶	۱۵	۱۲	۱۵
توصیف کننده	RTp	Mor29e	Ds	Mor11u	C001	Ss	Mor20m	
تعداد تکرار	۱۰	۱۲	۱۲	۱۳	۱۲	۱۰	۱۰	
توصیف کننده	Mor10p	RDF155u	GATS5m	Mor16m	RDF130v	Mor15m	R1m-A	
تعداد تکرار	۱۱	۱۱	۹	۸	۱۳	۹	۱۲	

جدول (۳-۱۸) طبقه و نام کامل توصیف کننده‌های تئوری انتخابی توسط الگوریتم ژنتیک

شماره	نشان	طبقه	نام کامل
۱	JGI8	Galvez topol charge indices	Mean topological charge index of order8
۲	MATS1m	2D autocorrelations	Moran autocorrelation log1 Weighted by atomic masses
۳	G2v	WHIM descriptors	2st component symmetry directional WHIM index weighted by atomic van der waals volumes
۴	Mor30u	3D MoRSE descriptors	3D MoRSE single30 unweighted
۵	HATS5m	GETAWAY descriptors	Leverage weighted autocorrelation of leg5 weighted by atomic masses
۶	RDF130v	RDF descriptors	Radial Distribution Function13.0 Weighted by atomic van der Waals volumes
۷	RTp	GETAWAY descriptors	R total index weighted by atomic polarizabilities
۸	Mor29e	3D MoRSE descriptors	3D MoRSE single29 weighted by atomic sanderson electronegativities
۹	Ds	WHIM descriptors	D total accessibility index weighted by atomic electrotopological states
۱۰	Mor11u	3D MoRSE descriptors	3D MoRSE single11 unweighted
۱۱	C001	Atom centrad fragments	CH ₃ R CH ₄
۱۲	Ss	Constitiutional descriptors	Sum of kier-Hall electrotopological states
۱۳	Mor20m	3D MoRSE descriptors	3D MoRSE single20 weighted by atomic masses
۱۴	R2m-A	GETAWAY descriptors	R maximal autocorrelation of leg2 weighted by atomic masses
۱۵	X5A	Topological descriptors	Average connectivity index chi5
۱۶	Mor10p	3D MoRSE descriptors	3D MoRSE single10 weighted by atomic polarizabilities
۱۷	RDF155u	RDF descriptors	Radial Distribution Function15.5 unweighted
۱۸	GATS5m	2D autocorrelations	Geary autocorrelation leg5 weighted by atomic masses
۱۹	Mor16m	3D MoRSE descriptors	3D MoRSE single16 weighted by atomic masses
۲۰	H047	Atom centrad fragments	H attached to C1(sp ³) / C0(sp ²)
۲۱	Mor15m	3D MoRSE descriptors	3D MoRSE single15 weighted by atomic masses
۲۲	R1m-A	GETAWAY descriptors	R maximal autocorrelation of leg1 weighted by atomic masses

جدول (۳-۱۹) همبستگی بین توصیف کننده‌های انتخابی الگوریتم ژنتیک

	temperature	pressure	mass	JGI8	MATS1m	G2v	Mor30u	HATS5m	RDF130v	RTp	Mor29e	Ds
temperature	۱											
pressure	۰/۲۴۳	۱										
mass	۰/۳۱۵	۰/۰۹۸	۱									
JGI8	۰/۴۳۴	۰/۱۳۷	۰/۶۸۵	۱								
MATS1m	-۰/۴۳۱	-۰/۲۵۷	-۰/۲۱۳	-۰/۳۸۳	۱							
G2v	۰/۳۲۷	۰/۰۴	۰/۴۷۲	۰/۵۶۶	-۰/۳۰۱	۱						
Mor30u	-۰/۴۲۵	۰/۳۱	۰/۵۴۱	۰/۴۱۶	-۰/۷۳۷	۰/۴۰۲	۱					
HATS5m	-۰/۱۴۶	-۰/۰۹۲	-۰/۰۵۲	-۰/۱۰۴	۰/۴۰۸	۰/۸۱	-۰/۲	۱				
RDF130v	۰/۲۳۲	۰/۰۱۷	۰/۳۴۱	۰/۷۸۸	-۰/۲۵۴	۰/۳۶۵	۰/۲۰۵	-۰/۰۶	۱			
RTp	-۰/۴۳۱	-۰/۲۴۲	-۰/۵۷۵	-۰/۶۱۷	۰/۸۱۲	-۰/۴۵۸	-۰/۶۸۵	۰/۲۶۲	۰/۳۵۵	۱		
Mor29e	-۰/۴۰۶	-۰/۰۹۴	-۰/۶۶۱	-۰/۵۶۵	۰/۵۱۱	-۰/۶۲۳	-۰/۷۳۳	۰/۱۴۳	-۰/۲۱۵	۰/۵۷۴	۱	
Ds	-۰/۳۶۲	-۰/۲۲۶	-۰/۰۳۶	-۰/۲۳۶	۰/۸۸۳	-۰/۱۳۷	-۰/۶۳۱	۰/۴۶۵	-۰/۱۴۹	۰/۶۴۷	۰/۴۱۵	۱
Mor11u	-۰/۰۴۶	-۰/۰۸۲	-۰/۰۴۴	۰/۰۸۸	۰/۳۱۵	-۰/۰۵۵	-۰/۴۸۲	-۰/۰۳۲	-۰/۰۷۶	۰/۰۳۴	۰/۱۲	۰/۱۶۵
C001	۰/۲۵۶	۰/۱۵۷	۰/۱۷۳	۰/۲۸۱	-۰/۵۵۷	۰/۰	۰/۵۱۹	-۰/۲۵۵	-۰/۲۰۹	-۰/۳۷۷	-۰/۲۵۶	-۰/۵۳۶
Ss	۰/۱۰۴	۰/۰۱	۰/۷۲۹	۰/۵۰۲	۰/۱۱۷	۰/۱۴۴	۰/۱۱۷	۰/۰۵۳	۰/۲۲۴	-۰/۴	-۰/۲۵۳	۰/۱۴۵
Mor20m	-۰/۴۱۶	-۰/۰۹۷	-۰/۴۶۷	-۰/۵۴۴	۰/۷۷	-۰/۵۳۳	-۰/۶۵۹	۰/۴۰۸	-۰/۲۸۶	۰/۷۵۱	۰/۷۹۹	۰/۶۱۱
R2m-A	-۰/۲۸۵	-۰/۲۴۸	-۰/۰۰۴	-۰/۲۰۳	۰/۷۶۵	-۰/۰۰۳	-۰/۴۰۲	۰/۲۸۸	-۰/۱۷۲	۰/۵۶۶	۰/۰۷۶	۰/۷۴۸
X5A	۰/۲۲۷	۰/۱۲۸	۰/۳۹۹	۰/۲۷۹	-۰/۳۷۹	۰/۵۰۶	۰/۵۰۲	۰/۲۱۹	۰/۲۲	-۰/۴۵۸	-۰/۳۴۳	-۰/۱۵۹
Mor10p	۰/۳۳۱	۰/۱۲۲	۰/۶۶۳	۰/۵۱۲	-۰/۴۳۱	۰/۵۷	۰/۵۹	-۰/۱۰۱	۰/۲۲۷	-۰/۵۵۲	-۰/۷۸۷	-۰/۲۴۴
RDF155u	۰/۰۲۶	۰/۰۹۵	۰/۱۷۳	۰/۲۷۶	-۰/۰۴۵	۰/۱۱۷	۰/۰۵۵	-۰/۰۰۹	۰/۱۱۸	-۰/۱۳۱	۰/۰۷۸	۰/۰۰۶
GATS5m	-۰/۰۴۸	-۰/۰۲۵	۰/۱۴۱	۰/۰۱۲	۰/۰۵۹	۰/۰۲۷	۰/۱۸۸	۰/۳۳۲	۰/۰۳۱	۰/۱۲۸	۰/۰۴۲	۰/۱۹۹
Mor16m	-۰/۴	-۰/۲۱۱	-۰/۶۳۵	-۰/۵۵۵	۰/۷۱۹	-۰/۳۳۶	-۰/۷۶۵	۰/۳۹۹	-۰/۲۷	۰/۸۷۷	۰/۶۹۷	۰/۶۰۱
H047	-۰/۳۵	-۰/۰۸۴	-۰/۲۰۴	-۰/۳۶۹	۰/۷۱۷	-۰/۴۱۶	-۰/۵۸۹	۰/۳۷۸	-۰/۲۳	۰/۵۰۲	۰/۶۸۹	۰/۵۵۶
Mor15m	۰/۲۷۸	۰/۳۷۵	۰/۵۲۸	۰/۴۵۶	-۰/۴۲۵	۰/۰۸۷	۰/۴۵۲	-۰/۱۹۲	۰/۲۴۱	-۰/۶۷۱	-۰/۲۳۷	-۰/۴۳۶
R1m-A	-۰/۳۰۴	-۰/۲۴۶	۰/۰۸۶	-۰/۲۰۳	۰/۸۰۳	۰/۰۵۳	-۰/۳۹۱	۰/۴۹۳	-۰/۱۶۴	۰/۵۶۱	۰/۱۰۸	۰/۸۵۹

