

صَلَاةُ الْإِسْلَامِ



دانشگاه صنعتی شاهرود

دانشکده‌ی شیمی

گروه شیمی فیزیک

عنوان:

**مطالعه‌ی ارتباط کمی ساختار-خاصیت برای دانسیته‌ی هیدروکربن‌ها در محدوده‌ی  
وسعی از دما و فشار**

دانشجو:

منصوره متولی

اساتید راهنما:

دکتر زهرا کلانتر

دکتر محسن سرگلزایی

پایان‌نامه جهت اخذ درجه‌ی کارشناسی ارشد

دی ۹۶

شماره: ۱۷۱۰ د. س  
تاریخ: ۹۶/۱۱/۱۷

باسمه تعالی



مدیریت تحصیلات تکمیلی

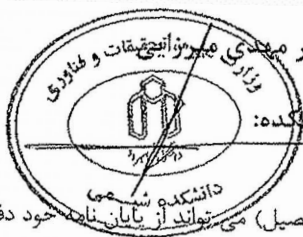
### فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای منصوره متولی با شماره دانشجویی ۹۴۱۵۶۴۴ رشته شیمی گرایش شیمی فیزیک تحت عنوان مطالعه‌ی ارتباط کمی ساختار- خاصیت برای دانسیته‌ی هیدروکربن‌ها در محدوده‌ی وسیعی از دما و فشار که در تاریخ ۹۶/۱۰/۲۶ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می‌گردد:

قبول (با درجه: .....):  مردود

نوع تحقیق: نظری  عملی

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنمای اول	دکتر زهرا کلاتر	استادیار	
۲- استاد راهنمای دوم	دکتر محسن سرگلزایی	استادیار	
۳- استاد مشاور	-	-	-
۴- نماینده تحصیلات تکمیلی	دکتر اسماعیل سلیمانی	دانشیار	
۵- استاد ممتحن اول	دکتر حسین نیکوفرد	دانشیار	
۶- استاد ممتحن دوم	دکتر زینب موسوی تکیه	استادیار	



نام و نام خانوادگی رئیس دانشکده: دکتر مهدی میرزوانیچقیقات و فناوری

تاریخ و امضاء و مهر دانشکده:

تیسره: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در مدت مجاز تحصیل) می‌تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه برگزار شود).

تقدیم به:

پدر و مادرم

که از نگاهشان صلابت

از رفتارشان محبت

و از صبرشان ایستادگی را آموختم

و برادر عزیزم

همراه، همیشگی و پشتیبانی زندگیم

## شکر و قدردانی

نخستین سپاس ستایش از آن خداوندی است که بنده ی کوچکش را در دریای میکران اندیشه، قطره ای ساخت تا وسعت آن را از دریچه ی اندیشه های ناب آموزگاران بزرگ به تماشا نشیند.

اینک که با لطف خداوند متعال تدوین این پایان نامه را به اتمام رسانده ام، لازم می دانم از استاد ارجمند و عالی قدرم، سرکار خانم دکتر زهرا کلاسر که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ گلی در این عرصه بر من دریغ ننمودند، شکر و قدردانی نمایم. با شکر فراوان از استاد گرانقدرم، جناب آقای دکتر سرگلزایی که با نظرات و راهنمودهایشان مرا در انجام این پایان نامه یاری نمودند.

در پایان نیز از تمامی دوستان خوبم که در این مدت با محبت و و همی بایشان همراه من بودند کمال شکر را دارم.

## تعهد نامه

اینجانب منصوره متولی دانشجوی دوره کارشناسی ارشد رشته شیمی فیزیک دانشکده شیمی دانشگاه صنعتی شاهرود نویسنده پایان نامه مطالعه‌ی ارتباط کمی ساختار-خاصیت برای دانسیته هیدروکربن‌ها در محدوده‌ی وسیعی از دما و فشار تحت راهنمایی دکتر زهرا کلانتر و دکتر محسن سرگلزایی 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 شاهرود » و یا « University of Shahrood » 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 اخلاقی انسانی رعایت شده است.

### تاریخ

### امضای دانشجو

### 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 کار یک مطالعه‌ی QSPR برای پیش‌بینی دانسیته‌ی هیدروکربن‌های آلیفاتیک اشباع و غیراشباع شامل آلکان‌ها، سیکلوآلکان‌ها، آلکن‌ها، سیکلوآلکن‌ها و همچنین هیدروکربن‌های آروماتیک در محدوده‌ی وسیعی از دما و فشار انجام شده است. اولین توصیف‌کننده‌های انتخاب شده دما ( $T$ )، فشار ( $P$ ) و جرم مولکولی ( $M$ ) بودند زیرا دانسیته به آن‌ها وابسته است. سایر توصیف‌کننده‌ها از روی ساختار شیمیایی ترکیبات با استفاده از دو رویکرد مختلف بدست آمدند. در رویکرد اول، بهترین توصیف‌کننده‌ها بر مبنای روش سهم‌گروه انتخاب شدند. پس از آنالیز ساختار شیمیایی ۷۹ هیدروکربن مورد مطالعه، ۱۵ گروه عاملی مفید پیدا شدند. در روش دوم، تعداد زیادی توصیف‌کننده مولکولی (۳۲۲۴ توصیف‌کننده) به کمک نرم‌افزار دراگون از روی ساختار شیمیایی ترکیبات محاسبه شدند. سپس زیر مجموعه‌ای از ۲۲ گروه توصیف‌کننده‌های دراگون با روش الگوریتم ژنتیک بر مبنای حداقل مربعات جزئی (GA-PLS) به عنوان تکنیک انتخاب ویژگی انتخاب گردیدند. ۱۱ توصیف‌کننده توسط الگوریتم ژنتیک (GA) به عنوان بهترین توصیف‌کننده‌ها بدست آمدند. توصیف‌کننده‌های انتخاب شده توسط دو روش انتخاب ویژگی و ۳ متغیر ( $M$  و  $P$  و  $T$ ) به عنوان گره‌های ورودی برای ساختار شبکه‌های عصبی مصنوعی مختلف استفاده شدند. ۴ نوع شبکه عصبی مصنوعی با ترکیب الگوریتم‌های آموزشی لونبرگ-مارکواریت (LM) و تنظیم بایزین (BR) و دو تابع انتقال لگاریتم سیگموئید و تانژانت سیگموئید طراحی شدند. پارامترهای مختلف شبکه‌های طراحی شده شامل تعداد نرون‌ها در لایه ورودی، لایه مخفی و تعداد دور تکرار بهینه گردیدند. پس از بهینه‌سازی پارامترها برای شبکه‌های عصبی مصنوعی ایجاد شده با استفاده از دو نوع توصیف‌کننده‌ها، عملکرد مدل‌ها توسط تست خارجی مورد ارزیابی قرار گرفت. متوسط مربعات خطا (MSE) و ضریب تعیین ( $R^2$ ) برای سری تست خارجی با استفاده از مدل GCM-ANN بهینه شده به ترتیب ۹/۵۵۶ و ۰/۹۹۸۷ و با استفاده از مدل GA-ANN بهینه شده ۷۰/۱۶۱۲ و ۰/۹۹۸۲ بدست آمد. نتایج حاصل نشان می‌دهد که شبکه‌ی عصبی با الگوریتم آموزشی لونبرگ-مارکواریت و تابع انتقال لگاریتم سیگموئید و توصیف‌کننده‌های

حاصل از روش سهم گروه می تواند به درستی رابطه‌ی بین توصیف‌کننده‌های ساختاری و دانسیته‌ی هیدروکربن‌های مورد نظر را شبیه‌سازی کند.

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



نتایج حاصل از این پایان نامه در دو مقاله تحت عناوین

**Density prediction of aliphatic hydrocarbons using nonlinear group contribution method**

**GA-ANN modeling for density prediction of hydrocarbons over a wide range of temperature and pressure**

در بیستمین کنفرانس شیمی فیزیک ایران در مرداد ماه ۱۳۹۶ ارائه گردید.

## فهرست مطالب

### فصل اول: مقدمه

- ۱-۱- مقدمه..... ۲
- ۲-۱- دانسیته..... ۲
- ۳-۱- روش‌های پیش‌بینی دانسیته‌ی هیدروکربن‌ها..... ۲
- ۱-۳-۱- پیش‌بینی دانسیته به کمک معادلات حالت..... ۳
- ۱-۱-۳-۱- پیش‌بینی دانسیته ترکیبات آلی با استفاده از معادله ی حالت MLIR..... ۳
- ۲-۱-۳-۱- پیش‌بینی دانسیته سیالات خالص با استفاده از معادله حالت GCM-GMA..... ۵
- ۳-۱-۳-۱- پیش‌بینی دانسیته سیالات خالص با استفاده از معادله‌ی حالت سوآو-ردلیچ-کوانگ (SRK-EOS)..... ۷
- ۴-۱-۳-۱- پیش‌بینی دانسیته سیالات خالص با استفاده از معادله حالت پینگ-رابینسون..... ۸
- ۲-۳-۱- پیش‌بینی دانسیته از طریق مطالعات QSPR..... ۹
- ۱-۲-۳-۱- پیش‌بینی دانسیته‌ی آلکان‌های خطی و سیکلو آلکان‌ها به کمک شبکه عصبی موجک..... ۹
- ۲-۲-۳-۱- پیش‌بینی دانسیته هیدروکربن‌ها به کمک شبکه عصبی مصنوعی..... ۱۰
- ۴-۱- هدف تحقیق..... ۱۰

### فصل دوم: کمومتریکس

- ۱-۲- کمومتریکس..... ۱۲
- ۲-۲- ارتباط کمی ساختار-خاصیت و ساختار-فعالیت..... ۱۳
- ۱-۲-۲- رابطه کمی ساختار-فعالیت (QSAR)..... ۱۳
- ۲-۲-۲- رابطه کمی ساختار-خاصیت (QSPR)..... ۱۳
- ۳-۲- مراحل مدل‌سازی QSPR..... ۱۵
- ۱-۳-۲- فراهم کردن سری داده‌ها..... ۱۵

- ۲-۳-۲- رسم و بهینه‌سازی ساختار هندسی مولکول‌ها..... ۱۵
- ۳-۳-۲- محاسبه‌ی توصیف‌کننده‌های تئوری..... ۱۵
- ۴-۳-۲- حذف توصیف‌کننده‌های نامناسب..... ۱۶
- ۵-۳-۲- تقسیم‌بندی داده‌ها..... ۱۸
- ۶-۳-۲- انتخاب بهترین توصیف‌کننده‌ها..... ۱۸
- ۱-۶-۳-۲- انتخاب بهترین توصیف‌کننده‌ها توسط الگوریتم ژنتیک..... ۱۹
- ۱-۶-۳-۲- الف- مقدمه..... ۱۹
- ۱-۶-۳-۲- ب- الگوریتم ژنتیک..... ۱۹
- ۱-۶-۳-۲- ج- نحوه‌ی انتخاب توصیف‌کننده‌ها و شرایط توقف الگوریتم ژنتیک..... ۲۱
- ۷-۳-۲- انتخاب توصیف‌کننده‌ها بر مبنای روش سهم‌گروه (GCM)..... ۲۲
- ۸-۳-۲- ساختن مدل..... ۲۴
- ۱-۸-۳-۲- شبکه عصبی مصنوعی..... ۲۴
- ۲-۸-۳-۲- تاریخچه‌ی شبکه عصبی مصنوعی..... ۲۶
- ۲-۸-۳-۲- الف- ساختار شبکه عصبی مصنوعی..... ۲۸
- ۲-۸-۳-۲- ب- انواع شبکه‌ها از نظر ارتباطات بین نرونی..... ۲۹
- ۲-۸-۳-۲- ج- روش‌های آموزش..... ۳۱
- ۲-۸-۳-۲- د- محاسن شبکه عصبی مصنوعی..... ۳۳
- ۲-۸-۳-۲- ه- معایب شبکه عصبی مصنوعی..... ۳۳
- ۹-۳-۲- ارزیابی قدرت پیش‌بینی مدل..... ۳۴
- ۹-۳-۲- الف- استفاده از پارامترهای آماری..... ۳۴
- ۹-۳-۲- ب- استفاده از نمودار برگشتی..... ۳۷
- ۹-۳-۲- ج- استفاده از نمودار خطای باقیمانده..... ۳۷

۳۸-۳-۹-د- استفاده از سری تست خارجی.....

### فصل سوم: محاسبات

۳-۱- انتخاب سری داده‌ها..... ۴۰

۳-۲- مدل سازی شبکه عصبی..... ۴۵

۳-۲-۱- مدل سازی شبکه عصبی مصنوعی با استفاده از روش سهم گروه..... ۴۵

۳-۲-۱-۱- محاسبه‌ی توصیف کننده‌های هیدروکربن‌های خالص با استفاده از روش سهم گروه..... ۴۵

۳-۲-۱-۲- دسته بندی داده‌ها..... ۴۷

۳-۲-۱-۳- مدل سازی شبکه عصبی مصنوعی با توصیف کننده‌های حاصل از روش سهم گروه -GCM-ANN..... ۴۸

۳-۲-۱-۴- ارزیابی شبکه GCM-ANN بهینه شده..... ۵۴

۳-۲-۲- مدل سازی شبکه عصبی مصنوعی با استفاده از توصیف کننده‌های تئوری..... ۵۸

۳-۲-۲-۱- رسم و بهینه سازی ساختار مولکول‌ها و محاسبه توصیف کننده‌ها..... ۵۸

۳-۲-۲-۲- انتخاب بهترین توصیف کننده‌ها به روش الگوریتم ژنتیک..... ۵۹

۳-۲-۲-۳- بهینه سازی شبکه عصبی مصنوعی با توصیف کننده های الگوریتم ژنتیک..... ۶۲

۳-۲-۲-۴- ارزیابی شبکه‌ی بهینه شده‌ی GA-ANN..... ۶۵

۳-۳- ارزیابی قدرت پیش بینی مدل های غیر خطی بهینه شده برای سری تست خارجی..... ۶۸

۳-۴- استفاده از پارامترهای آماری..... ۷۲

۳-۵- بررسی ارتباط توصیف کننده‌های انتخاب شده در مدل GA-ANN با دانسیته‌ی ترکیبات..... ۷۳

۳-۵-۱- توصیف کننده‌ی X0Av..... ۷۳

۳-۵-۲- توصیف کننده‌های Constitutional..... ۷۴

۳-۵-۳- توصیف کننده‌های GETAWAY..... ۷۶

۳-۵-۴- توصیف کننده‌های توپولوژیکی..... ۷۸

۳-۶- بررسی میزان مشارکت توصیف‌کننده‌های منتخب شبکه‌ی عصبی مصنوعی..... ۸۰

۳-۷- نتیجه‌گیری..... ۸۲

۳-۸- آینده‌نگری..... ۸۴

منابع..... ۸۵

## فهرست اشکال

- شکل (۲-۱) - ساختار یک نرون زیستی..... ۲۶
- شکل (۲-۲) - ساختار یک شبکه عصبی مصنوعی سه لایه..... ۲۹
- شکل (۳-۲) - نمونه‌ای از یک شبکه پیش‌خور..... ۳۰
- شکل (۴-۲) - نمونه‌ای از شبکه پس‌خور برگشتی..... ۳۱
- شکل (۱-۳) - نمودار میانگین مربعات خطا بر حسب  $\mu$  حول نقطه بهینه برای سری ارزیابی..... ۵۴
- شکل (۲-۳) - نمودار برگشتی مقادیر پیش‌بینی شده‌ی دانسیته‌ی هیدروکربن‌ها بر حسب مقادیر تجربی برای (الف) سری ارزیابی و (ب) تست داخلی..... ۵۶
- شکل (۳-۳) - نمودار خطای باقیمانده بر حسب مقادیر تجربی دانسیته‌ی هیدروکربن‌ها (الف) سری ارزیابی و (ب) تست داخلی..... ۵۷
- شکل (۴-۳) - نمودار میانگین مربعات خطای حاصل از پیش‌بینی دانسیته سری ارزیابی بر حسب پارامتر  $\mu$  حول نقطه‌ی بهینه..... ۶۴
- شکل (۵-۳) - نمودار برگشتی مقادیر پیش‌بینی شده‌ی دانسیته‌ی هیدروکربن‌ها بر حسب مقادیر تجربی برای (الف) سری ارزیابی و (ب) تست داخلی..... ۶۶
- شکل (۶-۳) - نمودار خطای باقیمانده برای پیش‌بینی دانسیته‌ی هیدروکربن‌ها بر حسب مقادیر تجربی دانسیته برای (الف) سری ارزیابی و (ب) تست داخلی..... ۶۷
- شکل (۷-۳) - نمودار برگشتی مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست خارجی برای مدل‌های بهینه شده‌ی (الف) GCM-ANN و (ب) GA-ANN..... ۶۹
- شکل (۸-۳) - نمودار خطای باقیمانده مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست خارجی برای مدل‌های بهینه شده‌ی (الف) GCM-ANN و (ب) GA-ANN..... ۷۰
- شکل (۹-۳) - مولکول هیستیدین و گراف مولکولی آن. مقادیر احتمالی محاسبه شده مربوط به هر پیوند از رأس  $C_{\alpha}$ ..... ۷۹
- شکل (۱۰-۳) - درصد مشارکت توصیف‌کننده‌ها در مدل GA-ANN..... ۸۱