ادامه جدول (۳-۱۹) - همبستگی بین توصیف‌کننده‌های انتخابی الگوریتم ژنتیک

	Mor11u	C001	Ss	Mor20m	R2m-A	X5A	Mor10p	RDF155u	GATS5m	Mor16m	H047	Mor15m	R1m-A
temperature													
pressure													
mass													
JGI8													
MATS1m													
G2v													
Mor30u													
HATS5m													
RDF130v													
RTp													
Mor29e													
Ds													
Mor11u	۱												
C001	-۰/۴۲	۱											
Ss	۰/۳۷۴	۰/۰۶۵	۱										
Mor20m	-۰/۱۹	-۰/۳۷۷	-۰/۰۸۷	۱									
R2m-A	۰/۲۰۶	-۰/۴۴۹	۰/۱۸	۰/۳۶۲	۱								
X5A	-۰/۳۷۶	۰/۰۳۳	۰/۱۲۱	-۰/۲۶۵	-۰/۲۱۹	۱							
Mor10p	-۰/۱۹۸	-۰/۰۲	۰/۲۷۴	-۰/۶۶	-۰/۰۲۴	۰/۵۲۹	۱						
RDF155u	۰/۰۳۸	۰/۰۴	۰/۱۴۴	-۰/۰۸۷	۰/۰۲۲	۰/۰۶۳	۰/۱۰۸	۱					
GATS5m	-۰/۴۷۶	۰/۲۲۵	-۰/۰۶۱	۰/۱۳۳	-۰/۱۲۴	۰/۲۳۵	-۰/۰۰۸	۰/۰۱۴	۱				
Mor16m	۰/۱	-۰/۳۹۸	-۰/۴۵۹	۰/۷۳۹	۰/۴۲۱	-۰/۳۳۶	-۰/۶۴۵	-۰/۰۷۲	۰/۰۴۷	۱			
H047	۰/۳۳۷	-۰/۵۴۷	۰/۱۸۵	۰/۸۳	۰/۳۲۵	-۰/۲۷۶	-۰/۵۰۸	-۰/۰۷۱	-۰/۰۹	۰/۴۸۹	۱		
Mor15m	۰/۱۴۹	۰/۲۴۸	۰/۶۰۶	-۰/۱۸۱	-۰/۴۶۵	۰/۲۹۲	۰/۳۲۶	۰/۰۷۸	-۰/۰۸۵	-۰/۷۰۱	۰/۷۹	۱	
R1m-A	۰/۱۲۱	-۰/۶۱۲	۰/۱۷۱	۰/۴۵۲	۰/۸۸۶	-۰/۰۰۵	۰/۰۲۲	۰/۰۲۱	۰/۱۴۵	۰/۴۵	۰/۴۳۶	-۰/۴۳۶	۱

۳-۲-۴-۳- بهینه‌سازی شبکه‌ی عصبی مصنوعی با توصیف‌کننده‌های الگوریتم ژنتیک

شبکه‌ی عصبی مورد مطالعه در این بخش، شبکه‌ای سه لایه است. در این مرحله نیز شبکه‌هایی از تلفیق دو الگوریتم آموزشی متفاوت لونبرگ-مارکوات (trainlm) و تنظیم بایزین (trainbr) و دو تابع انتقال لگاریتم سیگموئید (logsig) و تانژانت سیگموئید (tansig) طراحی شدند و سپس در هریک از شبکه‌های طراحی شده پارامترهای شبکه، بهینه‌سازی گردیدند. برای این منظور شبکه‌هایی با مقدار μ ثابت و تعداد ورودی‌های از ۳ تا ۲۵ توصیف‌کننده با گام ۱، تعداد گره لایه‌ی پنهان از ۱۰ تا ۳۰ با گام ۱ و تعداد دور آموزشی از ۵۰ تا ۴۰۰ با گام ۱۰ در نظر گرفته شدند و به ازای همه ترکیب‌های ممکن از این سه پارامتر، چهار شبکه عصبی مصنوعی طراحی شده، بهینه‌سازی گردیدند. لازم به ذکر است که در اینجا توصیف‌کننده‌های انتخاب شده توسط روش GA که در جدول (۳-۱۷) نمایش داده شده‌اند، به روش رتبه‌بندی همبستگی، CR، به شبکه عصبی وارد شدند. در جدول (۳-۲۰) مقادیر همبستگی توصیف‌کننده‌های تئوری انتخاب شده توسط روش GA با متغیر وابسته گزارش شده است.

جدول (۳-۲۰)- مقادیر همبستگی توصیف‌کننده‌های تئوری انتخابی با روش GA با دانسیته

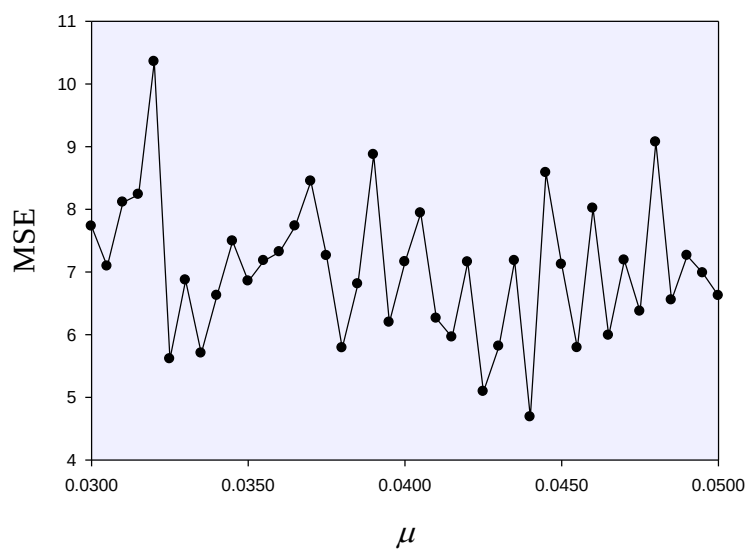
توصیف کننده	X5A	RDF130v	R2m-A	JGI8	MATS1m	G2v	Mor30u	HATS5m
مقدار همبستگی	-۰/۲۱۳	-۰/۲۰۸	۰/۷۳۵	-۰/۲۸۲	۰/۹۰۶	-۰/۱۹۱	-۰/۶۴۶	۰/۴۹۹
توصیف کننده	RTp	Mor29e	Ds	Mor11u	C001	Ss	Mor20m	
مقدار همبستگی	۰/۵۹۱	۰/۴۶۲	۰/۸۴۲	۰/۳۷۹	-۰/۶۷۱	۰/۲۷	۰/۷۰۸	
توصیف کننده	Mor10p	RDF155u	GATS5m	Mor16m	H047	Mor15m	R1m-A	
مقدار همبستگی	-۰/۲۵۸	-۰/۰۱	-۰/۰۵۷	۰/۵۴۳	۰/۷۷۱	-۰/۲۱۱	۰/۸۰۴	

مقدار پارامترهای بهینه شدهی هر یک از شبکه‌های GA-ANN طراحی شده در جدول (۳-۲۱) گزارش شده است. نتایج موجود در این جدول نشان می‌دهد که از میان چهار شبکه بهینه شده، شبکه‌ای با الگوریتم آموزشی تنظیم بایزین و تابع انتقال لگاریتم سیگموئیدی با ۷ توصیف‌کننده در لایه‌ی ورودی، ۲۶ گره در لایه‌ی پنهان و دور آموزشی ۳۳۰، کمترین میانگین مربع خطا را داشته و به عنوان شبکه‌ی بهینه انتخاب می‌شود. ۷ توصیف‌کننده‌ی انتخاب شده توسط شبکه‌ی بهینه عبارتند از دما، فشار، جرم‌مولی کل، MATS1m، Ds، H047، R1m-A.

جدول (۳-۲۱) - توابع و پارامترهای بهینه‌ی شبکه GA-ANN

تابع انتقال	تابع آموزش	تعداد توصیف کننده	تعداد نرون لایه پنهان	تعداد دور آموزش	MSE
لگاریتم سیگموئید	تنظیم بایزین	۷	۲۶	۳۳۰	۶/۱۷۰
لگاریتم سیگموئید	لونبرگ-مارکوارت	۸	۲۶	۳۹۰	۶/۳۰۲
تانژانت سیگموئید	تنظیم بایزین	۷	۲۱	۳۵۰	۱۰/۶۸۱
تانژانت سیگموئید	لونبرگ-مارکوارت	۷	۲۶	۳۲۰	۷/۱۸۴

در ادامه بهینه کردن شبکه مورد بررسی، مقدار پارامتر μ نیز بایستی بهینه شود که بدین منظور، در ساختار شبکه‌ی بهینه با ۷ متغیر ورودی، ۲۶ گره در لایه‌ی پنهان، الگوریتم آموزشی تنظیم بایزین، تابع انتقال لگاریتم سیگموئیدی و دور آموزشی ۳۳۰، 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 ۰/۰۴۴۰ به عنوان مقدار بهینه‌ی μ انتخاب شد. نمودار مقدار میانگین مربعات خطا بر حسب μ حول نقطه‌ی بهینه در شکل (۳-۱۷) رسم شده است.



شکل (۳-۱۷) - نمودار میانگین مربع خطای حاصل از سری ارزیابی بر حسب پارامتر μ

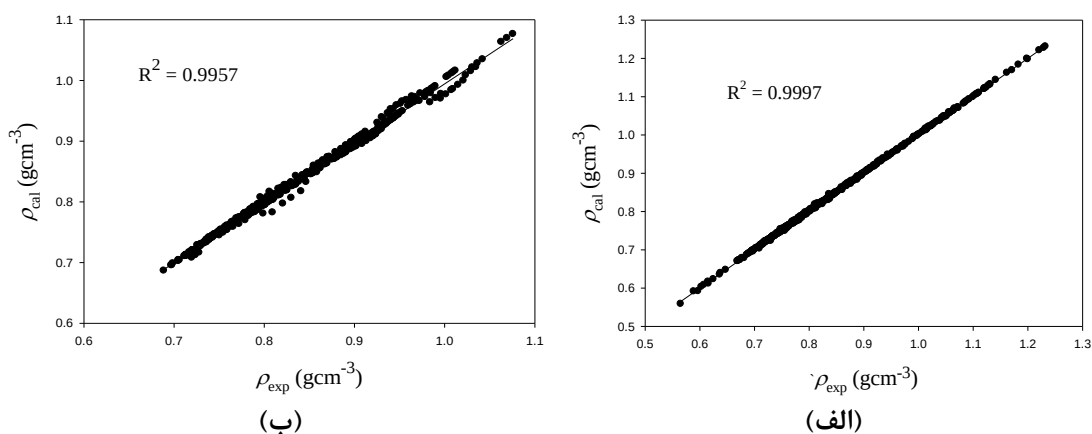
با توجه به روند بهینه‌سازی ارائه شده، مقدار پارامترهای بهینه شده‌ی شبکه‌ی عصبی مصنوعی با استفاده از توصیف‌کننده‌های منتخب الگوریتم ژنتیک (GA-ANN) در جدول (۳-۲۲) گزارش شده است.

جدول (۳-۲۲) - توابع و پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های تئوری انتخابی GA

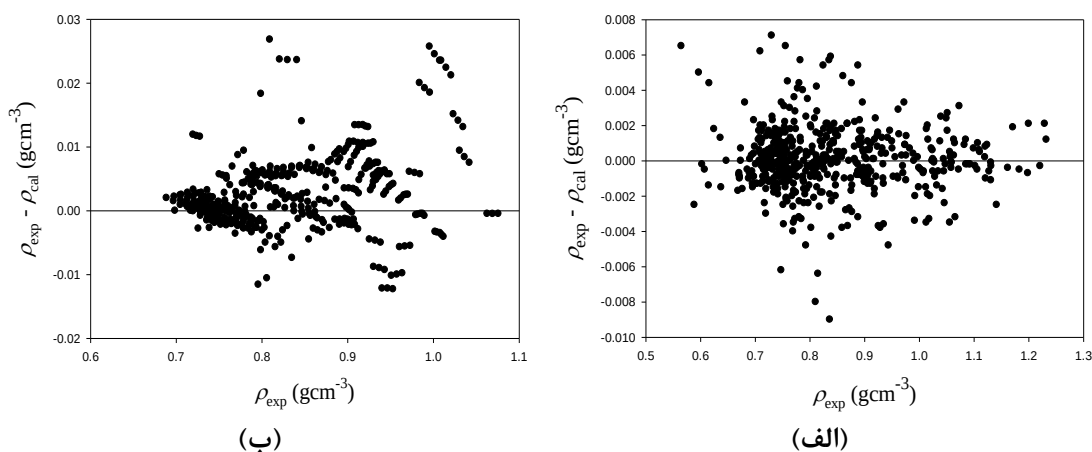
تابع آموزش	Train br
تابع انتقال لایه‌ی پنهان	Log sig
تابع انتقال لایه‌ی خروجی	pure line
تعداد نرون لایه‌ی پنهان	۲۶
تعداد متغیرهای ورودی	۷
تعداد دوره‌های آموزش	۳۳۰
پارامتر μ	۰/۰۴۴
مقدار MSE سری ارزیابی	۴/۶۸۳

۴-۲-۳- ارزیابی شبکه‌ی بهینه شده‌ی GA-ANN

مجدداً برای ارزیابی شبکه‌ی بهینه شده‌ی GA-ANN، نمودارهای برگشتی و خطای باقیمانده برای دو سری ارزیابی و تست ترسیم گردیدند. این نمودارها در شکل‌های (۳-۱۸) و (۳-۱۹) نشان داده شده‌اند.



شکل (۳-۱۸)- نمودار برگشتی مقادیر پیش‌بینی شده‌ی دانسیته‌ی مخلوط‌ها بر حسب مقادیر تجربی (الف) سری ارزیابی و (ب) سری تست



شکل (۳-۱۹)- نمودار خطای باقیمانده دانسیته‌ی مخلوط‌ها بر حسب مقادیر تجربی (الف) سری ارزیابی و (ب) سری تست

مقایسه‌ی نمودارهای (۱۸-۳-الف) و (۱۸-۳-ب) با نمودارهای (۱۳-۳-ب) و (۱۴-۳-ب) نشان می‌دهد که مدل‌های بهینه شده به دو روش GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل و GA-ANN در پیش‌بینی داده‌های سری ارزیابی که به روش ردمخلوط‌ها انتخاب شده‌اند چندان تفاوت ندارند، اما در پیش‌بینی سری تست که به روش ردترکیبات انتخاب شده‌اند روش GCM-ANN کارایی بالاتری دارد. این موضوع در مقایسه بین نمودارهای خطای باقیمانده‌ی دو روش بیشتر نمایان می‌شود. خطای باقیمانده سری تست در روش GCM-ANN -0.008 تا $+0.008$ است، اما این مقدار در روش GA-ANN -0.02 تا $+0.03$ است.

۳-۵- ارزیابی قدرت پیش‌بینی مدل‌های غیر خطی بهینه شده

در این مرحله قدرت پیش‌بینی مدل‌های به‌دست آمده از شبکه‌های عصبی مصنوعی که با روش‌های متفاوت انتخاب توصیف‌کننده، طراحی و بهینه‌سازی شده‌اند، مورد ارزیابی قرار گرفته است. برای این منظور از نمودار برگشتی و نمودار خطای باقیمانده و پارامترهای آماری استفاده شده است. در نمودارهای برگشتی مدل بهینه شده‌ی GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل در شکل‌های (۱۳-۳-ب) و (۱۴-۳-ب) و همچنین نمودارهای برگشتی مدل بهینه شده‌ی GA-ANN در شکل‌های (۱۸-۳-الف) و (۱۸-۳-ب) به ترتیب برای دو سری ارزیابی و تست دیدیم که هر دو روش در سری ارزیابی کارایی خوبی دارند، اما مدل GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل برای پیش‌بینی سری تست از کارایی بسیار بالاتری برخوردار است زیرا در نمودارهای سری تست R^2 برای مدل GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل دارای مقدار 0.9996 درحالی‌که، R^2 برای مدل GA-ANN دارای مقدار 0.9957 است که نشان از برتری روش سهم‌گروه دارد.

این موضوع در شکل‌های (۱۵-۳-ب)، (۱۶-۳-ب) و (۱۸-۳-ب)، (۱۹-۳-ب) که خطای باقیمانده‌ی حاصل از دو روش فوق در مقابل مقدار تجربی دانسیته رسم شده است، بهتر نمایان می‌شود. مقدار

خطای باقیمانده در مدل GCM-ANN بهینه شده در حضور توصیف‌کننده‌ی جرم‌مولی کل بسیار کمتر از مقدار خطای باقیمانده در مدل GA-ANN بهینه شده است.

۳-۶- استفاده از پارامترهای آماری

پارامترهای آماری مقادیر تعیین‌کننده‌ای برای ارزیابی مدل‌های بهینه شده هستند. این مقادیر برای ارزیابی مدل‌های بهینه شده با روش‌های غیرخطی GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل و GA-ANN محاسبه شده که نتایج این محاسبات در جدول (۳-۲۳) نشان داده شده است همان‌طور که در بخش (۳-۳) توضیح داده شد نتایج سری تست در مدل‌سازی مخلوط‌ها از اهمیت ویژه‌ای برخوردار است. اگر به اعداد جدول (۳-۲۳) برای سری تست در دو مدل بهینه شده دقت کنید می‌بینید مقدار R^2 برای مدل GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل بیشتر از روش GA-ANN است همچنین سایر پارامترهای آماری روش GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل مثل MSE کمتر از روش‌های GA-ANN است. با توجه به این مقادیر می‌توان نتیجه گرفت روش سهم‌گروه برای پیش‌بینی دانسیته‌ی مخلوط‌ها روش کارآمدتری است. از طرفی روش سهم‌گروه یک روش ساده و قابل دسترس است و نیاز به نرم‌افزارهای متعدد برای محاسبه‌ی توصیف‌کننده‌ها ندارد.

جدول (۳-۲۳) - پارامترهای آماری مدل‌های بهینه شده‌ی GA-ANN و GCM-ANN در حضور توصیف‌کننده‌ی جرم‌مولی کل برای دو سری ارزیابی و تست

پارامتر آماری	سری داده	GA-ANN	GCM-ANN
R^2	ارزیابی	۰/۹۹۹۷۸۶	۰/۹۹۹۹۵۶۹
	تست	۰/۹۹۵۶۸۱۱	۰/۹۹۹۶۳۳۵
MSE	ارزیابی	۴/۶۸۲۰۸	۱/۰۴۳۱۶۶
	تست	۴۰/۶۶۳۹۶	۱۱/۰۵۱۴۰۲
PRESS	ارزیابی	۲۴۲۵/۳۲	۵۴۰/۳۶
	تست	۱۷۹۷۳/۴۷	۴۸۸۴/۷۲
MAE	ارزیابی	۱/۳۵۵۵۹۸	۰/۴۷۶۰۶۲
	تست	۴/۳۵۷۶۹۲	۱/۳۲۷۱۴۹
MPE	ارزیابی	۰/۰۱۵۷۶۵	۰/۰۰۶۶۹۳
	تست	۰/۳۰۷۷۲۶۰	-۰/۰۰۷۹۹
MAPE	ارزیابی	۰/۱۶۶۸۲	۰/۰۶۰۴۵۳
	تست	۰/۵۰۱۹۶۱	۰/۱۵۴۲۷۵
D_{max}	ارزیابی	۲/۴۴۶۷۴۷	۱/۶۱۴۰۶
	تست	۳/۳۱۱۵۰۴	۴/۲۸۷۶۵۶

۷-۳- بررسی ارتباط توصیف‌کننده‌های انتخاب شده در مدل GA-ANN با دانسیته‌ی

مخلوط‌ها

در این قسمت به طور خلاصه ارتباط بین توصیف‌کننده‌های تئوری انتخاب شده در مدل GA-ANN و دانسیته مخلوط‌ها مورد بررسی قرار خواهد گرفت. این توصیف‌کننده‌ها شامل MATS1m، Ds، R1m-A، H047 می‌باشد.

۱-۷-۳- توصیف‌کننده‌ی MATS1m

توصیف‌کننده‌ی MATS1m^۱ جزء توصیف‌کننده‌های گروه Moran Autocorrelation هستند که از طریق رابطه زیر محاسبه می‌شوند.

$$I(d) = \frac{\frac{1}{2} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (\omega_i - \bar{\omega})(\omega_j - \bar{\omega})}{\frac{1}{A} \sum_{i=1}^A (\omega_i - \bar{\omega})^2} \quad (۳-۵)$$

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). در اینجا در حالتی که $d_{ij} = d$ باشد مقدار $\delta_{ij} = 1$ است در غیر این صورت صفر است. مقدار این توصیف‌کننده از جنس فاصله است و از ۱- تا ۱- متغیر است [۶۶]. توصیف‌کننده‌ی انتخاب شده‌ی MATS1m^۲ جزء توصیف‌کننده‌های دوبعدی خودهمبستگی است که با جرم اتمی وزن دار شده است.