## فهرست جدول‌ها

- جدول (۱-۲) - انواع توصیف‌کننده‌های محاسبه شده توسط نرم افزار Dragon.....۱۷
- جدول (۱-۳) - نام ترکیبات، محدوده‌ی دما و فشار و تعداد نقاط مورد استفاده برای هر ترکیب به همراه مرجع مورد استفاده برای هر یک از آنها.....۴۰
- جدول (۲-۳) - گروه‌های عاملی موجود در اجزای هیدروکربن‌های مورد مطالعه در این پایان‌نامه طبق روش سهم‌گروه.....۴۶
- جدول (۳-۳) - مقادیر همبستگی توصیف‌کننده‌های روش سهم‌گروه با دانسیته.....۴۷
- جدول (۴-۳) - 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.....۴۹
- جدول (۵-۳) - 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.....۵۰
- 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.....۵۱
- جدول (۷-۳) - 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.....۵۲
- جدول (۸-۳) - توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN.....۵۳
- جدول (۹-۳) - توابع و مقادیر پارامترهای شبکه‌ی GCM-ANN بهینه شده برای پیش‌بینی دانسیته‌ی هیدروکربن‌ها بر حسب دما و فشار.....۵۴
- جدول (۱۰-۳) - تعداد دفعات تکرار توصیف‌کننده‌ها در روش انتخابی الگوریتم ژنتیک پس از ۱۶ بار اجرا.....۵۹
- جدول (۱۱-۳) - طبقه و نام کامل توصیف‌کننده‌های تئوری انتخابی توسط الگوریتم ژنتیک.....۶۰
- جدول (۱۲-۳) - همبستگی بین توصیف‌کننده‌های انتخابی الگوریتم ژنتیک.....۶۱
- جدول (۱۳-۳) - مقادیر همبستگی توصیف‌کننده‌های تئوری انتخابی با روش GA با دانسیته.....۶۲
- جدول (۱۴-۳) - توابع و پارامترهای بهینه‌ی شبکه GA-ANN.....۶۳

- جدول (۳-۱۵) - توابع و پارامترهای بهینه شده شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های تئوری انتخابی GA ..... ۶۴
- جدول (۳-۱۶) - پارامترهای آماری مدل‌های بهینه شده‌ی GCM-ANN و GA-ANN برای دو سری ارزیابی و تست ..... ۷۲
- جدول (۳-۱۷) - مثال‌هایی از مقدار توصیف‌کننده X0Av در تعدادی از ترکیبات ..... ۷۴
- جدول (۳-۱۸) - مثال‌هایی از مقدار توصیف‌کننده‌های AMW و Ms در تعدادی از ترکیبات ..... ۷۶
- جدول (۳-۱۹) - مثال‌هایی از مقادیر توصیف‌کننده R7u در تعدادی از ترکیبات ..... ۷۸
- جدول (۳-۲۰) - مثال‌هایی از مقادیر توصیف‌کننده PW3، PW4 و GGI3 در تعدادی از ترکیبات ..... ۸۰



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 در شرایط مختلف دما و فشار برای طراحی بهینه‌ی واحدهای عملیاتی صنایع شیمیایی مختلف که در آن‌ها انتقال جرم و انتقال گرما وجود دارد ضروری است. با وجود اهمیت دانسیته‌ی سیالات خالص در صنایع شیمیایی، اندازه‌گیری تجربی دانسیته‌ی سیالات خالص در شرایط دما و فشار گوناگون مشکل می‌باشد [۲۱]. در این شرایط، استفاده از روش‌های پیش‌بینی دانسیته از اهمیت ویژه‌ای برخوردار هستند.

## ۱-۲- دانسیته

نسبت جرم به حجم یک ماده در دما و فشار مشخص دانسیته نامیده می‌شود. واحد دانسیته در سیستم SI،  $\text{kg m}^{-3}$  است، اما واحدهای  $\text{g cm}^{-3}$  یا  $\text{g ml}^{-1}$  نیز ممکن است استفاده شوند. داده‌های دانسیته برای مواد مختلف در فشار یک بار و دمای ۲۹۸/۱۵ کلوین در کتاب‌های مرجع برای داده‌های ترمودینامیکی گزارش شده است [۳].

## ۱-۳- روش‌های پیش‌بینی دانسیته‌ی هیدروکربن‌ها

همان‌طور که در مقدمه شرح داده شد خواص ترمودینامیکی سیالات خالص، ابزار مهمی برای طراحی مهندسی و آنالیز فرآیندهای شیمیایی است. اگرچه داده‌های تجربی دقیق هستند اما اندازه‌گیری آن‌ها در یک دامنه وسیع دما و فشار کاری وقت‌گیر و هزینه‌بر است. لذا باید از روش‌های پیش‌بینی استفاده نمود. مهم‌ترین روش‌های پیش‌بینی دانسیته استفاده از معادلات حالت و مطالعات QSPR می‌باشد که در ادامه راجع به هر یک توضیح داده خواهد شد.

### ۱-۳-۱- پیش بینی دانسیته به کمک معادلات حالت

یکی از مهم‌ترین روش‌های پیش‌بینی دانسیته‌ی سیالات خالص استفاده از معادلات حالت است. معادله‌ی حالت یک رابطه‌ی ریاضی است که فشار را به صورت تابعی از دما و دانسیته بیان می‌کند. با استفاده از معادله‌ی حالت و یک مشتق مناسب از آن می‌توان سایر خواص ترمودینامیکی را بدست آورد. چهار معادله‌ی حالتی که در سال‌های اخیر برای پیش‌بینی خواص ترمودینامیکی هیدروکربن‌ها برحسب دما و فشار معرفی شده‌اند و از کارایی بالایی برخوردارند معادلات حالت MLIR<sup>۱</sup> و GCM-<sup>۲</sup> و SRK<sup>۳</sup> و PR<sup>۴</sup> هستند. در ادامه به‌طور مختصر راجع به هر کدام از آن‌ها توضیح داده خواهد شد.

### ۱-۱-۳-۱- پیش‌بینی دانسیته ترکیبات آلی با استفاده از معادله‌ی حالت MLIR

قاعده‌ی همدمای خطی (LIR) قاعده‌ای است که در سال ۱۹۹۳ برای سیالات چگال زیر بحرانی و فوق بحرانی براساس مفهوم متوسط پتانسیل جفت موثر (AEPP) به دست آمده و دارای مبنای نظری است. این قاعده بیان می‌کند که کمیت  $(Z-1)v^2$  بر حسب  $\rho^2$  برای هر همدمای سیال چگال خطی است، که  $\rho = \frac{1}{v}$  دانسیته‌ی مولی سیال است.

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

وابستگی دمایی پارامترهای  $A$  و  $B$  نیز به صورت زیر آمده است:

- 
- 1-Modified Linear Isotherm Regularity (MLIR)
  - 2- Group Contribution Method-Goharshadi-Morsali-Abbaspour (GCM-GMA)
  - 3-Soave- Redlich- Kwong
  - 4-Peng-Rabinson
  - 5- Linear Isotherm Regularity (LIR)
  - 6- Average Effective Pair Potential (AEPP)

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

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

رفتار خطی معادله‌ی حالت LIR، برای سیالات چگال قطبی و غیرقطبی با مولکول‌های کوچک در محدوده‌های  $T > 2T_B$  و  $\rho > \rho_B$  نشان می‌دهد که  $T_B$  و  $\rho_B$  به ترتیب دانسیته‌ی بویل و دمای بویل سیال می‌باشند [۴]. در سال ۲۰۰۳ پارسافر 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 [۵]. آن‌ها برای ارائه‌ی این معادله، هر ترکیب آلی را به صورت یک مخلوط فرضی از گروه‌های عاملی تشکیل دهنده‌اش در نظر گرفتند که تصادفی در مخلوط توزیع شده‌اند. سپس طبق تقریب تک سیال و اندروالس، معادله‌ی حالت LIR را عیناً برای چنین مخلوط فرضی در نظر 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 دمای  $T$  و فشار  $p$ ، دانسیته‌ی مولی سیال  $\rho$  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:

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

که در آن  $A_m$  و  $B_m$  پارامترهای LIR به ازای یک گروه در سیال مورد نظر هستند. سپس آن‌ها با تعریف دو پارامتر  $A'$  و  $B'$  به صورت معادله‌های (۵-۱) و (۶-۱)،

$$A' = A_m n^2 \quad (5-1)$$

$$B' = B_m n^4 \quad (6-1)$$

معادله‌ی (۷-۱) را به صورت کلی زیر تعریف نمودند و آن را قاعده‌ی همدمای خطی اصلاح شده (MLIR) نامیدند.

$$\left(\frac{Z}{n}-1\right)v^2 = A' + B'\rho^2 \quad (7-1)$$

طبق این معادله نمودار  $\left(\frac{Z}{n}-1\right)v^2$  بر حسب  $\rho^2$  برای هر همدمای ترکیب آلی چگال، خطی است. برای تمامی ترکیبات آلی مورد مطالعه شامل آلکان‌های خطی، الکل‌های نوع اول و دوم و سوم، کتون‌ها، اسیدهای کربوکسیلیک، آمین‌ها و استرها [۶-۸] 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.

### ۱-۳-۲- پیش‌بینی دانسیته سیالات خالص با استفاده از معادله حالت GCM-

#### GMA

در سال ۲۰۰۵ گوهرشادی و همکارانش بر اساس مفهوم متوسط انرژی پتانسیل جفت مؤثر، توانستند معادله حالت عمومی GMA را برای پیش‌بینی خواص ترمودینامیکی سیالات چگال ارائه کنند [۹]. این معادله حالت بیان می‌کند که، نمودار  $(2Z-1)v_m^3$  بر حسب  $\rho$  برای هر همدمای سیال چگال خطی است که  $\rho = 1/v_m$  دانسیته‌ی مولی سیال است.

$$(2Z-1)v_m^3 = A + B\rho \quad (8-1)$$

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

$$A = A_0 - \frac{2A_1}{RT} + \frac{2A_2 \ln T}{R} \quad (9-1)$$

$$B = B_0 - \frac{2B_1}{RT} + \frac{2B_2 \ln T}{R} \quad (10-1)$$

در روابط (۹-۱) و (۱۰-۱) 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،  $\rho_c$  و  $T_c$  به ترتیب دانسیته و دمای بحرانی سیال هستند [۱۰-۱۳].

موسوی در سال ۲۰۱۱ توانست از ترکیب روش سهم‌گروه با معادله حالت GMA که توسط گوهرشادی و همکارانش ارائه شده بود، دانسیته‌ی ترکیبات آلی مختلف و مخلوط‌های آن‌ها را پیش‌بینی کند [۱۴].

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

$$\left(\frac{2Z}{n} - 1\right)v_m^3 = A_m n^3 + B_m n^4 \rho \quad (11-1)$$

در این معادله که GCM-GMA نامیده شده،  $n$  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.

### ۱-۳-۱-۳- پیش‌بینی دانسیته سیالات خالص با استفاده از معادله‌ی حالت

#### سوآو-ردلیچ-وانگ (SRK-EOS)

معادله حالت SRK که در سال ۱۹۴۹ ارائه شده در مقایسه با سایر معادلات حالت ساده، از دقت بیشتری برخوردار است [۱۵]. فرم کلی معادله‌ی حالت SRK به صورت زیر است:

$$p = \frac{RT}{v-b} - \frac{a(T)}{v(v+b)} \quad (12-1)$$

که  $R$  ثابت گازها،  $v$  حجم مولی،  $T$  دما و  $P$  فشار و  $a$  و  $b$  ثوابت قابل محاسبه می‌باشد. در رابطه فوق ضرایب  $a$  و  $b$  با استفاده از روابط زیر تعریف می‌شوند:

$$a = 0.42748 \frac{R^2 T_c^2}{P_c} \alpha(T_r)$$

$$b = 0.08664 \frac{RT_c}{P_c}$$

$$\alpha(T_r) = \left[ 1 + c_1 \left( 1 - (T/T_c)^{0.5} \right) \right]^2$$

$$c_1 = 0.480 + 1.574\omega - 0.176\omega^2$$

صحت معادله‌ی حالت SRK جهت تخمین خواص ترمودینامیکی سیالات خالص در مقایسه با داده‌های تجربی مورد تایید قرار گرفته است. از این معادله برای محاسبه‌ی سریع خواص ترمودینامیکی سیالات و پیش‌بینی رفتار فیزیکی و خواص حجمی مایعات در طراحی‌های مربوط به نفت و گاز استفاده می‌شود.

### ۱-۳-۱-۴- پیش بینی دانسیته سیالات خالص با استفاده از معادله حالت پینگ-

#### رابینسون

پینگ و رابینسون با چند هدف معادله‌ی توسعه یافته‌ی دو پارامتری مکعبی خود را ارائه کردند:

۱- محاسبه‌ی پارامترها بر حسب خصوصیات بحرانی و فاکتور شکل  $(\omega)$

۲- محاسبه‌ی فاکتور تراکم‌پذیری و چگالی مایع در نزدیکی نقطه‌ی بحرانی

۳- قابلیت کاربرد برای محاسبه‌ی همه خصوصیات سیالات

فرم کلی معادله‌ی حالت پینگ-رابینسون [۱۶] به صورت زیر است:

$$p = \frac{RT}{v-b} - \frac{a(T)}{[v^2 + 2vb - b^2]} \quad (13-1)$$

که  $R$  ثابت گازها،  $v$  حجم مولی،  $T$  دما،  $p$  فشار و  $a$  و  $b$  ثابت‌های معادله می‌باشند. پارامترهای  $a$  و  $b$  عبارتند از:

$$a(T) = a(T_c) \cdot \alpha(T_r)$$

$$a(T_c) = -0.45724 \frac{R^2 T_c^2}{p_c}$$

$$\alpha(T_r, \omega) = [1 + (0.37464 + 1.54226\omega - 0.26992\omega^2)(1 - \sqrt{T_r})]^2$$

$$b = 0.0778 \frac{RT_c}{p_c}$$

---

1- cubic  
2- Acentric factor



## ۱-۳-۲- پیش‌بینی دانسیته از طریق مطالعات QSPR

مهم‌ترین مشکل استفاده از معادلات حالت وجود ثابت‌های وابسته به نوع ماده در معادله است. این بدان معنی است که اگر برای ماده‌ای داده‌های تجربی وجود نداشته باشد، ثابت‌های معادله نیز وجود ندارد و در نتیجه معادله حالت برای ماده مورد نظر قابل استفاده نیست. راه‌حل خوبی که در سال‌های اخیر برای حل این مشکل ارائه شده مطالعات QSPR است. این دسته از روش‌های پیش‌بینی می‌توانند بین ساختار ماده و خواص فیزیکی شیمیایی آن ارتباط برقرار کنند. در این نوع مطالعات از دو نوع توصیف‌کننده‌ی ساختاری استفاده می‌شود: روش سهم‌گروه و توصیف‌کننده‌های تئوری. در این صورت برای موادی که برای آن‌ها اندازه‌گیری انجام نشده یا حتی سنتز نشده‌اند نیز می‌توان توصیف‌کننده‌ها را محاسبه و خواص آن‌ها را پیش‌بینی نمود.

## ۱-۳-۲-۱- پیش‌بینی دانسیته‌ی آلکان‌های خطی و سیکلو آلکان‌ها به کمک

### شبکه عصبی موجک

در سال ۲۰۰۹ شعبانی و کلانتر یک مدل شبکه عصبی موجک با تعداد لایه‌ی ورودی ۵، تعداد گره لایه‌ی پنهان ۸ و خروجی ۱ بر مبنای روش سهم‌گروه برای پیش‌بینی دانسیته آلکان‌های خطی و سیکلو آلکان‌ها ارائه نمودند که می‌توانست دانسیته این ترکیبات را با متوسط خطای نسبی کمتر از ۱٪ پیش‌بینی کند. در این مدل از دو متغیر دما و فشار و سه متغیر (تعداد گروه‌های متیل، متیلن و متین) بر مبنای ساختار مولکول‌های مورد مطالعه جهت مدل‌سازی استفاده شده است [۱۷].