در جدول (۳-۲۴) چگونگی ارتباط این توصیف‌کننده با دانسیته‌ی شش مخلوط دوتایی در کسرمولی‌های متفاوت، در دما و فشار معین به عنوان مثال آورده شده است. بررسی اثر این

^۱2D-Autocorrelation descriptor

^۲Moran autocorrelation log1 Weighted by atomic masses

توصیف کننده نشان میدهد که با افزایش مقدار توصیف کننده ی MATS1m، برای هر مخلوط دانسیته مخلوط افزایش می یابد.

جدول (۳-۲۴) - مثال‌هایی از مقدار توصیف‌کننده MATS1m در تعدادی از کسرمولی‌ها

شماره مخلوط	جزء ۱	جزء ۲	دما (K)	فشار (MPa)	مقدار توصیف‌کننده جزء ۱	مقدار توصیف‌کننده جزء ۲	کسرمولی جزء ۱	مقدار محاسبه شده توصیف‌کننده MATS1m	Density (kg m ⁻³)
۱	Dodecane	Ethyl chloroacetate	۲۹۸/۱۵	۰/۱	۰/۸۴۲	۰/۹۵۶	۰/۹۰۳۹	۰/۸۵۲۹۵۵	۷۶۲/۴
							۰/۶۹۸۸	۰/۸۷۶۳۳۷	۸۰۷/۸
							۰/۴۹۷۱	۰/۸۹۹۳۳۱	۸۶۷/۲
							۰/۳۹۷۱	۰/۹۱۰۷۳۱	۹۰۴
							۰/۰۹۸۳	۰/۹۹۴۴۷۹	۱۰۶۵/۵
۲	Octane	1-Chlorohexane	۲۹۸/۱۵	۶/۱۸	۰/۸۳۹	۰/۹	۰/۱۰۱۵۸	۰/۸۹۳۸۰۴	۸۵۷/۶۷
							۰/۱۹۹۵۵	۰/۸۸۷۸۲۷	۸۳۸/۲۵
							۰/۳۱۵۳۳	۰/۸۸۰۷۶۵	۸۱۶/۳۷
							۰/۳۹۴۸۳	۰/۸۷۵۹۱۵	۸۰۱/۶۸
							۰/۴۸۶۲	۰/۸۷۰۳۴۲	۷۸۵/۴۹
							۰/۵۹۳۸۵	۰/۸۶۳۷۷۵	۷۶۷/۰۸
							۰/۶۹۶۵۹	۰/۸۵۷۵۰۸	۷۵۰/۱۵
							۰/۷۹۴۰۲	۰/۸۵۱۵۶۵	۷۳۴/۶۱
							۰/۹۰۰۱۲	۰/۸۴۵۰۹۳	۷۱۸/۳۳
۳	Octane	1-Chlorobutane	۲۹۸/۱۵	۲/۱۳	۰/۸۳۹	۰/۹۱۶	۰/۱۲۷۲۱	۰/۹۰۶۲۰۵	۸۴۷/۹۹
							۰/۲۹۷۸۴	۰/۸۹۳۰۶۶	۸۰۸/۳
							۰/۴۸۹۳۳	۰/۸۷۸۳۲۲	۷۷۱/۴۹
							۰/۷۰۶۵۳	۰/۸۶۱۵۹۷	۷۳۷/۰۹
							۰/۸۸۹۸۴	۰/۸۴۷۴۸۲	۷۱۳/۱۷
۴	Hexane	1-Chlorobutane	۲۸۳/۱۵	۲/۵	۰/۲۳۷	۰/۹۱۶	۰/۱۰۹۸۵	۰/۸۴۳۸۹۸	۶۹۰/۵۱
							۰/۲۹۲۴۸۸	۰/۸۵۸۶۹۱	۷۲۶/۴۹
							۰/۴۰۴۹۹	۰/۸۶۷۸۰۴	۷۵۰/۱۹
							۰/۵۹۷۵	۰/۸۸۳۳۹۸	۷۹۳/۷۶
							۰/۸۲۹۳۹	۰/۹۰۲۱۸۱	۸۵۲/۰۴
۵	1,4-Butanediol	1-Dodecanol	۲۹۸/۱۵	۲۴/۵۲	۰/۸۸۶	۰/۸۸۵	۰/۰۰۱۶۶	۰/۸۸۵۹۴۹	۱۰۲۲/۵۲
							۰/۰۴۹۹۵	۰/۸۸۴۴۵۲	۱۰۱۰/۱۳
							۰/۲۵۲۹۲	۰/۸۷۸۱۵۹	۹۵۶/۰۸
							۰/۵۵۶۷۴	۰/۸۶۸۷۴۱	۸۹۳/۵۹
۶	2-Ethoxyethanol	1,4-Dioxane	۳۱۳/۱۵	۰/۱	۰/۸۸۶	۰/۹۰۳	۰/۰۵۹۴	۰/۹۰۱۹۹	۱۰۰۳/۶۴
							۰/۱۴۴۳	۰/۹۰۰۵۴۷	۹۹۳/۶۱
							۰/۳۰۴۱	۰/۸۹۷۸۳	۹۷۵/۹
							۰/۴۵۶۸	۰/۸۹۵۲۳۴	۹۶۰/۲۵
							۰/۶۴۶۸	۰/۸۹۲۰۰۴	۱/۹۴۲
							۰/۸۶۰۸	۰/۸۸۸۳۷	۹۲۳/۱۶

۳-۷-۲- توصیف‌کننده‌ی Ds^۱

توصیف‌کننده‌ی انتخاب شده‌ی Ds جزء توصیف‌کننده‌های WHIM^۲ است، توصیف‌کننده‌های WHIM از مختصات کارتزین سه بعدی مولکول، با استفاده از صورتبندی با حداقل انرژی محاسبه می‌شود و شامل اطلاعاتی درباره اندازه، شکل، تقارن و توزیع اتمی ساختار سه بعدی مولکول می‌باشد. این توصیف‌کننده‌ها از رابطه‌ی (۳-۶) به دست می‌آیند.

$$S_{jk} = \frac{\sum_{i=1}^A \omega_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k)}{\sum_{i=1}^A \omega_i} \quad (3-6)$$

که S_{jk} کوواریانس وزن دار شده بین کئوردینه j ام و k ام، A تعداد اتم‌ها، ω_i وزن i امین اتم، q_{ij} و q_{ik} به ترتیب j امین و k امین کئوردینه‌های اتم i ام و $\bar{q}$ مقدار میانگین متناظر است [۶۶]. جدول (۳-۲۵) چگونگی ارتباط این توصیف‌کننده با دانسیته‌ی شش مخلوط دوتایی در کسرمولی‌های متفاوت، در دما و فشار معین به عنوان مثال آورده شده است. بررسی اثر این توصیف‌کننده نشان می‌دهد که با افزایش مقدار توصیف‌کننده‌ی Ds، برای هر مخلوط دانسیته مخلوط افزایش می‌یابد.

^۱-D total accessibility index weighted by atomic electrotopological states

^۲-Weighted Holistic Invariant Molecular descriptors

جدول (۳-۲۵) - مثال‌هایی از مقدار توصیف‌کننده Ds در تعدادی از کسرمولی‌ها

شماره مخلوط	جزء ۱	جزء ۲	دما (K)	فشار (MPa)	مقدار توصیف‌کننده جزء ۱	مقدار توصیف‌کننده جزء ۲	کسرمولی جزء ۱	مقدار محاسبه شده توصیف‌کننده Ds	Density (kg m ⁻³)
۱	Dodecane	Ethyl chloroacetate	۲۹۸/۱۵	۰/۱	۰/۱۶۶	۰/۵۵۴	۰/۹۰۳۹	۰/۲۰۳۲۸۷	۷۶۲/۴
							۰/۶۹۸۸	۰/۲۸۲۸۶۶	۸۰۷/۸
							۰/۴۹۷۱	۰/۳۶۱۱۲۵	۸۶۷/۲
							۰/۳۹۷۱	۰/۳۹۹۹۲۵	۹۰۴
							۰/۰۹۸۳	۰/۵۱۵۸۶	۱۰۶۵/۵
۲	Octane	1-Chlorohexane	۲۹۸/۱۵	۶/۱۸	۰/۱۵	۰/۳۱۴	۰/۱۰۱۵۸	۰/۲۹۷۳۴۱	۸۵۷/۶۷
							۰/۱۹۹۵۵	۰/۲۸۱۲۷۴	۸۳۸/۲۵
							۰/۳۱۵۳۳	۰/۲۶۲۲۸۶	۸۱۶/۳۷
							۰/۳۹۴۸۳	۰/۲۴۹۲۴۸	۸۰۱/۶۸
							۰/۴۸۶۲	۰/۲۳۴۲۶۳	۷۸۵/۴۹
							۰/۵۹۳۸۵	۰/۲۱۶۶۰۹	۷۶۷/۰۸
							۰/۶۹۶۵۹	۰/۱۹۹۷۵۹	۷۵۰/۱۵
							۰/۷۹۴۰۲	۰/۱۸۳۷۸۱	۷۳۴/۶۱
۳	Octane	1-Chlorobutane	۲۹۸/۱۵	۲/۱۳	۰/۱۵	۰/۳۶۴	۰/۱۲۷۲۱	۰/۳۳۶۷۷۷	۸۴۷/۹۹
							۰/۲۹۷۸۴	۰/۳۰۰۲۶۲	۸۰۸/۳
							۰/۴۸۹۳۳	۰/۲۵۹۲۸۳	۷۷۱/۴۹
							۰/۷۰۶۵۳	۰/۲۱۲۸۰۳	۷۳۷/۰۹
							۰/۸۸۹۸۴	۰/۱۷۳۵۷۴	۷۱۳/۱۷
۴	Hexane	1-Chlorobutane	۲۸۳/۱۵	۲/۵	۰/۱۳۵	۰/۳۶۴	۰/۱۰۹۸۵	۰/۱۶۰۱۵۶	۶۹۰/۵۱
							۰/۲۹۲۴۸۸	۰/۲۰۱۹۷۸	۷۲۶/۴۹
							۰/۴۰۴۹۹	۰/۲۲۷۷۴۳	۷۵۰/۱۹
							۰/۵۹۷۵	۰/۲۷۱۸۲۸	۷۹۳/۷۶
							۰/۸۲۹۳۹	۰/۳۲۴۹۳	۸۵۲/۰۴
۵	1,4-Butanediol	1-Dodecanol	۲۹۸/۱۵	۲۴/۵۲	۰/۲۶۴	۰/۲۲۶	۰/۰۰۱۶۶	۰/۲۶۳۳۹۳۷	۱۰۲۲/۵۲
							۰/۰۴۹۹۵	۰/۲۶۲۱۰۲	۱۰۱۰/۱۳
							۰/۲۵۲۹۲	۰/۲۵۴۳۸۹	۹۵۶/۰۸
							۰/۵۵۶۷۴	۰/۲۴۲۸۴۴	۸۹۳/۵۹
۶	2-Ethoxyethanol	1,4-Dioxane	۳۱۳/۱۵	۰/۱	۰/۱۷۳	۰/۳۱۷	۰/۰۵۹۴	۰/۳۰۸۴۴۶	۱۰۰۳/۶۴
							۰/۱۴۴۳	۰/۲۹۶۲۲۱	۹۹۳/۶۱
							۰/۳۰۴۱	۰/۲۷۳۲۱	۹۷۵/۹
							۰/۴۵۶۸	۰/۲۵۱۲۲۱	۹۶۰/۲۵
							۰/۶۴۶۸	۰/۲۲۳۸۶۱	۹۴۲/۱
							۰/۸۶۰۸	۰/۱۹۳۰۷۴	۹۲۳/۱۶