### ۱-۳-۲-۲- پیش‌بینی دانسیته هیدروکربن‌ها به کمک شبکه عصبی مصنوعی

در سال ۲۰۱۳ موسوی و همکارانش یک مدل شبکه عصبی مصنوعی برای پیش‌بینی دانسیته‌ی آلکان‌های خطی و غیرخطی، سیکلو آلکان‌ها و هیدروکربن‌های آروماتیک تک حلقه‌ای ارائه کرده‌اند. آن‌ها در این کار با شناسایی ۹ گروه عاملی بر مبنای روش سهم‌گروه‌ها و سه متغیر دما، فشار و جرم مولی ترکیبات توانستند یک شبکه عصبی مصنوعی با تعداد لایه‌ی ورودی ۱۲، تعداد گره لایه‌ی پنهان ۲۱ و خروجی ۱ ارائه کنند که دانسیته را با متوسط خطای نسبی کمتر از ۰/۴٪ پیش‌بینی می‌کرد [۱۸].

### ۱-۴- هدف تحقیق

در این پایان‌نامه پیش‌بینی دانسیته‌ی طیف وسیعی‌تری از هیدروکربن‌ها شامل آلکان‌های خطی و غیرخطی، سیکلو آلکان‌ها، آلکن‌ها، سیکلو آلکن‌ها و ترکیبات آروماتیک تک حلقه‌ای و ۲ حلقه‌ای از ۵ تا ۴۰ کربن بر حسب دما و فشار با استفاده از توصیف‌کننده‌های ساختاری این ترکیبات و روش شبکه‌ی عصبی مصنوعی مورد مطالعه قرار گرفت. توصیف‌کننده‌های ساختاری در این ترکیبات با استفاده از دو روش سهم‌گروه و توصیف‌کننده‌های تئوری محاسبه شده بر مبنای ساختار مولکولی ترکیبات استخراج شد.

فصل دوم

# کمومتریکس

## ۲-۱- کمومتریکس

نام کمومتریکس اولین بار توسط دانشمند سوئدی جوان، اسوانت ولد<sup>۱</sup> در اوایل سال ۱۹۷۰ که در زمینه‌ی شیمی فیزیک آلی فعالیت داشت، مطرح گردید. همکاری او با کووالسکی<sup>۲</sup> که در آن زمان روی روش‌های تشخیص الگو در شیمی کار می‌کرد منتهی به پایه‌گذاری انجمن بین‌المللی کمومتریکس<sup>۳</sup> (ICS) در سال ۱۹۷۴ شد. چندین تعریف برای کمومتریکس بیان شده است 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) فرآیندهای بهینه را طراحی و انتخاب کند، (ب) ماکزیمم اطلاعات شیمیایی مناسب را به وسیله آنالیز اطلاعات شیمیایی فراهم کند و (ج) اطلاعات بیشتری در مورد سیستم‌های شیمیایی در اختیار قرار دهد [۱۹]."

با توجه به رشد سریع تجهیزات مورد استفاده در شیمی و حجم بسیار زیاد داده‌ها، هنگام پردازش، تفسیر اطلاعات و استخراج نتایج مفید از آن‌ها، نیاز به کامپیوتر امری اجتناب‌ناپذیر است. از طرف دیگر شیمی‌دانان، گاهی اوقات با موادی سر و کار دارند که بسیار گران، سمی و خطرناک بوده یا در مواردی به راحتی قابل دسترس نیستند. در این موارد می‌توان از روش‌های ریاضی و آمار جهت توصیف و توجیه نتایج آزمایش‌های مختلف استفاده نمود [۲۰]. می‌توان گفت که موارد فوق مهم‌ترین دلایل افزایش کاربرد روش‌های کمومتریکس نزد شیمی‌دانان است.

ظهور زمینه‌ی تحقیقاتی کمومتریکس در ایران به سال ۱۹۹۷ برمیگردد. اولین تحقیق درباره کمومتریکس در ایران در سال ۱۹۹۸ منتشر شد [۲۱].

---

1-Svante wold

2-Kowalski

3-International Chemometrics Society

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

### ۲-۲-۱- رابطه کمی ساختار-فعالیت<sup>۱</sup> (QSAR)

یکی از کاربردهای خاص روش‌های کمومتریکس در رابطه با توسعه‌ی ارتباط کمی ساختارفعالیت (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 و ویژگی‌های مولکولی دارد تا بتواند از این قواعد برای ارزیابی فعالیت ترکیبات جدید استفاده کند.

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 فعالیت مولکول چه عواملی باید بهینه شوند؟

دلیل استفاده از کلمه کمی<sup>۲</sup> مدل ریاضی است که برای بررسی فعالیت مورد استفاده قرار می‌گیرد [۲۲].

### ۲-۲-۲- رابطه کمی ساختار-خاصیت<sup>۳</sup> (QSPR)

یکی دیگر از زمینه‌های مهم کاربرد کمومتریکس، مطالعاتی است که خواص 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 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 خواص نظیر نقطه‌ی جوش و ذوب، فشار بخار و غیره. نتایج این مطالعات علاوه بر شفاف سازی نحوه‌ی ارتباط

---

1-Quantitative Structure-Activity Relationship

2-Quantitative

3-Quantitative Structure-Property Relationship

خواص مولکول‌ها و ویژگی‌های ساختمانی آن‌ها با خاصیت مورد نظر، 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<sup>۱</sup> (OLS)، تحلیل مؤلفه‌های اساسی<sup>۲</sup> (PCA)، رگرسیون خطی چندگانه<sup>۳</sup> (MLR)، رگرسیون اجزای اصلی<sup>۴</sup> (PCR) و حداقل مربعات جزئی<sup>۵</sup> (PLS)، اشاره نمود. روش‌های دیگری مانند شبکه عصبی<sup>۶</sup> (ANN) و ماشین بردار پشتیبان<sup>۷</sup> (SVM)، ارتباط غیر خطی میان ساختار و خواص ترکیبات را مورد مطالعه قرار می‌دهند.

به‌طور کلی مطالعات QSPR و همچنین QSAR شامل مراحل زیر است [۲۳]:

- ۱- فراهم کردن سری داده‌ها
- ۲- رسم و بهینه‌سازی ساختار هندسی مولکول‌ها
- ۳- محاسبه توصیف‌کننده‌ها
- ۴- حذف توصیف‌کننده‌های نامناسب
- ۵- تقسیم‌بندی سری داده‌ها به سری‌های آموزش، ارزیابی و تست
- ۶- آنالیز آماری مدل‌ها و انتخاب بهترین مدل
- ۷- ارزیابی اعتبار مدل‌های انتخاب شده
- ۸- پیش‌بینی خاصیت مورد نظر

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

---

1-Ordinary Least Square  
2-Principle Component Analysis  
3-Multiple Linear Regression  
4-Principal Component Regression  
5-Partial Least Squares  
6-Artificial Neural Network  
7-Support Vector Machine

## ۲-۳-۳- مراحل مدل سازی QSPR

### ۲-۳-۱- فراهم کردن سری داده‌ها

در یک مطالعه QSPR اولین مرحله جمع‌آوری و انتخاب سری داده‌هاست که خاصیت مورد مطالعه برای آن‌ها در شرایط عملی یکسانی به دست آمده باشد. هر چي سری داده‌ها وسیع‌تر و متنوع‌تر باشد، مدل به دست آمده مدل کلی‌تر و معتبرتری بوده و محدود به یک گروه خاص از ترکیبات نخواهد بود. داده‌های جمع‌آوری شده می‌توانند در آزمایشگاه تولید و یا از طریق مقادیر منتشر شده در مقالات استخراج شوند.

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

ایجاد توصیف‌کننده‌های هندسی بر مبنای ساختار و هندسه‌ی دقیق مولکولی استوار است. اگر ساختارها به صورت صورت‌بندی با حداقل انرژی نباشند مقادیر غیر صحیحی برای توصیف‌کننده‌ها ایجاد می‌شود. در این مرحله، ساختار ترکیبات مربوط به سری داده‌ها با استفاده از نرم افزارهای تجاری در دسترس رسم شده و سپس توسط روش‌های کوانتوم-مکانیکی مختلف نظیر AM1، PM3، MNDO و.... بهینه می‌شود [۲۴]. نرم افزاری که برای این منظور معمولاً مورد استفاده قرار می‌گیرد نرم افزار Hyper Chem است.

### ۲-۳-۳- محاسبه‌ی توصیف‌کننده‌های تئوری

توصیف‌کننده‌ها پارامترهای 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 مورد نظر مؤثر باشد [۲۵]. این توصیف‌کننده‌ها به دو گروه کلی توصیف‌کننده‌های تجربی و تئوری تقسیم‌بندی می‌شوند.

برخی از ویژگی‌های یک توصیف‌کننده‌ی مناسب عبارت است از:

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

توصیف‌کننده‌های مولکولی به دو دسته اصلی تقسیم‌بندی می‌شوند: توصیف‌کننده‌های حاصل از اندازه‌گیری‌های تجربی (مانند قطبش‌پذیری و ممان دوقطبی) و توصیف‌کننده‌های تئوری که با استفاده از ساختار مولکول و بدون نیاز به داده‌های تجربی محاسبه می‌شوند که باعث صرفه جویی در وقت، هزینه، مواد و تجهیزات خواهند شد و برای هر مولکول واقعی یا فرضی که هنوز سنتز نشده‌اند، در دسترس می‌باشند [۲۶]. در این تحقیق ۳۲۲۴ توصیف‌کننده‌ی تئوری توسط نرم افزار درآگون ۵/۵ محاسبه شد که این توصیف‌کننده‌ها از نظر بیان چگونگی خصوصیات مولکول به ۲۲ دسته‌ی متفاوت تقسیم می‌شوند، که در جدول (۱-۲) نام این توصیف‌کننده‌ها آورده شده است.

## ۲-۳-۴- حذف توصیف‌کننده‌های نامناسب

در این مرحله باید توصیف‌کننده‌هایی که اطلاعات کمتری در مورد تغییرات خواص شیمیایی و یا بیولوژیکی مواد مورد نظر دارند، حذف شوند. برای این منظور لازم است که از فرآیند کاهش متغیر استفاده شود. زیاد بودن تعداد توصیف‌کننده‌ها و همچنین همبستگی بالای برخی از آنها به یکدیگر، باعث طولانی و دشوار شدن فرایند مدل‌سازی می‌شود. 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 غیر مفید، ماتریس همبستگی از طریق نرم‌افزار Matlab بین توصیف‌کننده‌ها تشکیل می‌شود. همبستگی بیش از ۹۰٪ توصیف‌کننده‌ها



بیانگر رابطه‌ی خطی بین آنهاست. در این صورت توصیف‌کننده‌ی ای که ارتباطش با متغیر وابسته مشکل‌تر بوده و ضریب همبستگی‌اش کمتر باشد، حذف می‌شود.

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

توصیف‌کننده‌های توپولوژی <sup>۲</sup>	۲	توصیف‌کننده‌های زیرساختاری <sup>۱</sup>	۱
شاخص‌های اتصال <sup>۴</sup>	۴	شمارنده‌های مولکولی <sup>۳</sup>	۳
خود ارتباطی‌های دو بعدی <sup>۶</sup>	۶	شاخص‌های اطلاعاتی <sup>۵</sup>	۵
خصوصیات بارگذاری <sup>۸</sup>	۸	شاخص‌های مجاورت لبه <sup>۷</sup>	۷
شاخص‌های مبتنی بر مقادیر ویژه <sup>۱۰</sup>	۱۰	شاخص بار توپولوژیکی <sup>۹</sup>	۹
توصیف‌کننده‌های هندسی <sup>۱۲</sup>	۱۲	پروفایل‌های مولکولی راندیک <sup>۱۱</sup>	۱۱
توصیف‌کننده‌های سه بعدی موریس <sup>۱۴</sup>	۱۴	توصیف‌کننده‌های RDF <sup>۱۳</sup>	۱۳
توصیف‌کننده‌های GETAWAY <sup>۱۶</sup>	۱۶	توصیف‌کننده‌های WHIM <sup>۱۵</sup>	۱۵
اجزای میان اتمی <sup>۱۸</sup>	۱۸	گروه‌های عاملی <sup>۱۷</sup>	۱۷
خصوصیات مولکولی <sup>۲۰</sup>	۲۰	توصیف‌کننده‌های باردار <sup>۱۹</sup>	۱۹
اثر فرکانس دو بعدی <sup>۲۲</sup>	۲۲	اثر دو بعدی باینری <sup>۲۱</sup>	۲۱

- 
- 1- Constitutional Descriptors
  - 2-Topological Descriptors
  - 3- Walk and path Counts
  - 4- Connectivity Indices
  - 5- Information Indices
  - 6- 2D Autocorrelations
  - 7- Edge adjacency Indices
  - 8- Burden eigenvalues
  - 9- Topological charge indices
  - 10- Eigenvalue-based indices
  - 11- Randic molecular profiles
  - 12-Geometrical descriptors
  - 13- Radial Distribution Function descriptors
  - 14- 3D-MORSE descriptors
  - 15- Weighted Holistic Invariant Molecular descriptors
  - 16- Geometry, Topology and Atom Weighted Assembly descriptors
  - 17- Functional group conts
  - 18- Atom-centred fragments
  - 19- Charge descriptors
  - 20- Molecular properties
  - 21- 2D binary fingerprints
  - 22- 2D frequency fingerprints

## ۲-۳-۵- تقسیم‌بندی داده‌ها

قبل از ساختن مدل، سری داده‌ها به سه زیر مجموعه تقسیم می‌شوند. اولین زیر مجموعه، سری آموزش<sup>۱</sup> می‌باشد که اکثریت داده‌ها را در بر گرفته و از آن برای ساخت مدل‌های خطی و غیرخطی استفاده می‌شود. دومین زیر مجموعه، سری ارزیابی<sup>۲</sup> است که مدل‌های به دست آمده از سری آموزش توسط آن ارزیابی شده و به این ترتیب بهترین مدل انتخاب می‌شود. زیر مجموعه‌ی سوم سری تست<sup>۳</sup> یا آزمایش نام دارد که در طول فرآیند مدل‌سازی دخالتی نداشته و از آن برای مقایسه‌ی مدل‌های متفاوت استفاده می‌شود. تقسیم بندی این سه زیر مجموعه معمولاً به صورت تصادفی و یا روش تحلیل مؤلفه‌های اساسی<sup>۴</sup> (PCA) استفاده می‌شود.

## ۲-۳-۶- انتخاب بهترین توصیف‌کننده‌ها

اگر چه 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. هر چه در مدل نهایی تعداد توصیف‌کننده‌ها کمتر باشد، مدل ساده‌تر خواهد شد. بنابراین باید به ترتیبی، بهترین توصیف‌کننده‌ها را از میان توصیف‌کننده‌های باقی مانده، انتخاب نمود. دو روش مرسوم برای انتخاب توصیف‌کننده‌ها، روش خطی رگرسیون مرحله‌ای (MLR) و روش غیرخطی الگوریتم ژنتیک<sup>۵</sup> 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.

---

1-Train  
2-Validation  
3-Test  
4- Principal Components Analysis  
5-Genetic algorithm

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

### ۲-۳-۶-۱-الف- مقدمه

امروزه با پیشرفت سریع دانش، بهینه‌سازی از اهمیت بالایی در علوم مختلف مهندسی برخوردار شده است. بهینه‌سازی یکی از مفاهیمی است که به خوبی از پل ارتباطی بین تئوری و عمل عبور کرده و دارای کاربرد گسترده‌ای می‌باشد. الگوریتم ژنتیک یک روش بهینه‌سازی الهام گرفته از طبیعت جاندار است و می‌توان در طبقه‌بندی‌ها، آن را به عنوان یک روش عددی، جستجوی مستقیم و تصادفی معرفی کرد. این الگوریتم در مسائل متنوعی نظیر بهینه‌سازی، شناسایی و کنترل سیستم، پردازش تصویر و مسائل ترکیبی، تعیین توپولوژی و آموزش شبکه‌های عصبی مصنوعی و سیستم‌های مبتنی بر تصمیم و قاعده به کار می‌رود.

### ۲-۳-۶-۱-ب- الگوریتم ژنتیک

الگوریتم ژنتیک (GA) یک تکنیک جستجو در علم رایانه برای یافتن راه‌حل بهینه و مسائل جستجو است. الگوریتم‌های ژنتیک یکی از انواع الگوریتم‌های تکاملی هستند که از مبانی علم زیست‌شناسی مانند وراثت، جهش، انتخاب ناگهانی، انتخاب طبیعی و ترکیب الهام گرفته‌اند. الگوریتم ژنتیک که به عنوان یکی از روش‌های بهینه‌یابی شناخته شده، توسط جان هالند در سال ۱۹۶۷ ابداع شد که او این روش را در سال ۱۹۷۵ در کتابی که درباره الگوریتم ژنتیک منتشر کرد آورده است [۲۷]. ساز و کار الگوریتم ژنتیک بر مبنای قوانین موجود در علم ژنتیک است. در این بخش با برخی مفاهیم و تعاریف 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<sup>۱</sup> مناسب انتخاب گردد. به این ساختار داده‌ای اصطلاحاً ژن یا کروموزوم<sup>۲</sup> گفته می‌شود. هر ژن مبین یک متغیر یا یک پارامتر است که تمامی اطلاعات لازم در مورد یک جواب را در خود دارد [۲۸].

**افراد<sup>۳</sup> یا کروموزوم<sup>۴</sup>:** در الگوریتم ژنتیکی، هر فرد تنها یک کروموزوم دارد و هر کروموزوم به معنای یک راه حل مسئله به صورت رمزی شده می‌باشد. کروموزوم‌ها از ترکیب مقادیر مختلف متغیرها یا ژن‌ها (هر ژن مبین یک متغیر است) به وجود می‌آیند [۲۹].

**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 به دست می‌آید که به آن نسل بعدی می‌گویند [۲۹].

**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.

**تابع شایستگی<sup>۵</sup>:** 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 محدوده مجاز صورت می‌گیرد.

---

1-Data Structure  
2-Chromosome  
3-Individuals  
4-Chromosome  
5- Fitness function

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

- انتخاب چرخ رولت<sup>۱</sup>؛ در این روش هر عنصری که عدد برازش (تناسب) بیشتری داشته باشد، شانس انتخاب بیشتری دارد.
- انتخاب نخبه‌گرا یا ممتازگرا<sup>۲</sup>؛ در این روش مناسب‌ترین عضو هر اجتماع انتخاب می‌شود.
- انتخاب مقیاسی<sup>۳</sup>؛ 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<sup>۴</sup>؛ 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 [۳۰].

## 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 مورد نظر را در یک جهت در یک لحظه جستجو کنند و اگر راه‌حل 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 نقطه 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 در اختیار قرار می‌دهند. الگوریتم ژنتیک تغییرات تصادفی در راه‌حل‌های

---

1- Roulette wheel  
2- Elitist  
3- Scalin  
4- Tournament

کандید شده را نشان می‌دهد سپس از تابع برازش برای سنجش این که آیا این 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:

۱- تعداد نسل‌ها: الگوریتم زمانی متوقف می‌شود که تعداد نسل‌ها به مقدار معینی برسد.

۲- محدودیت زمانی: در جعبه ابزار GA می‌توان این محدودیت را برحسب واحد ثانیه تعیین نمود.

۳- محدودیت شایستگی: الگوریتم زمانی متوقف می‌شود که بهترین مقدار شایستگی در جمعیت حاضر کوچکتر یا مساوی یک مقدار معین گردد.

۴- رکود نسلی: اگر در نسل‌ها به تعداد معینی هیچ گونه پیشرفتی حاصل نشود، الگوریتم متوقف می‌گردد.

۵- رکود زمانی: در صورتی که در تابع هدف برای زمان معینی هیچ گونه پیشرفتی حاصل نشود، الگوریتم متوقف می‌گردد.

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

## ۲-۳-۷- انتخاب توصیف‌کننده‌ها بر مبنای روش سهم‌گروه (GCM)

روش سهم‌گروه روشی ساده، کارآمد و در عین حال معتبر است که به طور گسترده‌ای در پیش‌بینی خواص فیزیکی و ترمودینامیکی مواد به کار 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 اگر چه هزاران ترکیب شیمیایی در دنیای علم و تکنولوژی وجود دارد اما تعداد گروه‌های ساختاری و عاملی سازنده‌ی همه‌ی این ترکیبات، بسیار کمتر از تعداد این ترکیبات می‌باشد. با این ایده، روش سهم‌گروه‌ها فرض می‌کند که خاصیت فیزیکی یا ترمودینامیکی هر ماده را می‌توان از سهم‌گروه‌های

---

۱- Groupe Contribution Method

موجود در مولکول آن ماده در خاصیت موردنظر، به دست آورد. به این ترتیب یک روش عملی برای پیش‌بینی خواص تعداد زیادی ماده برحسب تعداد کمتری پارامتر به دست می‌آید که این پارامترها، سهم گروه‌های ساختاری را در خاصیت موردنظر مشخص می‌کنند.

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

در مطالعات QSPR که به طور گسترده‌ای برای مدل‌سازی خواص فیزیکی و شیمیایی مواد به کار می‌رود، خاصیت مورد نظر به توصیف‌کننده‌های مولکولی مرتبط می‌شود. توصیف‌کننده‌های مولکولی مشخصه‌های عددی ماده‌ی خالص هستند که مستقیماً از روی ساختار بهینه شده‌ی مولکول و با استفاده از الگوریتم‌ها و نرم‌افزارهای خاص محاسبه می‌شوند. اما مهم‌ترین عیب این روش، مسیر طولانی و پیچیده‌ی محاسبه‌ی توصیف‌کننده از روی ساختار مولکولی و سپس انتخاب بهترین آن‌هاست. در نتیجه مدل‌های حاصل از این مطالعات برای استفاده، چندان ساده و راحت 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 سهم گروه‌ها در پیش‌بینی خواص مختلفی از جمله برای پیش‌بینی نقطه‌ی اشتعال آلکان‌ها [۳۲]، دانسیته‌ی مایعات یونی [۳۳]، خصوصیات نقطه‌ی بحرانی تعدادی از هیدروکربن‌ها [۳۴]، هدایت یونی و ویسکوزیته‌ی مایعات یونی [۳۵] استفاده شده است.

## ۲-۳-۸- ساختن مدل

پس از انتخاب بهترین توصیف‌کننده‌ها، نوبت به ساخت مدل می‌رسد. مدل، یک رابطه‌ی ریاضی است که بیان‌کننده‌ی ارتباط بین خاصیت مورد نظر 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 ترکیبات آلی استفاده شده است. در ادامه این روش به اختصار شرح داده خواهد شد.

## ۲-۳-۸-۱- شبکه عصبی مصنوعی

تحقیقات در مورد شبکه‌های عصبی مصنوعی از زمانی شروع شد که نگرش نوینی در مورد 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<sup>۱</sup> تشکیل شده که هر کدام به هزاران نرون دیگر متصل هستند. اگرچه همه 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- بدنه ی سلول یا جسم سلولی

۲- دندریت<sup>۲</sup>

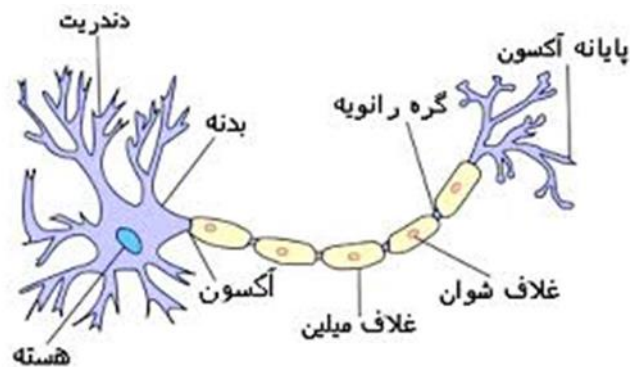
۳- آکسون<sup>۳</sup>

دندریت و آکسون، عناصر ارتباطی نرون را تشکیل می‌دهند. دندریت‌ها 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. دندریت‌ها گیرنده‌های نرون بوده و وظیفه‌ی آن‌ها انتقال سیگنال‌های الکتریکی به هسته 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 بر خلاف دندریت از سطح صاف‌تری برخوردار بوده و شاخه‌های کمتری دارد. آکسون‌ها، بر عکس دندریت‌ها از سطحی هموارتر و تعداد شاخه‌های کمتری برخوردار هستند. آنها طول بیشتری دارند و سیگنال الکتروشیمیایی دریافتی از هسته سلول را به نرون‌های دیگر منتقل می‌کنند. محل اتصال آکسون یک سلول به دندریت‌های سلول‌های دیگر را سیناپس<sup>۴</sup> می‌گویند. پیام‌های عصبی تنها به صورت یک طرفه از دندریت‌ها به بدنه سلول و از آنجا به آکسون حرکت می‌کنند. وقتی پیام به انتهای آکسون می‌رسد، موجب آزاد شدن یک ماده‌ی شیمیایی به نام انتقال دهنده نرونی از انتهای آکسون می‌شود و پس از

---

1- Neuron  
2- Dendrit  
3- Axon  
4- Synapse

نفوذ در سیناپس‌ها، گیرنده‌های سلول‌های مجاور را فعال می‌کند [۳۶]. شکل (۲-۱) ساختار کلی یک نرون زیستی را نشان می‌دهد.



شکل (۲-۱) - ساختار یک نرون زیستی

شبکه عصبی مصنوعی یک برنامه نرم‌افزاری است که با روشی مشابه با مغز، اطلاعات را پردازش می‌کند. شبکه‌های عصبی 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 گفته می‌شود زیرا براساس محاسبات روی داده‌های عددی یا مثال‌های قوانین کلی را فرا می‌گیرند [۳۷].

## ۲-۳-۸-۲- تاریخچه‌ی شبکه عصبی مصنوعی

از قرن نوزدهم به طور همزمان اما جداگانه از سوی نروفیزیولوژیست‌ها سعی کردند سیستم یادگیری و تجزیه و تحلیل مغز را کشف کنند و از سوی دیگر ریاضیدانان تلاش کردند تا مدل ریاضی بسازند که قابلیت فراگیری و تجزیه و تحلیل عمومی مسائل را دارا باشد. اولین کوشش‌ها در شبیه‌سازی با استفاده از یک مدل منطقی توسط مک کلوک<sup>۱</sup> و والتر پیتز<sup>۲</sup> انجام شد که امروزه بلوک

1- McCulloch  
2- Walter Pits

اصلی سازنده‌ی اکثر شبکه‌های مصنوعی است. اگر چه این مدل فرضیه‌هایی در مورد عملکرد نرون‌ها ارائه می‌کند اما محدودیت‌های تکنولوژی در آن زمان اجازه کار بیشتر را به آن‌ها نداد [۳۸]. نه تنها نروفیزیولوژیست‌ها بلکه روانشناسان و مهندسان نیز در پیشرفت شبیه‌سازی شبکه‌های عصبی تاثیر داشتند. در سال ۱۹۵۸ شبکه‌ی پرسپترون<sup>۱</sup> توسط فرانک روزنبلات<sup>۲</sup> معرفی گردید. پرسپترون دارای ۳ لایه به همراه یک لایه‌ی وسط که به عنوان لایه‌ی پیوند شناخته می‌شده است. این سیستم می‌تواند یاد بگیرد که به ورودی داده شده خروجی تصادفی متناظر را اعمال کند. سیستم دیگر مدل خطی تطبیقی نرون (آدالاین) می‌باشد که در سال ۱۹۶۰ توسط برنارد ویدرو<sup>۳</sup> و تد هوف<sup>۴</sup> به وجود آمد که اولین شبکه‌های عصبی به کار گرفته شده در مسائل واقعی بودند [۳۹]. هر دوی این شبکه‌ها، پرسپترون و آدالاین، محدودیت داشتند و فقط قادر به طبقه‌بندی الگوهایی بودند که به طور خطی متمایز می‌شدند. این محدودیت‌ها در سیستم‌های تک لایه و چند لایه در سال ۱۹۶۹ توسط مینسکی<sup>۵</sup> و پاپرت<sup>۶</sup> در کتابی مورد بررسی قرار گرفت. پیشرفت‌هایی که در سال‌های ۱۹۷۰ تا ۱۹۸۰ بدست آمد برای جلب توجه به شبکه‌های عصبی بسیار مهم بود. در دهه‌ی هشتاد ایده به کارگیری شبکه‌ها در زمینه‌های جدید مطرح شد. دهه‌ی هشتاد قرن بیستم را می‌توان زمان تولد دوباره شبکه عصبی دانست چرا که در این دوران دو کار بسیار بزرگ صورت گرفت.

۷ در سال ۱۹۸۲ فیزیکدان آمریکایی جان هاپفیلد<sup>۷</sup> شبکه‌های چند لایه و الگوریتم

یادگیری دارای بازخورد را معرفی کرد که از این شبکه‌ها برای ذخیره سازی اطلاعات

استفاده می‌شود [۴۰].

---

1- Perceptron  
2- Frank Rosenblat  
3- Bernard Widrow  
4- Ted Hoff  
5- Minsky  
6- Papert  
7- John Hopfield

۷ در سال ۱۹۸۶ دیوید راملهارت<sup>۱</sup> و جیمز مککلند<sup>۲</sup> الگوریتم پس انتشار را مطرح کردند و از

آن زمان تاکنون این زمینه ی علمی روند رشد و پیشرفت را ادامه داده است.

پس از آن زمان کاربردهای شبکه‌ی عصبی مصنوعی در شیمی و علوم وابسته به آن مانند بیوشیمی، مهندسی شیمی و داروسازی رواج پیدا کرده است [۴۱] و مسلماً در آینده نیز این شبکه ها خواهند توانست به عنوان یک ابزار علمی برای حل مسائل جدید آن زمان مطرح شوند.

## ۲-۳-۸-۲-الف-ساختار شبکه عصبی مصنوعی

مهم‌ترین جزء تشکیل‌دهنده‌ی یک شبکه‌ی عصبی مصنوعی نرون می‌باشد. نرون یک واحد محاسباتی-پردازشی است که در حقیقت مدل ساده شده‌ای از نرون طبیعی است. هر نرون از تعدادی ورودی، خروجی، وزن‌های مربوط به ایجاد سیگنال تحریک و تابع تحریک تشکیل شده است. 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 ارتباط دارد بفرستد. توابع تحریک به کار رفته در شبکه‌ی مصنوعی می‌توانند خطی و یا غیرخطی باشند، البته برای مقاصد مختلف، توابع محرک متفاوتی به کار می‌روند. تعداد توابع محرک زیادی در ساختار شبکه‌ی عصبی می‌توانند مورد استفاده قرار گیرند که سه مورد از معروف‌ترین توابع محرک عبارتند از: تابع محرک خطی،<sup>۳</sup> تابع محرک آستانه‌ی حدی<sup>۴</sup> و تابع محرک سیگموئید<sup>۵</sup>. در سیستم‌های پردازش موازی سه نوع واحد پردازشگر قابل تفکیک وجود دارد که عبارتند از: واحدهای ورودی، واحدهای خروجی و واحدهای میانی یا پنهان.

---

1- David Rummelhart

2- James Mcland

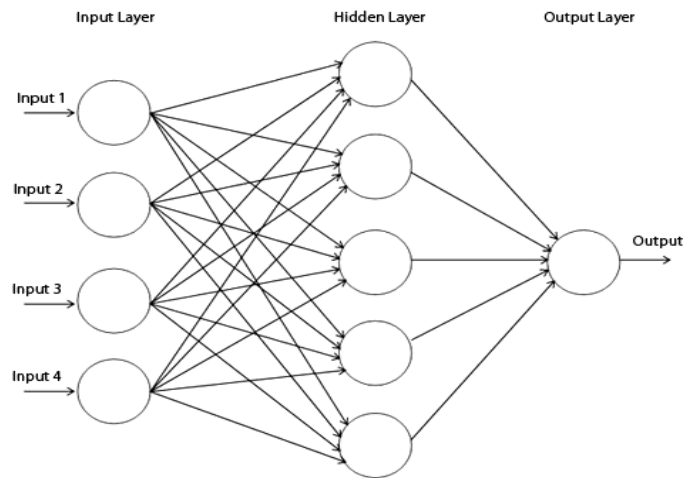
3- Linear Transfer Function

4- Hard Limit Transfer Function

5- Log-sigmoid Transfer Function

- 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 [۳۶].

واحدهای یک شبکه‌ی عصبی مصنوعی در شکل (۲-۲) نشان داده شده است.



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

## ۲-۳-۸-۲-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<sup>۱</sup>

۲- شبکه‌های پس‌خور<sup>۲</sup>

شبکه‌های پیش‌خور، شبکه‌هایی هستند که مسیر پاسخ در آن‌ها، همواره رو به جلو پردازش می‌شود و به نرون‌های لایه قبل باز نمی‌گردد. در این نوع شبکه‌ها به سیگنال‌ها اجازه می‌دهند تنها از یک مسیر یک طرفه عبور کنند، یعنی از ورودی تا خروجی. بنابراین بازخوردی وجود ندارد به این معنی که خروجی هر لایه تاثیری بر همان لایه ندارد. در بدن انسان نیز، پیام‌هایی عصبی به صورت یک طرفه حرکت می‌کنند: از دندریت به بدنه سلول و سپس به آکسون. 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.