۳-۷-۳- توصیف‌کننده‌ی R1m-A^۱

این توصیف‌کننده که جزء توصیف‌کننده‌های گروه GETAWAY^۲ است، با توجه به مختصات فضایی اتم‌ها در یک مولکول به راحتی قابل محاسبه‌اند این توصیف‌کننده‌ها با استفاده ماتریس قدرت نفوذ^۳ تعریف می‌شوند و بیان‌کننده ویژگی‌های هندسی و توپولوژیکی مولکول‌ها هستند. ماتریس قدرت نفوذ یا ماتریس تاثیر مولکول^۴ (MIM) با رابطه زیر بیان می‌شود.

$$H = M.(M^T.M)^{-1}.M^T \quad (۳-۷)$$

که M ماتریس مختصات اتمی، T به معنای ماتریس ترانپوز و H ماتریس نفوذ مولکولی است که یک ماتریس $A \times A$ بوده که A نیز تعداد اتم‌هاست. عناصر قطری ماتریس H لوریج‌ها نام دارند که هریک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. اتم‌های سطحی اعداد لوریج بزرگتری نسبت به اتم‌های مرکزی دارند. همچنین مولکول‌های کروی اعداد لوریج پایین‌تری نسبت به مولکول‌های کشیده دارند [۶۶]. توصیف‌کننده‌های GETAWAT به دو زیر مجموعه‌ی H و R تقسیم می‌شوند. توصیف‌کننده‌های H-GETAWAY به اندازه و شکل مولکول وابسته بوده به گونه‌ای که با افزایش اندازه اتم و فاصله‌ی اتم از مرکز مولکول، مقدار آنها افزایش می‌یابد. گروه دیگر توصیف‌کننده‌های R-GETAWAY هستند که براساس ماتریس فاصله-نفوذ به صورت زیر تعریف می‌شوند.

$$[R]_{ij} = \left[\frac{\sqrt{h_{ii} \cdot h_{jj}}}{r_{ij}} \right] \quad i \neq j \quad (۳-۸)$$

که h_{ii} و h_{jj} لوریج‌های دو اتم i و j و r_{ij} فاصله‌ی آن دو است. عناصر قطری این ماتریس نیز صفر هستند. این توصیف‌کننده‌ها به ساختار سه بعدی مولکول‌ها حساس بوده و در برگیرنده‌ی اطلاعاتی نظیر شکل، اندازه، تقارن مولکول و نحوه‌ی توزیع اتم‌ها در مولکول می‌باشد.

^۱-R maximal autocorrelation of leg1 weighted by atomic masses

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

^۳-Leverage Matrix

^۴-Molecular Influence Matrix

جدول (۳-۲۶) چگونگی ارتباط این توصیف‌کننده با دانسیته‌ی شش مخلوط دوتایی در کسرمولی‌های متفاوت، در دما و فشار معین به عنوان مثال آورده شده است. بررسی اثر این توصیف‌کننده نشان می‌دهد که با افزایش مقدار توصیف‌کننده‌ی R1m-A، برای هر مخلوط دانسیته مخلوط افزایش می‌یابد.

جدول (۳-۲۶)-مثال‌هایی از مقدار توصیف‌کننده R1m-A در تعدادی از کسرمولی‌ها

شماره مخلوط	جزء ۱	جزء ۲	دما (K)	فشار (MPa)	مقدار توصیف‌کننده جزء ۱	مقدار توصیف‌کننده جزء ۲	کسرمولی جزء ۱	مقدار محاسبه شده توصیف‌کننده R1m-A	Density (kg m ⁻³)
۱	Dodecane	Ethyl chloroacetate	۲۹۸/۱۵	۰/۱	۰/۱۶۶	۰/۵۵۴	۰/۹۰۳۹	۰/۲۰۳۲۸۷	۷۶۲/۴
							۰/۶۹۸۸	۰/۲۸۲۸۶۶	۸۰۷/۸
							۰/۴۹۷۱	۰/۳۶۱۱۲۵	۸۶۷/۲
							۰/۳۹۷۱	۰/۳۹۹۹۲۵	۹۰۴
							۰/۰۹۸۳	۰/۵۱۵۸۶	۱۰۶۵/۵
۲	Octane	1-Chlorohexane	۲۹۸/۱۵	۶/۱۸	۰/۲۶۳	۰/۴۴۴	۰/۱۰۱۵۸	۰/۴۲۵۶۱۴	۸۵۷/۶۷
							۰/۱۹۹۵۵	۰/۴۰۷۸۸۱	۸۳۸/۲۵
							۰/۳۱۵۳۳	۰/۳۸۶۹۲۵	۸۱۶/۳۷
							۰/۳۹۴۸۳	۰/۳۷۲۵۳۶	۸۰۱/۶۸
							۰/۴۸۶۲	۰/۳۵۵۹۹۸	۷۸۵/۴۹
							۰/۵۹۳۸۵	۰/۳۳۶۵۱۳	۷۶۷/۰۸
							۰/۶۹۶۵۹	۰/۳۱۷۹۱۷	۷۵۰/۱۵
							۰/۷۹۴۰۲	۰/۳۰۰۲۸۲	۷۳۴/۶۱
							۰/۹۰۰۱۲	۰/۲۸۱۰۷۸	۷۱۸/۳۳
۳	Octane	1-Chlorobutane	۲۹۸/۱۵	۲/۱۳	۰/۲۶۳	۰/۴۵۸	۰/۱۲۷۲۱	۰/۴۳۳۱۹۴	۸۴۷/۹۹
							۰/۲۹۷۸۴	۰/۳۹۹۹۲۱	۸۰۸/۳
							۰/۴۸۹۳۳	۰/۳۶۲۵۸۱	۷۷۱/۴۹
							۰/۷۰۶۵۳	۰/۳۲۰۲۲۷	۷۳۷/۰۹
							۰/۸۸۹۸۴	۰/۲۸۴۴۸۱	۷۱۳/۱۷
۴	Hexane	1-Chlorobutane	۲۸۳/۱۵	۲/۵	۰/۲۳۷	۰/۴۵۸	۰/۱۰۹۸۵	۰/۲۶۱۲۷۷	۶۹۰/۵۱
							۰/۲۹۲۴۸۸	۰/۳۰۱۶۳۸	۷۲۶/۴۹
							۰/۴۰۴۹۹	۰/۳۲۶۵۰۳	۷۵۰/۱۹
							۰/۵۹۷۵	۰/۳۶۹۰۴۸	۷۹۳/۷۶
							۰/۸۲۹۳۹	۰/۴۲۰۲۹۵	۸۵۲/۰۴
۵	1,4-Butanediol	1-Dodecanol	۲۹۸/۱۵	۲۴/۵۲	۰/۳۶۳	۰/۳۴۲	۰/۰۰۱۶۶	۰/۳۶۲۹۶۵	۱۰۲۲/۵۲
							۰/۰۴۹۹۵	۰/۳۶۱۹۵۱	۱۰۱۰/۱۳
							۰/۲۵۲۹۲	۰/۳۵۷۶۸۹	۹۵۶/۰۸
							۰/۵۵۶۷۴	۰/۳۵۱۳۰۸	۸۹۳/۵۹
۶	2-Ethoxyethanol	1,4-Dioxane	۳۱۳/۱۵	۰/۱	۰/۳۱۴	۰/۶۴۴	۰/۰۵۹۴	۰/۶۲۴۳۹۸	۱۰۰۳/۶۴
							۰/۱۴۴۳	۰/۵۹۶۳۸۱	۹۹۳/۶۱
							۰/۳۰۴۱	۰/۵۴۳۶۴۷	۹۷۵/۹
							۰/۴۵۶۸	۰/۴۹۳۲۵۶	۹۶۰/۲۵
							۰/۶۴۶۸	۰/۴۳۰۵۵۶	۹۴۲/۱
							۰/۸۶۰۸	۰/۳۶۰۰۲	۹۲۳/۱۶

۴-۷-۳- توصیف‌کننده‌ی H047^۱

توصیف‌کننده‌ی H047 جزء توصیف‌کننده‌های کدهای اجزای با مرکزیت اتمی (ACFC)^۲ یک کد با مرکزیت اتمی با دامنه‌ی کوتاه است که هر اتم را با نوع اتم، انواع پیوند و انواع اتم‌های همسایه توصیف می‌کند. هر مولکول کاملاً با یک کد مجزا که تعداد اتم‌های غیرهیدروژنی است معرفی می‌شود [۶۶]. این توصیف‌کننده‌ها اطلاعات شیمیایی زیادی را در رابطه با گروه‌های عاملی در اطراف یک اتم مرکزی و مولکول ارائه می‌دهد. توصیف‌کننده‌ی H047 (H متصل به کربن با هیبریداسیون sp^3 و sp^2) جزء این گروه از توصیف‌کننده‌ها است، مقدار این توصیف برای ترکیبات آلکان‌های خطی و غیرخطی دارای مقادیر صفر است و برای ترکیباتی که دارای گروه‌های عاملی هستند دارای مقدار است. جدول (۳-۲۷) چگونگی ارتباط این توصیف‌کننده با دانسیته‌ی شش مخلوط دوتایی در کسرمولی‌های متفاوت، در دما و فشار معین به عنوان مثال آورده شده است. بررسی اثر این توصیف‌کننده نشان می‌دهد که با افزایش مقدار توصیف‌کننده‌ی H047، برای هر مخلوط دانسیته مخلوط افزایش می‌یابد.