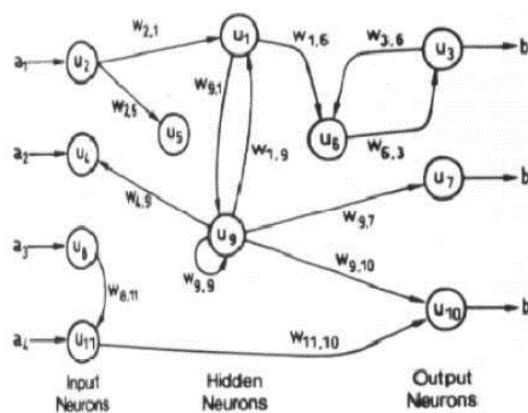
شکل (۲-۳)- نمونه‌ای از یک شبکه پیش‌خور

تفاوت شبکه‌های پس‌خور با پیش‌خور در آن است که در شبکه‌های پس‌خور حداقل یک سیگنال برگشتی از یک نرون به همان نرون یا نرون‌های همان لایه یا نرون‌های لایه‌های قبل وجود دارد و اگر نرونی دارای فیدبک باشد بدین مفهوم است که خروجی نرون در لحظه حال نه تنها به ورودی در آن لحظه بلکه به مقدار خروجی خود نرون، در لحظه‌ی گذشته نیز بستگی دارد. شبکه‌های برگشتی بهتر

---

1- Feed Forward  
2- Feed Back

می‌توانند رفتار مربوط به ویژگی‌های زمانی و پویایی سیستم‌ها را نشان دهند. این شبکه‌ها پویا هستند، وضعیت آن‌ها پیوسته در حال تغییر است تا این که به یک نقطه‌ی تعادل برسند. آن‌ها در این وضعیت تعادل باقی می‌مانند تا زمانی که ورودی تغییر کند و نیاز باشد تا تعادل تازه‌ای پیدا شود. 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 [۳۱].



شکل (۲-۴) - نمونه‌ای از شبکه پس‌خور برگشتی

## ۲-۳-۸-۲-ج-روش‌های آموزش

برای آموزش شبکه‌های عصبی دو روش آموزش با ناظر<sup>۱</sup> و آموزش بدون ناظر<sup>۲</sup> مورد استفاده قرار می‌گیرند. در یادگیری با ناظر، به ازای هر دسته از الگوهای ورودی خروجی‌های متناظر نیز به شبکه نشان داده می‌شود و تغییر وزن‌ها تا موقعی صورت می‌گیرد که اختلاف خروجی شبکه به ازای الگوهای آموزشی از خروجی واقعی در حد خطای قابل قبولی باشد. در اکثر شبکه‌های عصبی از این نوع یادگیری استفاده می‌شود. چنین شبکه‌ای به طور گسترده برای کارهای تشخیص الگو به کار گرفته می‌شود.

1- Supervised Learning  
2- Unsupervised Learning

نوع دیگری از یادگیری به نام یادگیری بدون ناظر یا خودسامان‌ده می‌باشد که وزن‌ها فقط بر اساس ورودی‌ها اصلاح می‌شوند و خروجی مطلوب وجود ندارد تا با مقایسه خروجی شبکه با آن و تعیین مقدار خطا وزن‌ها اصلاح شود. وزن‌ها فقط بر اساس اطلاعات الگوهای ورودی به هنگام می‌شوند. هدف این روش، استخراج مشخصه‌های الگوهای ورودی بر اساس راهبرد خوشه‌یابی یا دسته‌بندی و تشخیص شباهت‌ها (تشکیل گروه‌هایی با الگوی مشابه) است، بدون این که خروجی یا کلاس‌های متناظر با الگوهای ورودی از قبل مشخص باشد. این یادگیری معمولاً بر پایه‌ی شیوه‌ی برترین هم‌خوانی انجام می‌گیرد. شبکه‌ی بدون ناظر وزن‌های خود را بر پایه‌ی خروجی حاصل شده از ورودی تغییر می‌دهد تا در برخورد بعدی پاسخ مناسبی را برای این ورودی داشته باشد. در نتیجه شبکه یاد می‌گیرد چگونه به ورودی پاسخ بدهد. به عبارت دیگر هدف روش این است که با تکنیک نرون غالب نرونی که بیشترین تحریک آغازین را دارد برگزیده شود. بنابراین در شبکه‌های بدون ناظر یافتن نرون غالب یکی از مهمترین کارها است.

برای بهینه‌سازی عملکرد شبکه، ابتدا باید یک شاخص عملکرد پیدا کرد. 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) است. شاخص عملکرد را اصطلاحاً تابع هدف یا تابع خطا نیز می‌گویند، یعنی تابعی که کمینه کردن آن مورد نظر است. پس از تعیین شاخص عملکرد، باید پارامترهای شبکه برای کاهش مقدار این شاخص عملکرد تنظیم گردد [۴۲].

از رایج‌ترین الگوریتم‌های آموزشی، الگوریتم‌های آموزشی لونبرگ-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 روش‌ها صد برابر بیشتر از شیوه‌های استاندارد مشابه است. در این تحقیق نیز از این دو الگوریتم برای بهینه‌سازی عملکرد شبکه استفاده شده است [۴۳].

## ۲-۳-۸-۲-د-مزایای شبکه عصبی مصنوعی

- **قابلیت یادگیری:** یعنی توانایی تنظیم پارامترهای شبکه، به نحوی که با تغییر ورودی‌ها، شبکه بتواند خود را با شرایط جدید وفق دهد. در نتیجه در هر زمینه‌ای که مجموعه داده‌های آموزش را دریافت کند (بدون توجه به نوع ورودی و خروجی) در آن زمینه کارآمد باشد.
- **مقاوم بودن:** چون هر سلول کار مخصوص به خود را انجام می‌دهد و رفتار کلی شبکه، برآیند رفتارهای محلی سلول‌های متعدد است، اگر چند سلول به طور محلی خطا کنند، خروجی نهایی به خاطر درست عمل کردن اکثر نرون‌ها از این خطاها مصون می‌ماند. بعبارت دیگر نرون‌ها خطای یکدیگر را تصحیح می‌کنند. در این صورت شبکه توانایی تحمل خطا و مقاومت در مقابل اشتباه را دارد.
- **پردازش اطلاعات به صورت موازی:** سلول‌های هر لایه به طور همزمان روی ورودی‌های آن لایه اثر می‌کنند و علی‌رغم مستقل بودن، جمعاً ورودی را پردازش کرده و خروجی را تحویل می‌دهند. در نتیجه، کار بین چند پردازشگر تقسیم شده و سرعت پردازش بالا می‌رود.
- **قابلیت تعمیم:** بعد از آموزش شبکه با چند الگو، شبکه قادر است در مقابل یک ورودی تجربه نشده قرار بگیرد و خروجی مناسب آن را ارائه نماید. بدین معنی که شبکه از آموخته‌های خود برای پاسخ دادن به مجهولات استفاده کند [۴۴].

## ۲-۳-۸-۲-۵- معایب شبکه عصبی مصنوعی

با وجود برتری هایی که شبکه عصبی مصنوعی نسبت به سیستم های مرسوم دارند، معایبی نیز دارند که پژوهشگران این رشته تلاش دارند که آن ها را به حداقل برسانند، از جمله:

- قواعد یا دستورات مشخصی برای طراحی شبکه جهت یک کاربرد اختیاری وجود ندارد.
- در مورد مسائل مدلسازی، صرفا نمیتوان با استفاده از شبکه عصبی به فیزیک مسئله پی برد، بعبارتی دیگر مرتبط ساختن پارامترها یا ساختار شبکه به پارامترهای فرآیند معمولا غیرممکن است.
- دقت نتایج بستگی زیادی به اندازه مجموعه آموزش دارد.
- آموزش شبکه ممکن است مشکل و یا حتی غیر ممکن باشد.
- پیش بینی عملکرد آینده شبکه به سادگی امکان پذیر نیست.

## ۲-۳-۹-۲- ارزیابی قدرت پیش بینی مدل

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

برای اطمینان از این که مدل به دست آمده توانایی پیش بینی نمونه های مختلفی از یک جمعیت را داراست، باید مدل را ارزیابی کرد. این ارزیابی با محاسبه ی پارامترهای آماری صورت می گیرد. روابط ریاضی پارامترهای آماری استفاده شده در این پایان نامه، در ادامه توضیح داده خواهد شد.

**ضریب تعیین ( $R^2$ )**؛ به عنوان یک شاخص برای بیان دقت خط رگرسیون برآورده شده، به کار می رود و نشان دهنده ی 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:

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

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

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

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

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (3-2)$$

که در این روابط  $\hat{y}_i$  مقدار پیش‌بینی شده‌ی متغیر وابسته،  $y_i$  مقدار واقعی متغیر وابسته و  $\hat{y}$  در دو رابطه، میانگین مقادیر متغیر وابسته است. همچنین SSE<sup>۳</sup> مین مجموع مربعات انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است و طبق رابطه‌ی (۴-۲) تعریف می‌شود:

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

بنابراین با توجه به رابطه فوق می‌توان نوشت:

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

---

1- Sum Square Regression  
2- Sum Square Total  
3- Sum Square Error

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 هر گونه 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):

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n} \quad (۶-۲)$$

### مجموع مربع باقیمانده‌ها (PRESS):

$$PRESS = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (۷-۲)$$

### میانگین خطای مطلق (MAE):

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}|}{n} \quad (۸-۲)$$

### متوسط مطلق درصد خطای نسبی (MAPE):

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right| \quad (۹-۲)$$

### متوسط درصد خطای نسبی (MPE):

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{y_i - \hat{y}}{y_i} \quad (۱۰-۲)$$

- 
- 1- Predictive Residual Sum of Square
  - 2- Mean Absolute Error
  - 3- Mean Absolute Percentage Error
  - 4- Mean Percentage Error

## ماکزیمم مطلق درصد خطای نسبی ( $D_{max}$ )

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

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

### ۲-۳-۹-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 تا ۱ است. هر چقدر ضریب تعیین به ۱ نزدیک‌تر باشد، نتایج پیش‌بینی شده به مقادیر واقعی نزدیک‌ترند. اما اگر این مقدار برابر صفر باشد، بین داده‌ها هیچ‌گونه همبستگی وجود ندارد.

### ۲-۳-۹-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 نشان‌دهنده‌ی تصادفی بودن خطاست. ولی اگر عمده‌ی نقاط در این نمودار، در یک طرف خط صفر باشد، این بدان معناست که خطای جهت‌داری رخ داده است.

## ۲-۳-۹-۵- استفاده از سری تست خارجی

اعتبار و اهمیت مدل‌های پیش‌بینی کننده وقتی مشخص می‌شود که بتواند خاصیت مورد نظر را برای مولکول‌های جدید 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 مورد نظر مولکول‌های موجود در سری تست خارجی که در مدل‌سازی استفاده نشده‌اند، استفاده می‌شود.

فصل سوم

# محاسبات

در دو فصل قبل مطالبی راجع به دانسیته‌ی سیالات خالص 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 طیف وسیعی از ترکیبات هیدروکربنی خالص 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.

### ۳-۱- انتخاب سری داده‌ها

به منظور پیش‌بینی دانسیته‌ی هیدروکربن‌ها، ۶۰۸۳ نقطه داده‌ی مربوط به ۷۹ ترکیب هیدروکربنی شامل آلکان‌های خطی و شاخه‌دار، سیکلو آلکان‌ها، آلکن‌های خطی و شاخه‌دار، سیکلو آلکن‌ها و ترکیبات آروماتیک در محدوده‌ی وسیعی از دما و فشار جمع‌آوری گردید. نام این ترکیبات به همراه محدوده‌ی دما، فشار، تعداد نقاط مورد استفاده برای هر ترکیب و مرجع مربوط به هر یک از آن‌ها در جدول (۳-۱) گزارش شده است.



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

شماره ترکیب	نام ترکیب	$\Delta T$ (K)	$\Delta p$ (MPa)	$\Delta\rho$ (kg.m <sup>-3</sup> )	تعداد نقاط	مرجع
۱	<i>n</i> -Pentane	۳۰۳/۱۵ - ۳۲۳/۱۵	۲۱/۵ - ۲۵۶/۶	۶۰۳/۲ - ۷۴۴/۲	۱۹	۴۵
۲	<i>n</i> -Hexane	۲۹۸/۱۵ - ۳۷۳/۱۵	۰/۱ - ۵۶۴	۵۸۰/۵ - ۷۸۵/۹	۳۱	۴۶
۳	<i>n</i> -Heptane	۲۹۸/۱۵ - ۳۱۰/۶۵	۰/۱ - ۲۷۱/۱	۶۶۸/۷ - ۷۹۷/۲	۱۹	۴۷
۴	<i>n</i> -Octane	۲۹۸/۰۳ - ۳۴۸/۱۴	۰/۱ - ۴۶۷	۶۵۷/۱ - ۸۳۸/۳	۲۸	۴۸
۵	<i>n</i> -Nonane	۳۰۳/۱۵ - ۴۲۳/۱۵	۰ - ۸۰۰	۵۱۲/۵ - ۸۸۷/۳	۵۱۵	۴۷
۶	<i>n</i> -Decane	۲۸۳/۱۵ - ۳۹۸/۱۵	۰/۱ - ۷۰	۶۴۷/۵ - ۷۱۳/۵	۴۹	۴۸
۷	<i>n</i> -Undecane	۳۰۳/۱۵ - ۵۷۳/۱۵	۰ - ۵۰۰	۵۰۴/۱ - ۸۲۳/۵	۷۶	۴۷
۸	<i>n</i> -Dodecane	۲۹۸/۱۵ - ۴۷۳/۱۵	۰/۱ - ۱۶۰/۹۴	۶۳۰/۳ - ۷۴۸/۶	۴۴	۴۹
۹	<i>n</i> -Tridecane	۳۰۳/۱۵ - ۵۷۳/۱۵	۰ - ۵۰۰	۵۰۲/۴ - ۸۳۵/۳	۷۷	۴۷
۱۰	<i>n</i> -Tetradecane	۲۹۳/۱۵ - ۴۳۳/۱	۰/۱ - ۱۴۰	۷۴۶/۳ - ۸۰۹/۹	۵۹	۵۰
۱۱	<i>n</i> -Pentadecane	۲۹۳/۱۵ - ۳۸۳/۱۵	۰/۱ - ۱۴۹/۵۵	۷۰۵ - ۷۹۴/۹	۱۴۲	۵۱
۱۲	<i>n</i> -Hexadecane	۲۹۸/۱۵ - ۳۷۳/۱۵	۰/۱ - ۴۵۰/۵	۷۱۸/۶ - ۸۲۲/۲	۲۷	۴۶
۱۳	<i>n</i> -Heptadecane	۳۰۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۱۴۹/۵۵	۷۱۶ - ۷۰۲/۹	۱۲۴	۵۱
۱۴	<i>n</i> -Octadecane	۳۰۳/۱۵ - ۳۸۳/۱۵	۰/۱ - ۱۴۹/۵۵	۷۲۰/۶ - ۸۰۶/۵	۱۱۱	۵۲
۱۵	<i>n</i> -Nonadecane	۳۱۳/۱۵ - ۳۸۳/۱۵	۰/۱ - ۱۴۹/۵۵	۷۲۴/۸ - ۸۰۹/۶	۱۰۷	۵۲
۱۶	<i>n</i> -Eicosane	۳۷۳/۱۵ - ۵۷۳/۱۵	۰ - ۵۰۰	۵۸۳/۶ - ۸۵۸/۶	۵۵	۴۷
۱۷	<i>n</i> -Octacosane	۳۵۳/۱۵ - ۴۰۳/۱۵	۰/۱ - ۱۴۹/۵۵	۷۳۵/۱ - ۸۲۳/۶	۸۷	۵۳
۱۸	<i>n</i> -Triacontane	۳۷۳/۱۵ - ۵۷۳/۱۵	۰ - ۵۰۰	۶۲۵ - ۸۷۳/۴	۵۵	۴۷

ادامه جدول-(۱-۳)

شماره ترکیب	نام ترکیب	$\Delta T$ (K)	$\Delta p$ (MPa)	$\Delta \rho$ (kg.m <sup>-3</sup> )	تعداد نقاط	مرجع
۱۹	<i>n</i> -Hexatriacontane	۳۶۳/۱۵ - ۴۰۳/۱۵	۰/۱ - ۱۵۱/۳	۷۴۶/۱ - ۸۳۳/۵	۶۴	۵۳
۲۰	<i>n</i> -Tetracontane	۴۲۳/۱۵ - ۵۷۳/۱۵	۰ - ۵۰۰	۶۴۷/۶ - ۸۹۳/۶	۴۴	۴۷
۲۱	Isobutane	۲۰۰ - ۴۰۰	۰/۳ - ۳۰	۴۸۵/۶ - ۶۷۲/۱	۳۶	۵۴
۲۲	Neopentane	۳۴۳/۱۵ - ۳۸۸/۱۵	۰/۵ - ۳۱/۱	۴۸۴/۶ - ۵۶۴/۶	۵۵	۵۵
۲۳	2-Methylbutane	۲۰۱/۴ - ۲۵۹/۹	۰/۱ - ۳۰۰	۷۰۷/۳ - ۸۰۹/۱	۱۸۶	۵۶
۲۴	2,3-Dimethylbutane	۳۷۳/۱۵ - ۵۴۸/۱۵	۰/۳ - ۳۱/۵	۴۷۹/۱ - ۶۱۵/۱	۴۸	۵۷
۲۵	2-Methylpentane	۲۹۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۱۵۰	۵۹۲/۸ - ۷۰۹/۶	۱۳۹	۵۸
۲۶	2,3-Dimethylpentane	۲۹۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۱۵۰	۶۲۰/۶ - ۷۳۸/۷	۱۴۴	۵۸
۲۷	2,2,4-Trimethylpentane	۲۹۳/۱۵ - ۳۷۳/۱۵	۰/۱ - ۱۵۰	۶۹۱/۹ - ۷۴۰/۳	۱۴۴	۵۸
۲۸	Cyclopentane	۲۸۳/۱۵ - ۳۳۳/۱۵	۰/۱ - ۱۹/۵	۷۰۳/۹ - ۷۴۵/۵	۱۰۹	۵۹
۲۹	Methylecyclopentane	۲۸۳/۱۵ - ۳۳۳/۱۵	۰/۱ - ۱۹/۵	۷۲۰/۱ - ۷۴۹/۲	۹۵	۵۹
۳۰	Cyclohexane	۲۸۳/۱۵ - ۳۳۳/۱۵	۰/۱ - ۱۹/۵	۷۹۵/۳ - ۷۷۷/۲	۴۰	۵۹
۳۱	Methylcyclohexane	۲۸۳/۱۵ - ۳۳۳/۱۵	۰/۱ - ۱۹/۵	۷۳۴ - ۷۸۶/۱	۹۱	۵۹
۳۲	Ethylcyclohexane	۳۲۷/۱ - ۴۴۰/۹	۰/۴ - ۹/۷	۶۹۲ - ۷۷۸	۲۴	۶۰
۳۳	Cis-1,2-Dimethylcyclohexane	۳۲۵/۲ - ۴۶۱/۲	۰/۴ - ۹/۷	۷۱۵ - ۷۸۴	۲۱	۶۰
۳۴	Cyclooctane	۳۲۴/۴۵ - ۵۲۳/۸	۱۴ - ۲۷۵/۷	۶۹۹ - ۸۷۲	۲۹	۶۱
۳۵	Heptylcyclohexane	۲۹۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۱۰۰	۷۹۷ - ۸۴۴/۶	۴۲	۶۲
۳۶	Pentadecylcyclohexane	۳۰۳/۱۵ - ۳۸۳/۱۵	۰/۱ - ۱۵۰	۸۱۳/۹ - ۸۷۳/۲	۱۲۸	۶۳

ادامه جدول-(۱-۳)

شماره ترکیب ب	نام ترکیب	$\Delta T$ (K)	$\Delta p$ (MPa)	$\Delta \rho$ (kg.m <sup>-3</sup> )	تعداد نقاط	مرجع
۳۷	Nonadecylcyclohexane	۳۲۳/۱۵ - ۳۸۳/۱۵	۰/۱ - ۱۵۰	۷۸۵/۹ - ۸۵۰/۳۶	۹۵	۶۳
۳۸	Ethylene	۳۰۰ - ۳۵۰	۶۰ - ۲۶۰	۴۶۶/۱ - ۶۱۱/۲	۲۶	۶۴
۳۹	1-Butene	۲۷۳/۲۴ - ۵۰۸/۲۴	۰/۶ - ۳۹/۹	۴۶۱/۵۱ - ۵۳۲/۳	۳۱۰	۶۵
۴۰	Isobutene	۲۷۳/۲۴ - ۵۰۸/۲۴	۰/۶ - ۳۹/۹	۴۶۲ - ۶۲۱/۵	۲۷۹	۶۵
۴۱	Cis-2-butene	۲۷۳/۲۴ - ۵۲۳/۲۴	۰/۶ - ۳۹/۹	۵۰۴/۱۱ - ۶۵۹/۴	۳۴۶	۶۵
۴۲	Trans-2-butene	۲۷۳/۲۴ - ۵۱۸/۲۵	۰/۶ - ۳۹/۹	۴۶۴/۳ - ۶۱۵/۹	۳۲۸	۶۵
۴۳	1-Pentene	۳۵۳/۱۵ - ۴۹۸/۱۵	۰/۵ - ۳۱/۵	۴۶۳/۹ - ۶۲۳/۲	۵۱	۶۶
۴۴	1-Hexene	۲۹۸/۱۵ - ۴۷۱/۱۵	۰ - ۲۴۵/۱۵	۵۷۵/۴ - ۷۹۳/۴	۵۲	۶۷
۴۵	1-Heptene	۲۹۸/۱۵ - ۳۷۳/۱۵	۰ - ۲۴۵/۱۵	۵۷۸/۵ - ۸۰۰/۸	۵۶	۶۷
۴۶	1-Octene	۳۰۳/۱۵ - ۴۳۳/۱۵	۰/۱ - ۲۲۹/۲۱	۶۴۶/۶ - ۸۱۵/۸	۳۷	۶۸
۴۷	1-Tetradecene	۳۰۳/۱۵ - ۴۳۳/۱۵	۰/۱ - ۱۰۰	۶۷۷/۵ - ۷۶۳	۲۱۰	۶۹
۴۸	1-Hexadecene	۲۹۸/۱۵ - ۳۲۳/۲	۰/۱ - ۱۰۰	۷۷۰/۹ - ۸۰۹	۱۵۴	۷۰
۴۹	Cis-hexa-1,3,5-triene	۲۹۸/۱۵ - ۳۲۳/۲	۰/۱	۷۴۹/۳ - ۷۷۲/۲	۴	۷۱
۵۰	Trans-hexa-1,3-diene	۲۹۸/۱۵ - ۳۲۳/۲	۰/۱	۶۷۵/۹ - ۷۰۱/۱	۴	۷۱
۵۱	Trans-hexa-1,4-diene	۲۹۸/۱۵ - ۳۲۳/۲	۰/۱	۶۶۷/۸ - ۶۹۲	۴	۷۱
۵۲	Trans-trans-hexa-2,4-diene	۲۹۸/۱۵ - ۳۲۳/۲	۰/۱	۶۶۳/۶ - ۶۸۹/۴	۴	۷۱
۵۳	Trans-hexa-1,5-diene	۲۹۸/۱۵ - ۳۲۳/۲	۰/۱	۶۸۷/۴ - ۷۰۰/۹	۴	۷۱
۵۴	Cyclohexa-1-ene	۲۹۸/۱۵ - ۳۲۳/۲	۰/۱	۷۹۰/۷ - ۸۱۴/۵	۴	۷۱

ادامه جدول (۱-۳)

شماره ترکیب	نام ترکیب	$\Delta T$ (K)	$\Delta p$ (MPa)	$\Delta \rho$ (kg.m <sup>-3</sup> )	تعداد نقاط	مرجع
۵۵	Cyclohexa-1,3-diene	۲۹۸/۱۵ - ۲۲۳/۲	۰/۱	۸۱۰/۹ - ۸۳۷/۷	۴	۷۱
۵۶	Cyclohexa-1,4-diene	۲۹۸/۱۵ - ۲۲۳/۲	۰/۱	۸۲۴/۸ - ۸۴۹/۷	۴	۷۱
۵۷	Benzene	۲۹۸/۸۵ - ۳۷۳/۱۵	۰/۱ - ۲۶۰	۸۴۷ - ۹۳۶	۳۴	۷۲
۵۸	Toluene	۲۸۳/۱۵ - ۳۹۸/۱۵	۰/۱ - ۷۰	۷۶۵/۱ - ۹۱۶/۴	۴۸	۷۳
۵۹	<i>o</i> -xylene	۳۱۸/۱۵ - ۳۷۳/۱۵	۰/۱ - ۱۰۰	۸۳۷/۱ - ۸۶۵/۸	۳۶	۷۴
۶۰	<i>m</i> -xylene	۲۹۸/۱۵ - ۴۷۳/۱۵	۰/۱ - ۱۹۶/۴	۸۳۸/۹ - ۹۳۸/۳	۴۹	۴۸
۶۱	<i>p</i> -xylene	۳۲۳/۱۵ - ۴۴۸/۱۵	۰/۴ - ۴۹/۶	۷۹۰/۱ - ۸۲۰/۴	۸۲	۷۵
۶۲	Ethylbenzene	۲۹۳/۱۵ - ۳۵۳/۱۵	۱ - ۴۰	۸۳۵/۴ - ۸۶۱/۷	۳۶	۷۶
۶۳	Propylbenzene	۳۱۸/۱۵ - ۳۳۳/۱۵	۰/۱ - ۴/۹	۸۲۶/۷ - ۸۴۳/۴	۱۹	۷۷
۶۴	Isopropylbenzene	۳۱۸/۱۵ - ۳۳۳/۱۵	۰/۱ - ۴/۹	۸۲۶/۱ - ۸۳۴/۱	۲۱	۷۷
۶۵	Butylbenzene	۲۹۳/۱۵ - ۳۵۳/۱۵	۱ - ۴۰	۸۲۸/۷ - ۸۵۸	۳۶	۷۶
۶۶	Hexylbenzene	۲۹۳/۱۵ - ۳۵۳/۱۵	۱ - ۴۰	۸۱۶/۳ - ۸۷۱/۸	۳۶	۷۶
۶۷	Diphenylmethane	۳۲۳/۱۵ - ۴۱۳/۱۵	۰/۱ - ۳۰	۹۱۰/۱ - ۹۷۵/۴	۲۱	۷۸
۶۸	Heptylbenzene	۲۹۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۱۰۰	۸۱۲/۷ - ۸۸۴/۹	۴۲	۶۲
۶۹	Octylbenzene	۲۹۳/۱۵ - ۳۵۳/۱۵	۱ - ۴۰	۸۱۳/۱ - ۸۶۵/۶	۳۶	۷۶
۷۰	Nonylbenzene	۲۹۳/۱۵ - ۳۵۳/۱۵	۱ - ۴۰	۸۱۴/۴ - ۸۶۳/۲	۱۲	۷۶
۷۱	1-Phenyldecane	۲۹۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۶۵	۸۱۴ - ۸۶۵/۷	۵۶	۷۹
۷۲	1-Phenylundecane	۲۹۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۶۵	۸۴۲/۵ - ۸۵۴/۶	۵۶	۷۹

ادامه جدول (۱-۳)

شماره ترکیب	نام ترکیب	$\Delta T$ (K)	$\Delta P$ (MPa)	$\Delta \rho$ (kg.m <sup>-3</sup> )	تعداد نقاط	مرجع
۷۳	1-Phenyltridecane	۲۹۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۱۵۰	۸۴۰/۹ - ۹۰۶/۲	۸۵	۸۰
۷۴	1-Phenylpentadecane	۳۰۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۱۵۰	۸۴۲/۱ - ۹۱۰/۸	۸۸	۸۰
۷۵	1-Phenylheptadecane	۳۰۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۱۵۰	۸۴۰/۹ - ۸۹۷/۵	۷۸	۸۱
۷۶	1-Phenyloctadecane	۳۰۳/۱۵ - ۳۵۳/۱۵	۰/۱ - ۱۵۰	۸۲۷/۳ - ۸۷۸/۶	۵۶	۸۱
۷۷	Naphthalene	۳۷۸/۷۱ - ۷۲۸/۷	۲۰/۹ - ۳۹/۹	۷۶۰ - ۹۸۵/۵	۱۰	۸۲
۷۸	Tetraline	۲۹۸/۱۵ - ۴۴۸/۱۵	۰/۱ - ۲۰۰/۱	۹۲۵/۱ - ۱۰۲۳	۴۰	۴۸
۷۹	1-Methylnaphthalene	۲۹۸/۱۵ - ۴۷۳/۱۵	۰/۱ - ۲۰۰/۱	۹۶۲ - ۱۰۶۸	۵۷	۴۸

### ۳-۲- مدل سازی شبکه عصبی

#### ۳-۲-۱- مدل سازی شبکه عصبی مصنوعی با استفاده از روش سهم گروه

#### ۳-۲-۱-۱- محاسبه توصیف کننده های هیدروکربن های خالص با استفاده از

#### روش سهم گروه

برای محاسبه توصیف کننده های هیدروکربن ها با استفاده از روش سهم گروه، ابتدا باید گروه های

ساختاری ترکیبات را مشخص نمود. پس از آنالیز ساختار این ترکیبات ۱۵ گروه عاملی در این

ترکیبات شناسایی شدند که در جدول (۳-۲) گزارش شده است.

جدول (۲-۳)- گروه‌های عاملی موجود در هیدروکربن‌های مورد مطالعه در این پایان‌نامه طبق روش سهم‌گروه

#	شناسه گروه	توضیحات
۱	$-CH_3$	کربن ( $sp^3$ ) نوع اول انتهایی (متیل)
۲	$-CH_2-$	کربن ( $sp^3$ ) نوع دوم
۳	$\rangle CH-$	کربن ( $sp^3$ ) نوع سوم
۴	$\rangle C\langle$	کربن ( $sp^3$ ) نوع چهارم
۵	$-CH_2-\textcircled{R}$	کربن ( $sp^3$ ) نوع دوم داخل حلقه اشباع
۶	$\rangle CH-\textcircled{R}$	کربن ( $sp^3$ ) نوع سوم داخل حلقه اشباع
۷	$=CH_2$	کربن ( $sp^2$ ) نوع اول انتهایی
۸	$=CH-$	کربن ( $sp^2$ ) نوع دوم
۹	$=C\langle$	کربن ( $sp^2$ ) نوع سوم
۱۰	$=CH-\textcircled{R}$	کربن ( $sp^2$ ) نوع دوم داخل حلقه
۱۱	$=CH-(Ar)$	کربن حلقه بنزنی استخلاف نشده
۱۲	$=CR-(Ar)$	کربن حلقه بنزنی استخلاف شده
۱۳	$=CR_2(2Ar)$	کربن مشترک بین دو حلقه آروماتیک
۱۴	$=Cis$	فرم سیس
۱۵	$=Trans$	فرم ترانس

پس از شناسایی توصیف‌کننده‌های سهم‌گروه برای هیدروکربن‌ها مورد مطالعه سه متغیر دما، فشار و جرم‌مولی نیز به این توصیف‌کننده‌ها اضافه شدند. سپس داده‌های مربوط به توصیف‌کننده‌های سهم‌گروه پس از دسته‌بندی و مرتب‌سازی CR برای مدل‌سازی وارد شبکه‌ی عصبی مصنوعی شدند. در روش مرتب‌سازی CR همبستگی توصیف‌کننده‌ها با متغیر وابسته (دانشیته) توسط نرم‌افزار SPSS

محاسبه می‌شود. بعد توصیف‌کننده‌ها به ترتیب کاهش در قدر مطلق مقادیر همبستگی در ماتریس ورودی مرتب می‌شوند. توصیف‌کننده‌ای که بیشترین مقدار همبستگی با متغیر وابسته را دارد زودتر از سایر توصیف‌کننده‌ها به شبکه وارد می‌شود و سایر توصیف‌کننده‌ها پس از آن در ستون‌های بعدی ماتریس ورودی شبکه قرار می‌گیرند. لازم به ذکر است که به دلیل وابستگی دانسیته به دما، فشار و جرم مولی این سه متغیر به ترتیب در ستون‌های اول تا سوم ماتریس ورودی قرار می‌گیرند و سایر توصیف‌کننده‌های سهم‌گروه به ترتیب قدر مطلق مقادیر همبستگی در ستون‌های بعدی ماتریس قرار می‌گیرند. مقادیر همبستگی توصیف‌کننده‌های حاصل از روش سهم‌گروه با دانسیته در جدول (۳-۳) ارائه شده است.