^۱-H attached to C1 (sp^3) / C0 (sp^2)

^۲-Atom-Centred Fragment Code

جدول (۳-۲۷) - مثال هایی از مقدار توصیف کننده H047 در تعدادی از کسرمولی ها

شماره مخلوط	جزء ۱	جزء ۲	دما (K)	فشار (MPa)	مقدار توصیف کننده جزء ۱	مقدار توصیف کننده جزء ۲	کسرمولی جزء ۱	مقدار محاسبه شده توصیف کننده H047	Density (kg m ⁻³)
۱	Methyl metacrylate	2-butanol	۲۹۸/۱۵	۰/۱	۵	۱	۰/۰۵۹۴	۱/۲۳۷۶	۸۱۰/۶۷
							۰/۳۵۵۹	۲/۴۲۳۶	۸۵۰/۸۴
							۰/۵۰۱۹	۳/۰۰۷۶	۸۷۰/۶۶
							۰/۷۵۵۴	۴/۰۲۱۶	۹۰۵/۰۱
							۰/۹۵۴۵	۴/۸۱۸	۹۳۱/۶۱
۲	Octane	1-Chlorohexane	۲۹۸/۱۵	۶/۱۸	۰	۲	۰/۱۰۱۵۸	۰/۴۲۵۶۱۴	۸۵۷/۶۷
							۰/۱۹۹۵۵	۰/۴۰۷۸۸۱	۸۳۸/۲۵
							۰/۳۱۵۳۳	۰/۳۸۶۹۲۵	۸۱۶/۳۷
							۰/۳۹۴۸۳	۰/۳۷۲۵۳۶	۸۰۱/۶۸
							۰/۴۸۶۲	۰/۳۵۵۹۹۸	۷۸۵/۴۹
							۰/۵۹۳۸۵	۰/۳۳۶۵۱۳	۷۶۷/۰۸
							۰/۶۹۶۵۹	۰/۳۱۷۹۱۷	۷۵۰/۱۵
							۰/۷۹۴۰۲	۰/۳۰۰۲۸۲	۷۳۴/۶۱
							۰/۹۰۰۱۲	۰/۲۸۱۰۷۸	۷۱۸/۳۳
۳	Octane	1-Chlorobutane	۲۹۸/۱۵	۲/۱۳	۰	۲	۰/۱۲۷۲۱	۰/۴۳۳۱۹۴	۸۴۷/۹۹
							۰/۲۹۷۸۴	۰/۳۹۹۹۲۱	۸۰۸/۳
							۰/۴۸۹۳۳	۰/۳۶۲۵۸۱	۷۷۱/۴۹
							۰/۷۰۶۵۳	۰/۳۲۰۲۲۷	۷۳۷/۰۹
							۰/۸۸۹۸۴	۰/۲۸۴۴۸۱	۷۱۳/۱۷
۴	Hexane	1-Chlorobutane	۲۸۳/۱۵	۲/۵	۰	۲	۰/۱۰۹۸۵	۰/۲۶۱۲۷۷	۶۹۰/۵۱
							۰/۲۹۲۴۸۸	۰/۳۰۱۶۳۸	۷۲۶/۴۹
							۰/۴۰۴۹۹	۰/۳۲۶۵۰۳	۷۵۰/۱۹
							۰/۵۹۷۵	۰/۳۶۹۰۴۸	۷۹۳/۷۶
							۰/۸۲۹۳۹	۰/۴۲۰۲۹۵	۸۵۲/۰۴
۵	1,4-Butanediol	1-Dodecanol	۲۹۸/۱۵	۲۴/۵۲	۴	۲	۰/۰۰۱۶۶	۰/۳۶۲۹۶۵	۱۰۲۲/۵۲
							۰/۰۴۹۹۵	۰/۳۶۱۹۵۱	۱۰۱۰/۱۳
							۰/۲۵۲۹۲	۰/۳۵۷۶۸۹	۹۵۶/۰۸
							۰/۵۵۶۷۴	۰/۳۵۱۳۰۸	۸۹۳/۵۹
۶	2-Ethoxyethanol	1,4-Dioxane	۳۱۳/۱۵	۰/۱	۶	۸	۰/۰۵۹۴	۷/۸۸۱۲	۱۰۰۳/۶۴
							۰/۱۴۴۳	۷/۷۱۱۴	۹۹۳/۶۱
							۰/۳۰۴۱	۷/۳۹۱۸	۹۷۵/۹
							۰/۴۵۶۸	۷/۰۸۶۴	۹۶۰/۲۵
							۰/۶۴۶۸	۶/۷۶۴	۹۴۲/۱
							۰/۸۶۰۸	۶/۲۷۸۸	۹۲۳/۱۶

۸-۳- بررسی میزان مشارکت توصیف‌کننده‌های منتخب شبکه‌ی عصبی بهینه

میزان مشارکت توصیف‌کننده‌های انتخاب شده به صورت زیر تعیین شد:

۱- ابتدا وزن مربوط به توصیف‌کننده‌ی موردنظر 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 (۳-۱۱) محاسبه گردید:

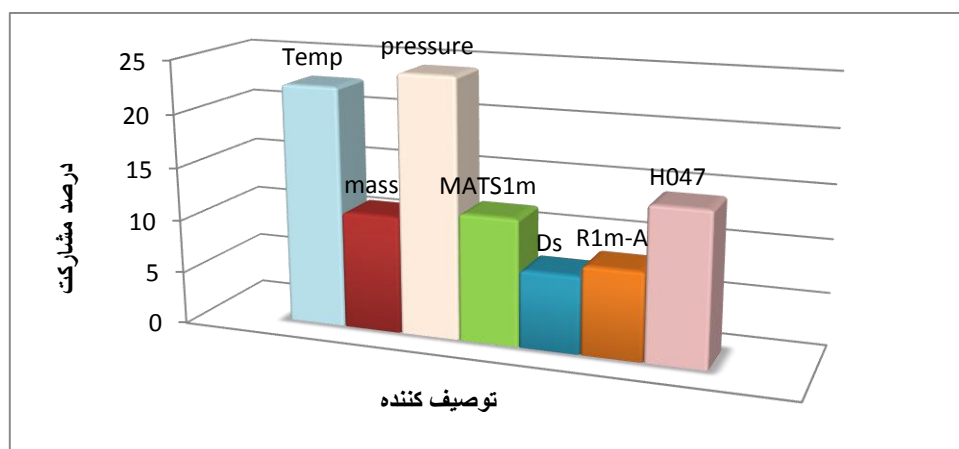
$$C_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad (3-11)$$

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. طبق

روش شرح داده شده درصد مشارکت توصیف‌کننده‌های منتخب در شبکه‌ی بهینه شده‌ی GA-ANN،

محاسبه شد که به صورت نموداری در شکل (۳-۲۰) نشان داده شده است.



شکل (۳-۲۰)- درصد مشارکت توصیف‌کننده‌ها در مدل GA-ANN

همان‌طور که در این شکل مشاهده می‌شود درصد مشارکت دو توصیف‌کننده‌ی دما و فشار بیشتر از توصیف‌کننده‌های تئوری و جرم‌مولی کل است که این به خوبی وابستگی شدید دانسیته به دما و فشار را نشان می‌دهد. همچنین وابستگی به فشار اندکی بیشتر از دماست که با بررسی داده‌های تجربی مشخص می‌شود که تغییرات دانسیته‌ی یک مخلوط معین با تغییر دما در فشار ثابت کمتر از تغییرات دانسیته‌ی همان مخلوط با تغییر فشار در دمای ثابت است.

۹-۳- نتیجه گیری

در این پایان نامه نشان دادیم که می توان روش مدل سازی شبکه ی عصبی مصنوعی را برای پیش بینی دانسیته مخلوط ترکیبات آلی در دما و فشارهای مختلف استفاده نمود. از آنجایی که اندازه گیری دانسیته ی مخلوط ها به صورت تجربی حتی در دما و فشار ثابت به صرف هزینه و زمان زیادی نیاز دارد لذا استفاده از روش های پیش بینی از اهمیت زیادی برخوردار است. در این پایان نامه مشخص شد که یکی از اصلی ترین مسائل در تعمیم مدل های QSPR به مخلوط ها در چگونگی تعریف توصیف کننده های مخلوط ها نهفته است. دو روش ارائه شده در این کار، مخلوط توسط توصیف کننده های وزن دار شده ی مولی و با استفاده از توصیف کننده های اجزاء خالص به شبکه وارد شدند. برای اجزاء مخلوط نیز از دو نوع توصیف کننده استفاده شد: توصیف کننده های برمبنای روش سهم گروه ها و توصیف کننده های تئوری که با استفاده از الگوریتم ژنتیک انتخاب شده بودند. مهم ترین مزیت این روش برای محاسبه ی توصیف کننده های مخلوط ها، سادگی معرفی مخلوط ها به شبکه است به گونه ای که می توان یک مدل پیش بینی کننده ی خوب ساخت. اما از آن طرف ایراد این روش این است که اثرات برهمکنش ها در آن دیده نمی شود. همچنین آن طور که می توان اثر و اهمیت یک توصیف کننده ی مولکولی خاص روی یک خاصیت معین در مواد خالص بررسی نمود، در اینجا روش کار مکانیزه ای برای این کار وجود ندارد. نتیجه مهم دیگر این کار اهمیت انتخاب تست در مخلوط ها بود که نسبت به مواد خالص، متفاوت و در عین حال بسیار سخت گیرانه است. در این پایان نامه از روش رد مخلوط ها برای سری ارزیابی و از روش رد ترکیبات برای سری تست استفاده شد تا مدل بهینه شده یک مدل کارآمد و معتبر باشد.