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

توصیف کننده	$-CH_3$	$-CH_2-$	$\rangle CH-$	$\rangle C\langle$	$-CH_2-\textcircled{R}$	$\rangle CH-\textcircled{R}$	$=CH_2$	$=CH-$
مقدار دانسیته با همبستگی	۰/۰۵۰	۰/۴۰۵	-۰/۲۹۴	-۰/۵۶۹	۰/۱۹۱	۰/۱۴۴	-۰/۱۰۵	-۰/۲۵۳
توصیف کننده	$=C\langle$	$=CH-\textcircled{R}$	$=CH-(Ar)$	$=CR-(Ar)$	$=CR_2(2Ar)$	$=Cis$	$=Trans$	
مقدار دانسیته با همبستگی	-۰/۱۰۹	۰/۰۳۴	۰/۵۱۷	۰/۴۲۹	۰/۲۷۷	-۰/۳۲۹	-۰/۳۳۳	

### ۳-۲-۱-۲- دسته بندی داده‌ها

در این مرحله، ابتدا از ۶۰۸۳ نقطه داده جمع‌آوری شده از مراجع مختلف برای ۷۹ هیدروکربن مختلف ۱۰ مولکول به عنوان سری تست خارجی (۴۷۳ نقطه) کنار گذاشته شدند و مابقی نقطه

داده‌ها به طور تصادفی به سه دسته‌ی آموزش (۳۴۲۰ نقطه)، ارزیابی (۱۱۴۹ نقطه) و تست داخلی (۱۰۴۱ نقطه) تقسیم بندی شدند.

### ۳-۲-۱-۳-مدل‌سازی شبکه عصبی مصنوعی با توصیف‌کننده‌های حاصل از

#### روش سهم گروه GCM-ANN

۱۵ توصیف‌کننده‌ای که از روش سهم‌گروه به دست آمدند به همراه سه متغیر دما، فشار و توصیف‌کننده‌ی جرم مولی به عنوان ورودی به شبکه عصبی داده شدند. خروجی شبکه نیز دانسیته‌ی متناظر ورودی‌ها را نشان می‌دهد. برای آموزش شبکه، دو الگوریتم آموزشی لورنبرگ-مارکوارت (lm) و تنظیم بایزین (br)، و برای انتخاب تابع انتقال مناسب، دو تابع انتقال تانژانت سیگموئید (tan sig) و لگاریتم سیگموئید (log sig) مورد استفاده قرار گرفتند که از ترکیب این دو الگوریتم و دو تابع انتقال، چهارنوع شبکه عصبی مختلف طراحی گردید. در هر یک از این شبکه‌ها باید تعداد نرون‌های لایه‌ی پنهان و تعداد دور آموزشی بهینه گردند که مقدار بهینه‌ی هر یک از این پارامترها زمانی است که میانگین مربع خطا (MSE) برای سری ارزیابی دارای کمترین مقدار باشد. برای بهینه‌سازی این دو پارامتر از روش تغییر هم‌زمان استفاده شد به این صورت که تعداد گره در لایه‌ی پنهان از ۹ تا ۲۷ با گام ۱ و تعداد دور آموزشی از ۶۰ تا ۵۰۰ با گام ۱۰ در نظر گرفته شدند. به ازای همه‌ی ترکیبات ممکن از این پارامترها، شبکه‌های طراحی شده به طور جداگانه آموزش داده شدند. بخشی از نتایج حاصل از این محاسبات حول مقدار بهینه، در جداول (۳-۴) تا (۳-۷) ارائه شده است. در هر یک از چهار شبکه‌ی طراحی شده، شبکه‌ی دارای کمترین مقدار MSE برای پیش‌بینی داده‌های سری ارزیابی، به عنوان شبکه‌ی بهینه انتخاب گردید.



جدول (۳-۴) - 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 لایه‌ی پنهان							
		۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
دور آموزشی	#	۱۱/۸۶۹	۱۰/۲۸۱	۱۳/۶۵۶	۱۳/۰۷۰	۹/۷۰۴	۹/۶۴۱	۹/۹۳۰	۱۱/۹۷۵
	۲۰۰	۱۱/۸۰۶۶	۱۰/۲۰۷	۱۳/۳۲۳	۱۲/۹۰۷	۹/۶۴۸	۹/۲۳۷	۹/۷۶۸	۱۱/۷۰۲
	۲۱۰	۱۱/۵۸۰	۱۰/۱۴۱	۱۲/۹۹۲	۱۲/۷۸۸	۹/۵۹۷	۹/۰۱۵	۹/۶۲۹	۱۱/۰۱۰
	۲۲۰	۱۱/۴۲۰	۱۰/۰۸۰	۱۲/۶۸۷	۱۲/۶۹۵	۹/۵۵۵	۹/۰۴۳	۹/۵۷۶	۱۰/۴۲۲
	۲۳۰	۱۱/۳۸۴	۱۰/۰۲۰	۱۲/۴۳۸	۱۲/۶۲۹	۹/۵۱۸	۹/۵۹۰	۹/۵۸۳	۹/۹۷۴
	۲۴۰	۱۱/۳۴۵	۹/۹۷۶	۱۲/۲۴۰	۱۲/۵۸۹	۹/۳۸۹	۹/۳۴۰	۹/۶۳۸	۹/۸۸۰
	۲۵۰	۱۱/۳۱۱	۹/۹۶۰	۱۲/۰۵۸	۱۲/۴۹۲	۹/۱۵۳	۱۰/۰۱۱	۹/۶۹۵	۹/۷۹۵
	۲۶۰	۱۱/۲۴۵	۹/۹۹۰	۱۱/۹۴۲	۱۲/۳۶۰	۸/۹۹۷	۱۰/۰۵۷	۹/۷۵۶	۹/۷۲۶
	۲۷۰	۱۱/۰۶۰	۱۰/۰۶۲	۱۱/۸۷۲	۱۲/۳۱۸	۸/۸۸۴	۱۰/۰۷۰	۹/۷۸۱	۹/۶۷۴
	۲۸۰	۱۰/۸۶۸	۱۰/۳۶۰	۱۱/۸۱۸	۱۲/۲۶۴	<b>۸/۸۲۳</b>	۱۰/۱۲۸	۹/۷۷۸	۹/۶۲۸
	۲۹۰	۱۰/۷۱۹	۱۰/۲۴۹	۱۱/۷۶۹	۱۲/۲۰۰	۸/۸۶۶	۱۰/۱۰۴	۹/۷۶۱	۹/۵۶۰
	۳۰۰	۱۰/۵۴۱	۱۰/۱۹۸	۱۱/۷۲۱	۱۱/۴۶۱	۸/۸۷۱	۱۰/۰۱۰	۹/۷۳۹	۹/۶۷۷
	۳۱۰	۱۰/۲۹۶	۱۰/۱۷۶	۱۱/۶۷۷	۱۱/۰۵۳	۸/۸۸۷	۱۰/۰۱۹	۹/۷۱۹	۹/۸۰۵
	۳۲۰	۱۰/۰۷۰	۱۰/۱۴۰	۱۱/۶۳۶	۱۰/۷۱۶	۹/۰۴۹	۱۰/۳۲۲	۹/۶۹۹	۹/۸۴۹
	۳۳۰	۹/۹۵۲	۱۰/۲۱۰	۱۱/۶۰۲	۱۰/۷۲۴	۹/۴۲۴	۱۰/۲۹۲	۹/۶۶۱	۹/۸۵۰
	۳۴۰	۹/۸۸۴	۱۰/۳۲۶	۱۱/۵۸۰	۱۰/۶۹۸	۹/۶۳۰	۱۰/۲۳۱	۹/۵۲۷	۹/۸۵۳
	۳۵۰								

جدول (۳-۵) - 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 لایه‌ی پنهان							
دور آموزشی	#	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
	۲۰۰	۱۱/۶۲۲	۷/۵۱۶	۱۴/۸۱۵	۳۳/۲۵۵	۱۲/۳۸۸	۱۰/۲۹۵	۸/۹۵۶	۹/۱۱۸
	۲۱۰	۱۱/۵۴۹	۷/۳۸۶	۱۴/۶۸۴	۳۲/۴۸۸	۱۲/۳۱۵	۱۰/۱۲۱	۸/۷۰۰	۹/۰۰۴
	۲۲۰	۱۱/۴۶۵	۷/۳۱۰	۱۳/۸۲۱	۳۱/۴۰۰	۱۲/۲۴۳	۹/۹۶۴	۸/۵۹۶	۹/۲۵۹
	۲۳۰	۱۱/۳۸۱	۷/۲۶۱	۱۳/۴۴	۳۰/۳۷۴	۱۱/۹۲۴	۹/۷۹۱	۸/۴۱۶	۹/۳۲۹
	۲۴۰	۱۱/۲۸۷	۷/۲۳۴	۱۲/۱۲۴	۲۹/۸۶۷	۱۱/۵۶۴	۹/۶۰۵	۸/۳۲۸	۹/۴۵۰
	۲۵۰	۱۱/۱۵۷	۷/۲۲۳	۱۰/۷۶۵	۲۹/۵۰۸	۱۱/۳۰۱	۹/۴۴۴	۸/۳۲۱	۹/۵۵۶
	۲۶۰	۱۰/۹۷۸	۷/۲۲۴	۱۰/۶۹۶	۲۹/۳۹۵	۱۱/۱۰۰	۹/۳۰۸	۸/۳۲۰	۹/۵۹۴
	۲۷۰	۱۰/۷۶۱	۷/۲۳۱	۱۰/۵۴۳	۲۹/۳۹۰	۱۱/۰۲۵	۹/۱۸۳	۸/۳۱۸	۹/۵۹۴
	۲۸۰	۱۰/۵۲۸	۷/۲۳۹	۱۰/۴۵۸	۲۹/۲۹۱	۱۰/۵۸۷	۹/۰۵۲	۸/۳۰۰	۹/۵۹۶
	۲۹۰	۹/۸۹۰	۷/۲۵۰	۱۰/۴۵۴	۲۹/۳۵۴	۱۰/۴۵۸	۸/۹۰۳	۸/۲۱۴	۹/۶۱۵
	۳۰۰	۸/۵۸۸	۷/۲۶۶	۱۰/۸۶۷	۲۸/۲۸۷	۱۰/۰۰۸	۸/۷۴۶	۸/۲۰۸	۹/۶۴۴
	۳۱۰	۸/۷۶۰	۷/۳۰۱	۱۱/۰۵۳	۲۷/۹۹۸	۹/۶۸۴	۸/۶۰۰	۸/۰۷۶	۹/۶۷۶
	۳۲۰	۱۰/۱۶۰	۷/۳۳۰	۱۱/۰۸۴	۲۷/۹۶۱	۹/۵۸۷	۸/۴۶۲	۸/۲۵۷	۹/۷۰۸
	۳۳۰	۱۱/۹۶۲	۷/۳۳۲	۱۱/۱۱۴	۲۷/۸۳۹	۹/۳۹۰	۸/۳۴۰	۸/۲۸۹	۹/۷۴۲
	۳۴۰	۱۱/۹۶۳	۷/۳۳۴	۱۱/۱۳۸	۲۷/۰۳۷	۹/۲۲۰	۸/۲۳۴	۸/۴۹۱	۹/۷۷۹
۳۵۰	۱۱/۹۶۶	۷/۳۴۷	۱۱/۱۲۲	۲۶/۷۳۰	۹/۱۰۳	۸/۱۵۸	۸/۵۰۷	۹/۸۲۰	

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 لایه‌ی پنهان								
دور آموزشی	#	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰	
	۲۰۰	۱۱/۳۰۲	۱۰/۱۹۰	۱۲/۲۸۴	۸/۵۷۷	۱۰/۰۸۰	۱۴/۰۹۶	۱۲/۷۸۸	۹/۶۴۳	
	۲۱۰	۱۱/۰۰۶	۹/۹۵۱	۱۲/۰۶۰	۸/۵۱۴	۱۰/۰۳۶	۱۲/۵۳۶	۱۲/۵۶۶	۹/۶۳۳	
	۲۲۰	۱۰/۷۹۰	۹/۷۳۵	۱۱/۵۸۲	۸/۵۱۷	۹/۸۸۳	۱۱/۱۱۹	۱۲/۳۳۷	۹/۶۲۴	
	۲۳۰	۱۰/۶۰۲	۹/۵۷۳	۱۱/۱۳۰	۸/۶۰۷	۹/۶۷۴	۱۰/۲۴۵	۱۲/۱۱۶	۹/۶۱۶	
	۲۴۰	۱۰/۳۷۰	۹/۴۹۴	۱۰/۸۶۵	۸/۷۱۸	۹/۴۹۱	۱۰/۰۱۷	۱۱/۹۱۳	۹/۴۷۱	
	۲۵۰	۱۰/۲۱۰	۹/۴۷۰	۱۰/۷۵۲	۸/۸۹۳	۹/۴۰۰	۱۰/۰۲۴	۱۱/۴۶۲	۹/۳۳۶	
	۲۶۰	۱۰/۰۹۰	۹/۴۷۶	۱۰/۶۷۶	۸/۹۳۷	۹/۴۰۳	۱۰/۰۸۰	۱۰/۶۵۹	۹/۲۰۳	
	۲۷۰	۱۰/۰۲۸	۹/۵۰۰	۱۰/۵۷۰	۸/۹۵۵	۹/۵۷۴	۱۰/۱۱۹	۱۰/۴۸۷	۹/۱۲۳	
	۲۸۰	۱۰/۰۱۲	۹/۵۱۱	۱۰/۴۰۵	۸/۹۶۵	۹/۷۵۱	۱۰/۱۳۵	۱۰/۵۱۱	۹/۱۲۲	
	۲۹۰	۱۰/۰۰۶	۹/۵۳۲	۱۰/۲۳۲	۸/۹۸۵	۱۰/۰۰۶	۱۰/۱۳۳	۹/۶۶۲	۹/۱۰۲	
	۳۰۰	۹/۹۸۱	۹/۵۶۵	۱۰/۰۸۸	۹/۰۱۸	۱۰/۱۸۹	۱۰/۱۰۸	۹/۴۷۴	۹/۱۹۹	
	۳۱۰	۹/۹۳۱	۹/۵۹۵	۹/۹۵۰	۹/۰۵۴	۱۰/۴۳۲	۱۰/۰۵۳	۹/۳۵۴	۹/۳۳۸	
	۳۲۰	۹/۸۷۰	۹/۶۴۴	۹/۹۰۲	۹/۰۸۶	۱۰/۷۳۰	۹/۹۸۸	۹/۳۱۲	۹/۵۰۶	
	۳۳۰	۹/۸۲۳	۹/۸۱۷	۹/۹۴۲	۹/۱۱۱	۱۰/۹۳۷	۹/۹۱۸	۹/۳۸۷	۹/۷۵۷	
	۳۴۰	۹/۸۱۰	۹/۸۸۸	۹/۹۸۰	۹/۱۳۱	۱۱/۰۸۲	۹/۸۲۶	۹/۵۵۱	۹/۹۹۳	
	۳۵۰	۹/۸۲۰	۹/۸۸۹	۱۰/۰۶۰	۹/۱۴۷	۱۱/۱۸۹	۹/۷۳۴	۹/۷۸۶	۱۰/۱۳۴	