بررسی نتایج حاصل از مدل بهینه شده ی GCM-ANN در حضور توصیف کننده ی جرم مولی کل و مدل GA-ANN برای پیش بینی داده های سری ارزیابی و سری تست مشخص گردید که مدل GCM-ANN کارایی بسیار بالاتری نسبت به مدل GA-ANN دارد. این در حالی است که روش سهم گروه برای محاسبه ی توصیف کننده های اجزای مخلوط بسیار ساده بوده و نیاز به نرم افزارهای خاصی ندارد.

۱۰-۳- آینده‌نگری

از روش ارائه شده می‌توان برای پیش‌بینی سایر خواص ترمودینامیکی مخلوط‌ها مثل حجم اضافی، آنتالپی اضافی یا خواص انتقالی مخلوط‌ها مثل ویسکوزیته و هدایت گرمایی و.... استفاده کرد. همچنین می‌توان به‌جای استفاده از روش شبکه‌عصبی مصنوعی از روش‌های غیرخطی دیگر مثل ماشین بردار پشتیبان برای مدل‌سازی استفاده کرد.

منابع

- [۱] آزمون فرن، (۱۳۹۲)، پایان نامه کاشناسی ارشد: "مدل سازی شبکه‌ی عصبی مصنوعی برای پیش-بینی دانسیته‌ی سیالات تجمعی با استفاده از توصیف‌کننده‌های مولکولی"، دانشکده‌ی شیمی، دانشگاه صنعتی شاهرود.
- [2] G. A. Parsafar, E.A. Mason, "Linear isotherms for dense fluids: A new regularity", *J. Phys. Chem.*, 97 (1993) 9048-9053.
- [3] G. A. Parsafar, E.A. Mason, "Linear isotherms for dense fluids : extension to mixtures", *J. Phys. Chem.*, 98 (1994) 1962-1967.
- [4] G. A. Parsafar, Z. Kalantar, "Extension of linear isotherm regularity to long chain alkanes", *Iran. J. Chem. Chem. Eng.*, 22 (2003) 1-8.
- [5] 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 Equilibri.*, 234 (2005) 19-29.
- [6] Z. Kalantar, H. Nikoofard, F. Javadi, "Extension of LIR equation of state to alkylamines using group contribution method", *ISRN Phys. Chem.*, 2 (2013) 1-9.
- [۷] مرتضوی فر آ، (۱۳۹۰)، پایان‌نامه کارشناسی ارشد: "پیش بینی خواص حجمی استرها با استفاده از معادله‌ی حالت MLIR و روش سهم گروه"، دانشکده شیمی، دانشگاه شاهرود.
- [8] E. K. Goharshadi, A. Morsali, M. Abbaspour, "New regularities and an equation of state for liquids", *Fluid Phase Equilibri.*, 230 (2005) 170-175.
- [9] E. K. Goharshadi, F. Moosavi, "Prediction of thermodynamic properties of some hydrofluoroether refrigerants using a new equation of state", *Fluid Phase Equilibri.* 238 (2005) 112-119.
- [10] E. K. Goharshadi, A.R. Berenji, "A new equation of state for predicting the thermodynamic properties of liquid alkali metals", *J. Nucl. Mater.* 348 (2006) 40-44.
- [11] E. K. Goharshadi, F. Moosavi, "Prediction of the volumetric and thermodynamic properties of some refrigerants using GMA equation of state", *Int. J. Refrig.* 30 (2007) 377-383.
- [12] E. K. Goharshadi, M. Moosavi, "Application of a new equation of state to liquid refrigerant mixtures", *Thermochim Acta.* 447 (2006) 64-68.

[13] M. Moosavi, E. K. Goharshadi, "Investigation of Volumetric Properties of Some Glycol Ethers Using a Simple Equation of State", *Int. J. Thermophys.*, 27 (2006) 1515-1526.

[14] E. K. Goharshadi, M. Moosavi, "Extension of a New Equation of State to the Liquid Mixtures", *Ind. Eng. Chem. Res.* 44 (2005) 6973-6980.

[15] E. K. Goharshadi, M. Moosavi, "Density calculation using GMA equation of state considering mixing and combining rules for some liquid mixtures", *Fluid Phase Equilibri.*, 245 (2006) 109-116.

[16] 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 Equilibri.*, 310 (2011) 63-73.

[۱۷] پایین محلی ح، (۱۳۸۸)، پایان نامه کارشناسی ارشد: "مطالعه اثربازدارندگی مشتقات 5- هیدروکسی-3-(2H)-پریدازینون بر روی ویروس هپاتیت نوع C با استفاده از روش های کمومتری کس"، دانشکده شیمی، دانشگاه صنعتی شریف.

[۱۸] کریمی م، (۱۳۹۰)، پایان نامه کارشناسی ارشد: "شناسایی ترکیبات اسانس های هل سبز و سیاه با استفاده از روش های GC-MS تلفیق شده با روش های تفکیک منحنی و تعیین کمی و گونه Taxus با استفاده از تکنیک کروماتوگرافی مایع سریع و تلفیق آن با روش های تفکیک منحنی"، دانشکده شیمی، دانشگاه صنعتی شریف.

[۱۹] کیانی آ، (۱۳۸۲)، پایان نامه کارشناسی ارشد: "مدلسازی کامپیوتری زمان بازداری (TR) ترکیبات آلی فرار با استفاده از روش های کمومتری کس"، دانشکده شیمی، دانشگاه صنعتی شریف.

[۲۰] محسنی ب، (۱۳۸۷)، پایان نامه کارشناسی 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.

[۲۱] کریمی س، (۱۳۸۵)، کارشناسی ارشد، "مدل سازی کمومتری کس اثرات مهارکنندگی مشتقات هارمان بعنوان ترکیبات ضد ایدز با استفاده از روش هیبریدی آنالیز اجزای اصلی، رگرسیون خطی چندگانه و شبکه عصبی مصنوعی"، دانشکده شیمی، دانشگاه صنعتی شریف.

- [22] www.iasbs.ac.ir/chemistry/chemometrics/.../8th/qsar_introduction.ppt
- [23] M. Arab Chamjangali, "Modeling of cytotoxicity data (CC50) of anti-HIV 1-[5-chlorophenyl) sulfonyl]-1H-pyrrole derivatives using calculated molecular descriptors and Levenberg-Marquadt artificial neural network", *J.Chem. Bio. Drug. Des.*, 73 (2009) 456-465.
- [24] L. Zhuo, J. Zheng, X. Li, F. Wang, B. Ai, J. Qian, "A Genetic Algorithm based wrapper feature selection method for classification of hyper spectral images using support vector machine", 7147 (2008) 397-402.
- [25] M. Melanie, (1999), "An introduction to genetic algorithms", Vol. 3, Cambridge, Massachusetts London, England, pp 1-144.
- [۲۶] باهر ۱، (۱۳۸۸)، پایان نامه دکتری: "پیش‌بینی و مطالعه QSPR بازداري ترکیبات آلي، دارويي و آلاينده‌ها با استفاده از رگرسيون بردارهاي پشتيبان"، دانشکده شيمي، دانشگاه مازندران.
- [۲۷] کلانتر ز، (۱۳۸۵)، پایان‌نامه دکتری: "پیش‌بینی خواص ترموديناميکي و انتقالی سيالات آلي چگال با استفاده از روش سهم گروه‌ها و شبکه‌ی عصبی موجک"، دانشکده‌ی شيمي، دانشگاه صنعتی اصفهان.
- [28] 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.
- [29] J. O. Valderrama, A. Reategui, R. E. Rojas, "Density of Ionic Liquids Using Group Contribution and Artificial Neural Networks", *Ind. Eng. Chem. Res.*, (2009).
- [30] 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.
- [31] 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 Equilibri.*, 261 (2007) 434-443.
- [۳۲] حسين كوچک م، (۱۳۸۴)، پایان‌نامه کارشناسی ارشد، "بهبود متدولوژی سطح پاسخ با استفاده از شبکه‌های عصبی و الگوریتم"، دانشکده مهندسی صنایع، دانشگاه صنعتی شریف.

[۳۳] طاهرزاده م، (۱۳۹۳)، پایان نامه کارشناسی ارشد: "مطالعه کمی ساختار- خاصیت ضریب فعالیت در رقت بی نهایت ترکیبات آلی و آب در محیط مایع یونی ۱-بوتیل-۱-متیل پیرولیدینیوم تریسیانومتانید"، دانشکده شیمی، دانشگاه صنعتی شاهرود.

[34] V. K. Stamations, Understanding neural networks and fuzzy logic IEEEPRESS, (1996).

[۳۵] کیانی آ، (۱۳۸۲)، پایان نامه کارشناسی ارشد: "مدل سازی کامپیوتری زمان بازداری ترکیبات آلی فرار با استفاده از روش های کمومتری کس"، دانشکده شیمی، دانشگاه صنعتی شریف.

[36] L. Moravkova, Z. Wagner, K. Aim, J. Linek, “ (ρ, V_m, T) Measurements of (octane + 1-chlorohexane) at temperatures from 298.15 K to 328.15 K and at pressures up to 40 MPa”, *J. Chem Thermodyn.*, 38 (2006) 861–870.

[37] J. H. Dymond, J. Robertson, J. D. Isdal, “ (ρ, p, T) of some pure n-alkanes and binary mixtures of n-alkanes in the range 298 to 373 K and 0.1 to 500 MPa”, *J. Chem. Thermodyn.*, 14 (1982) 51-59.

[38] L. Moravkova, J. Linek, “Excess molar volumes of (octane + 1-chloropentane) at temperatures between 298.15 K and 328.15 K and at pressures up to 40 MPa”, *J. Chem. Thermodyn.*, 35 (2003) 1119–1127.

[39] L. Moravkova, J. Linek, “Excess molar volumes of (octane + 1-chlorobutane) at temperatures between 298.15 and 328.15 K and at pressures up to 40MPa”, *J. Chem. Thermodyn.*, 35 (2003) 113–121.

[40] E. M. Alaoui, A. Montero, J. P. atrick Bazile, F. Aguilar, C. Boneda, “ (ρ, V_E, T) Measurements of mixtures (DBE + alcohol) at temperatures from (293.15 to 353.15) K and at pressures up to 140 MPa”, *Fluid Phase Equilibri.*, 363 (2014) 131– 148.

[41] P. Garcia-Gimenez, J. F. Martinez-Lopez, S. T. Blanco, I. Velasco, and S. Otin, “Densities and Isothermal Compressibilities at Pressures up to 20 MPa of the Systems *N,N*-Dimethylformamide or *N,N*-Dimethylacetamide + 1-Chloroalkane”, *J. Chem. Eng. Data.*, 52 (2007) 1693-1699.

[42] P. Garcia-Gimenez, L. Gil, S. T. Blanco, I. Velasco, and S. Otin, “Densities and Isothermal Compressibilities at Pressures up to 20 MPa of the Systems 1-Methyl-2-pyrrolidone + 1-Chloroalkane or + α,ω -Dichloroalkane”, *J. Chem. Eng. Data.*, 53 (2008) 66–72.

- [43] P. Garcia-Gimenez, J. Martinez-Lopez, S. T. Blanco, I. Velasco, and S. Otin, “Densities and Isothermal Compressibilities at Pressures up to 20 MPa of the Systems N,N-Dimethylformamide or N,N-Dimethylacetamide + α,ω -Dichloroalkane”, *J. Chem. Eng. Data.*, 52 (2007) 2368-2374.
- [44] A. Goldon, I. Malka, and Tadeusz Hofman, “Densities and Excess Volumes of the 1-Chlorobutane + n-Hexane System at Temperatures from (283.15 to 333.15) K and Pressures from (0.1 to 35) MPa”, *J. Chem. Eng. Data.*, 53 (2008) 1039–1045.
- [45] T. Regueira, W. Yan, and E. H. Stenby, “Densities of the Binary Systems n-Hexane + n-Decane and n-Hexane + n-Hexadecane Up to 60 MPa and 463 K”, *J. Chem. Eng. Data.*, 60 (2015) 3631-3645.
- [46] G. Watson, C. K. Zeberg-Mikkelsen, A. Baylaucq, and C. Boned, “High-Pressure Density Measurements for the Binary System Ethanol + Heptane”, *J. Chem. Eng. Data.*, 51 (2006) 112-118.
- [47] C. K. Zeberg-Mikkelsen, L. Lugo, J. Fernandez, “Density measurements under pressure for the binary system (ethanol + methylcyclohexane)”, *J. Chem. Thermodyn.*, 38 (2006) 640-649.
- [48] A. Compostizo, A. Diez-Pascual, A. Crespo Colin, and R. G. Rubio, “Density-Pressure Relationship in Hydrogen-Bonded Mixtures: 1,4-Butanediol + 1-Dodecanol”, *J. Chem. Eng. Data* 50 (2005) 591-595.
- [49] R. Malhotra, L. A. Woolf, “ (p, V_m, T, x) measurements for $x\text{CH}_3\text{CH}_2\text{COCH}_3 + (1-x)(\text{CH}_3)_3\text{CCH}_2\text{CH}(\text{CH}_3)_2$ at temperatures between 278 K and 338 K and at pressures between 0.1 MPa and 280 MPa”, *J. Chem. Thermodyn.*, 28 (1996) 701-715.
- [50] W. Weng, “Densities and Viscosities for Binary Mixtures of Butylamine with Aliphatic Alcohols”, *J. Chem. Eng. Data.*, 45 (2000) 606-609.
- [51] B. Gonzalez, A. Dominguez, and J. Tojo Dynamic “Viscosities, Densities, and Speed of Sound and Derived Properties of the Binary Systems Acetic Acid with Water, Methanol, Ethanol, Ethyl Acetate and Methyl Acetate at $T = (293.15, 298.15, \text{ and } 303.15)$ K at Atmospheric Pressure”, *J. Chem. Eng. Data.*, 49 (2004) 1590-1596.
- [52] A. J. Queimada, I. M. Marrucho, J. A. P. Coutinho, and E. H. Stenby, “Viscosity and Liquid Density of Asymmetric *n*-Alkane Mixtures: Measurement and Modeling”, *Int. J. Thermophys.*, 26 (2005) 47-61.
- [53] H. Djojoputro and S. Ismadji, “Density and Viscosity for a Binary Mixture of *cis*-3-Hexenyl Formate, Butyl Acetate, *trans*-2-Hexenyl Acetate, and *cis*-3-Hexenyl

Acetate with Ethanol at Several Temperatures”, *J. Chem. Eng. Data.*, 50(2005) 1009-1013.

[54] P. S. Nikam, T. R. Mahale, and M. Hasan, “Density and Viscosity of Binary Mixtures of Ethyl Acetate with Methanol, Ethanol, Propan-1-ol, Propan-2-ol, Butan-1-ol, 2-Methylpropan-1-ol, and 2-Methylpropan-2-ol at (298.15, 303.15, and 308.15) K”, *J. Chem. Eng. Data.*, 41(1996) 1055-1058.

[55] N. V. Sastry, R. Patel, S. Soni, “Excess molar volumes, excess isentropic compressibilities, excess viscosities, relative permittivity and molar polarization deviations for methyl acetate+, ethyl acetate+, butyl acetate+, isoamyl acetate+, methyl propionate+, ethyl propionate+, ethyl butyrate+, methyl methacrylate+, ethyl methacrylate+, and butyl methacrylate + cyclohexane at $T = 298.15$ and 303.15 K”, *J. Mol. Liq.*, 183 (2013) 102–112.

[56] L. Moravkova, Z. Wagner, J. Linek, “ (ρ, V_m, T) measurements of (cyclohexane + nonane) at temperatures from 298.15 K to 328.15 K and at pressures up to 40 MPa”, *J. Chem. Thermodyn.*, 39 (2007) 1637–1648.

[57] J. A. Amorim, O. Chiavone-Filho, L. L. Paredes, and K. Rajagopal, “High-Pressure Density Measurements for the Binary System Cyclohexane + *n*-Hexadecane in the Temperature Range of (318.15 to 413.15) K”, *J. Chem. Eng. Data.*, 52 (2007) 613-618.

[58] M. I. Aralaguppi, C. V. Jadar, and T. M. Aminabhavi, “Density, Refractive Index, Viscosity, and Speed of Sound in Binary Mixtures of Cyclohexanone with Hexane, Heptane, Octane, Nonane, Decane, Dodecane, and 2,2,4-Trimethylpentane”, *J. Chem. Eng. Data.*, 44 (1999) 435-440.

[59] N. V. Sastry and S. R. Patel, “Excess volumes and dielectric properties for (methyl methacrylate + a branched alcohol) at $T = 298.15$ K and $T = 308.15$ K”, *J. Chem. Thermodyn.*, 32 (2000) 1669–1682.

[60] Hiroyuki Ohji, Katsutoshi Tamura, Hideo Ogawa, “Excess thermodynamic properties of (2-ethoxyethanol C 1,4-dioxane or 1,2-dimethoxyethane) at temperatures between (283.15 and 313.15) K”, *J. Chem. Thermodyn.*, 32 (2000) 319–328.

[61] Jyoti N. Nayak, Mrityunjaya I. Aralaguppi, and Tejraj M. Aminabhavi, “Density, Viscosity, Refractive Index, and Speed of Sound for the Binary Mixtures of Ethyl Chloroacetate with *n*-Alkanes (C₆ to C₁₂) at (298.15, 303.15, and 308.15) K”, *J. Chem. Eng. Data.*, 46 (2001) 891-896.

- [62] J. George, N.V. Sastry, D.H.L. Prasad, "Excess molar enthalpies and excess molar volumes of methyl methacrylate + benzene, + toluene, + *p*-xylene, + cyclohexane and + aliphatic diethers (diethyl, diisopropyl and dibutyl) ", ***Fluid Phase Equilibri.***, 214 (2003) 39–51.
- [63] S. Ajmani, S. C. Rogers, M. H. Barley, and D. J. Livingstone, "Application of QSPR to Mixtures", ***J. Chem. Inf. Model.***, 46 (2006) 2043-2055.
- [64] I. Oprisiu, S. Novotarskyi and I. V. Tetko, "Modeling of non-additive mixture properties using the Online CHEmical database and Modeling environment (OCHEM) ", ***J. Cheminf.***, (2013) 2-7.
- [65] S. Ajmani, S. C. Rogers, M. H. Barley, A. N. Burgess, and D. J. Livingstone, "Characterization of Mixtures. Part 2: QSPR Models for Prediction of Excess Molar Volume and Liquid Density Using Neural Networks", ***Mol. Inf.***, 29 (2010) 645 – 653.
- [66] Todeschini, R., & Consonni, V. (2009). "Molecular descriptors for chemoinformatics". John Wiley & Sons.

Abstract

In this work, a quantitative structure-property relationship (QSPR) approach was employed to predict the density of organic liquid mixtures over a wide range of temperatures and pressures. Two types of descriptors for mixture components have been used for calculating of mixture descriptors: group contribution based descriptors and theoretically derived descriptors. In the first type, the descriptors of mixture components are selected based on group contribution method (GCM). After analyzing the chemical structure of all studied mixtures component in this work, 15 functional groups were recognized. In the second method, the 1481 descriptors of individual components of the mixture were calculated using dragon software. After calculating of descriptors for mixture components with two methods, the mixture descriptors were calculated as mole-weighted sums using the descriptor value and mole fraction of each pure component in the mixture. Therefore, 15 descriptors were obtained by GCM for mixtures. However, the best calculated descriptors from a large number of theoretically derived descriptors of mixtures may be selected by genetic algorithm based on partial least square (GA-PLS) method. 22 theoretically derived descriptors have been selected by GA-PLS method. The selected descriptors with two feature selection methods and 2 experimental variables (temperature and pressure) as well as total mass of mixture were used as input nodes for constructing different artificial neural network (ANNs). 4 feed forward artificial neural networks were designed by combination of Levenberg-Marquardt (LM) 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 Bayesian regularized 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 mixtures accurately.

The mean square error (MSE) and determination coefficient (R^2) for the test set are 11.051 and 0.9996 respectively.

Keywords: quantitative structure–property relationship (QSPR), artificial neural network (ANN), mixture density, Genetic algorithm (GCM), Genetic algorithm (GA).



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

Prediction of density for mixture of organic compounds using nonlinear methods

Bentolhoda Goodarzi

Supervisors:
Dr. Z. Kalantar

Advisor:
Dr. N. Goodarzi

January 2017