جدول (۳-۷) - 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

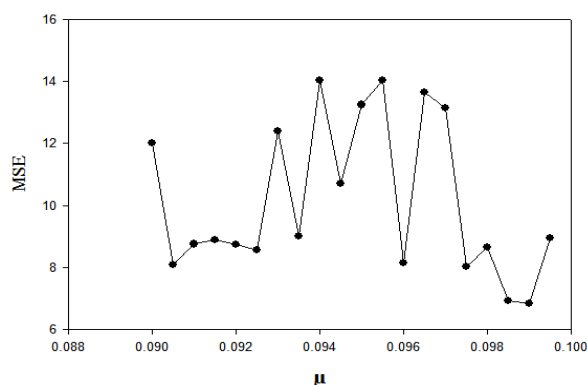
تعداد گره لایه‌ی پنهان									
دور آموزشی	#	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
	۲۰۰	۲۳/۹۲۶	۹/۶۵۵	۱۴/۷۴۶	۹/۳۸۸	۷/۹۸۹	۱۱/۸۰۲	۱۵/۰۱۱	۱۱/۴۱۵
	۲۱۰	۲۳/۷۳۰	۹/۶۴۷	۱۴/۶۵۳	۸/۹۸۳	۷/۷۵۳	۱۱/۸۵۷	۱۴/۰۲۶	۱۱/۲۷۹
	۲۲۰	۲۳/۴۷۷	۹/۶۱۸	۱۴/۱۷۶	۸/۴۴۰	۷/۶۵۶	۱۱/۹۱۲	۱۳/۰۱۱	۱۱/۱۶۷
	۲۳۰	۲۳/۲۶۴	۹/۵۷۶	۱۳/۹۸	۸/۳۷۵	۷/۶۵۱	۱۱/۹۴۰	۱۲/۳۲۹	۱۱/۱۵۹
	۲۴۰	۲۳/۱۰۰	۹/۵۲۹	۱۳/۶۷۱	۸/۳۵۵	۷/۵۹۲	۱۱/۹۴۰	۱۱/۸۶۷	۱۰/۱۳۹
	۲۵۰	۲۲/۹۴۰	۹/۴۸۶	۱۳/۱۹۲	۸/۳۱۳	۷/۵۰۵	۱۱/۹۲۵	۱۱/۴۹۲	۱۰/۰۷۹
	۲۶۰	۲۲/۸۰۲	۹/۴۴۹	۱۲/۸۴۷	۸/۲۷۴	۷/۴۱۸	۱۱/۹۰۶	۱۱/۲۱۰	۱۰/۱۷۲
	۲۷۰	۲۲/۶۷۶	۹/۲۵۱	۱۲/۵۹۷	۸/۲۴۱	۷/۳۸۷	۱۱/۸۸۹	۱۱/۰۴۷	۱۰/۱۷۶
	۲۸۰	۲۲/۵۷۷	۸/۸۹۷	۱۲/۳۹۹	۸/۱۹۰	۷/۳۰۱	۱۱/۸۷۲	۱۰/۶۲۱	۱۰/۲۲۶
	۲۹۰	۲۲/۴۹۲	۸/۸۷۴	۱۲/۲۱۱	۸/۱۶۲	<b>۷/۲۷۰</b>	۱۱/۸۵۳	۱۰/۴۱۹	۱۰/۳۶۵
	۳۰۰	۲۲/۴۱۳	۸/۸۵۵	۱۲/۱۹۳	۸/۱۴۵	۷/۴۴۱	۱۱/۷۱۸	۹/۹۶۲	۱۰/۵۶۰
	۳۱۰	۲۲/۳۳۸	۸/۸۳۸	۱۲/۲۱۱	۸/۱۲۸	۷/۶۰۶	۱۰/۹۴۹	۹/۲۱۰	۱۰/۶۸۷
	۳۲۰	۲۲/۲۵۴	۸/۶۰۷	۱۲/۲۱۶	۸/۱۴۱	۷/۷۱۶	۱۰/۶۱۵	۹/۱۱۹	۱۰/۷۴۱
	۳۳۰	۲۲/۰۸۶	۸/۳۵۵	۱۲/۲۱۸	۸/۱۶۸	۷/۸۰۸	۱۰/۵۷۸	۸/۹۸۰	۱۰/۷۴۷
	۳۴۰	۲۱/۸۷۵	۸/۲۲۴	۱۲/۲۱۸	۸/۱۸۱	۷/۸۹۲	۱۰/۷۰۲	۹/۰۷۶	۱۰/۷۴۸
	۳۵۰	۲۱/۳۳۹	۸/۱۳۷	۱۲/۲۱۸	۸/۱۸۵	۷/۹۸۶	۱۰/۷۰۰	۹/۱۲۹	۱۰/۷۴۸

توابع و مقادیر پارامترهای بهینه شده برای شبکه‌های عصبی مختلف با توصیف‌کننده‌های روش سهم گروه (GCM-ANN) برای پیش‌بینی دانسیته‌ی ترکیبات هیدروکربنی خالص در جدول (۸-۳) ارائه شده است.

جدول (۸-۳) - توابع و مقادیر بهینه‌ی شبکه‌های GCM-ANN

MSE	تعداد دور آموزش	تعداد نرون لایه ی پنهان	تعداد توصیف کننده	تابع آموزشی	تابع انتقال
۸/۸۲۳	۲۹۰	۱۷	۱۸	تنظیم بایزین	لگاریتم سیگموئید
۷/۲۲۳	۲۵۰	۱۴	۱۸	لونبرگ-مارکوارت	لگاریتم سیگموئید
۸/۵۱۴	۲۱۰	۱۶	۱۸	تنظیم بایزین	تانژانت سیگموئید
۷/۲۷۰	۲۹۰	۱۷	۱۸	لونبرگ-مارکوارت	تانژانت سیگموئید

با توجه به نتایج موجود در جدول (۸-۳) شبکه‌ی با الگوریتم آموزشی لونبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئیدی با تعداد ۱۴ نرون در لایه‌ی پنهان و دور آموزشی ۲۵۰ به عنوان شبکه‌ی بهینه انتخاب گردید. در مرحله‌ی بعد برای تعیین مقدار بهینه  $\mu$ ، با استفاده از شبکه‌ی بهینه، مقدار  $\mu$  از ۰/۰۰۰۵ تا ۰/۱ با گام ۰/۰۰۰۵ تغییر داده شده و بر اساس مقدار حداقل مربعات خطا، MSE، مقدار بهینه‌ی پارامتر  $\mu$ ، ۰/۰۹۹۰ به دست آمد. نمودار مقدار میانگین مربعات خطا بر حسب  $\mu$  حول نقطه‌ی بهینه نیز در شکل (۱-۳) رسم شده است. توابع و مقادیر پارامترهای بهینه شده‌ی GCM-ANN برای پیش‌بینی دانسیته‌ی هیدروکربن‌ها، در جدول (۹-۳) ارائه شده است.



شکل (۳-۱) - نمودار میانگین مربعات خطا بر حسب  $\mu$  حول نقطه بهینه برای سری ارزیابی

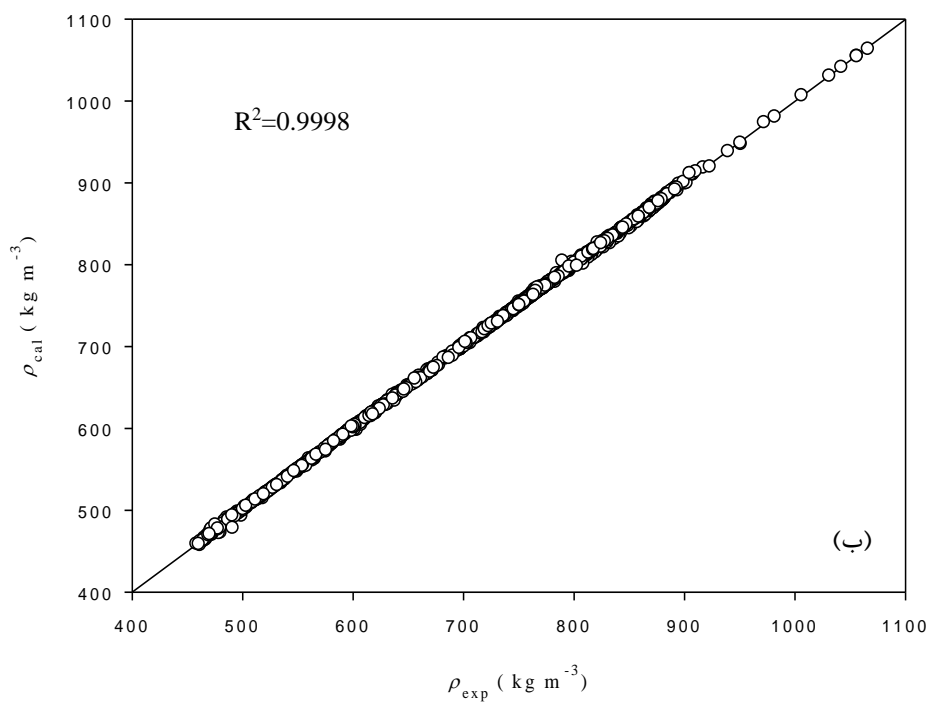
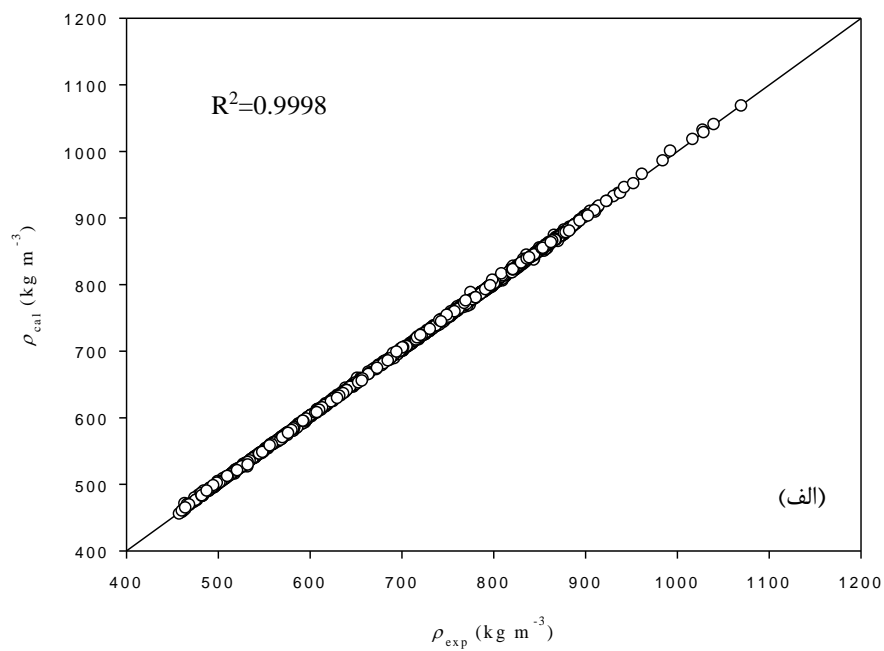
جدول (۳-۹) - توابع و مقادیر پارامترهای شبکه‌ی GCM-ANN بهینه شده برای پیش‌بینی دانسیته‌ی هیدروکربن‌ها بر حسب دما و فشار

Train lm	تابع آموزش
Log sig	تابع انتقال لایه پنهان
Pure line	تابع انتقال لایه خروجی
۱۴	تعداد نرون لایه پنهان
۱۸	تعداد متغیرهای ورودی
۲۵۰	تعداد دورهای آموزش
۰/۰۹۹۰	پارامتر $\mu$
۶/۸۲۴۰	مقدار MSE سری ارزیابی

### ۳-۲-۱-۴- ارزیابی شبکه GCM-ANN بهینه شده

برای ارزیابی شبکه‌ی GCM-ANN بهینه شده، از دو روش نمودار برگشتی و نمودار خطای باقیمانده استفاده گردید. در نمودار برگشتی که در شکل (۳-۲) آمده است مقادیر پیش‌بینی شده‌ی دانسیته بر حسب مقادیر تجربی رسم گردیده در این نمودارها مقدار ضریب تعیین  $R^2$  که پراکندگی نقاط در اطراف خط برگشت را نشان می‌دهد نیز آمده است. همانطور که در این شکل‌ها مشاهده

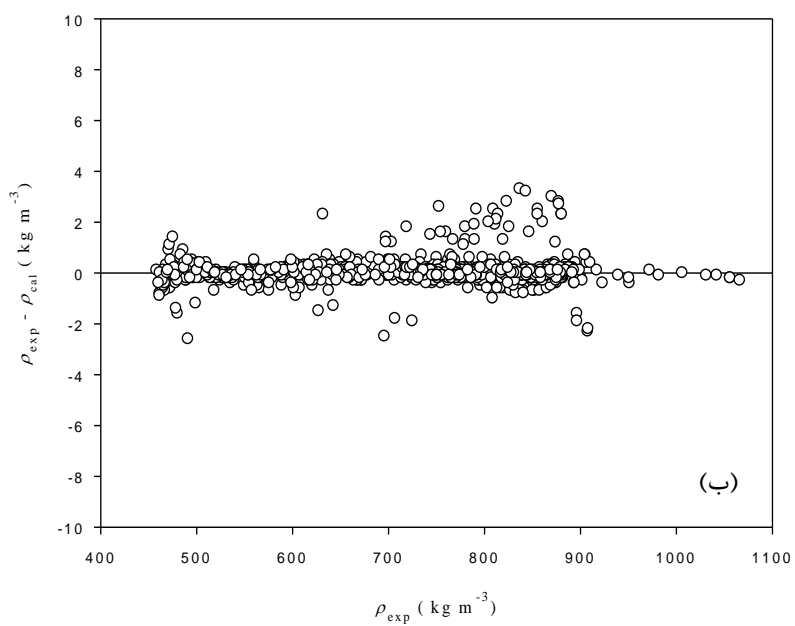
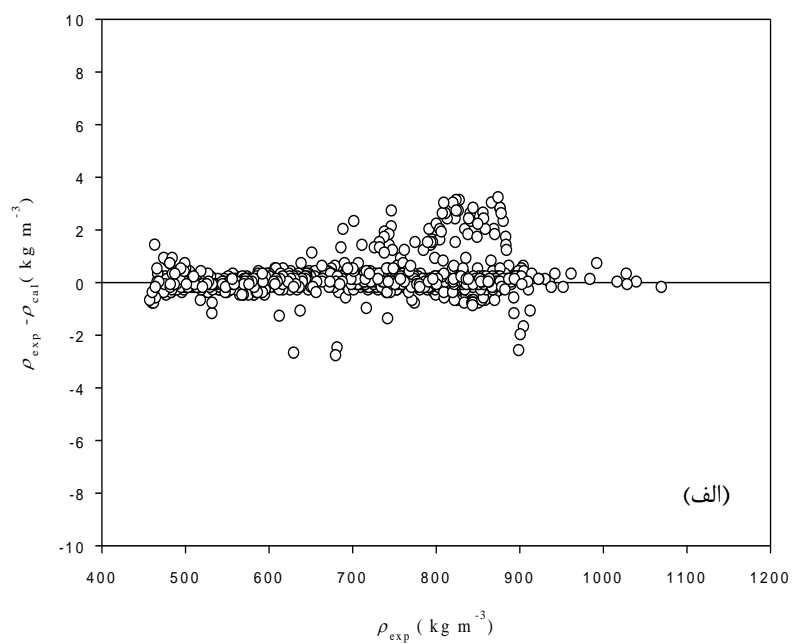
می‌شود، ضریب تعیین،  $R^2$ ، بدست آمده برای دو سری ارزیابی و سری تست داخلی نزدیک به یک است که این موضوع بر توافق عالی بین مقادیر دانسیته‌ی پیش‌بینی شده توسط این روش با مقادیر تجربی دلالت دارد. نمودار خطای باقیمانده نیز در شکل (۳-۲) آمده است که پراکندگی متقارن داده‌ها حول خطای صفر در این شکل‌ها نشان از عدم وجود خطای سیستماتیک در داده‌ها دارد.



شکل (۲-۳) - نمودار برگشتی مقادیر پیش بینی شده‌ی دانسیته‌ی هیدروکربن‌ها بر حسب مقادیر تجربی برای

(الف) سری ارزیابی و (ب) سری تست داخلی





شکل (۳-۳) - نمودار خطای باقیمانده بر حسب مقادیر تجربی دانسیته‌ی هیدروکربن‌ها

(الف) سری ارزیابی و (ب) سری تست داخلی

### ۳-۲-۲- مدل سازی شبکه عصبی مصنوعی با استفاده از توصیف کننده های تئوری

#### ۳-۲-۲-۱- رسم و بهینه سازی ساختار مولکول ها و محاسبه توصیف کننده ها

ابتدا ساختار مولکول ها با استفاده از نرم افزار هایپرکم رسم و بعد با استفاده از روش نیمه تجربی AM1 ساختار سه بعدی ترکیبات بهینه شد. در این مرحله ۱۹ توصیف کننده برای هر مولکول استخراج گردید. در ادامه ساختار بهینه شده ی مولکول ها به عنوان ورودی به نرم افزار دراگون داده شد و برای هر مولکول ۳۲۲۴ توصیف کننده تئوری به دست آمد.

بنابراین ماتریس نهایی توصیف کننده ها شامل ۳۲۴۳ ردیف است. از آنجا که این تعداد توصیف کننده برای مدل سازی بسیار زیاد است، باید به دنبال روشی جهت کاهش تعداد توصیف کننده ها بود تا از انجام محاسبات وقت گیر و بیهوده جلوگیری شده و از پیچیدگی کار پرهیز شود. روش مورد استفاده باید قادر باشد تا توصیف کننده هایی که بیشترین ارتباط را با متغیر وابسته، یعنی دانسیته، دارند انتخاب کرده و سایر توصیف کننده ها را حذف نماید. به همین منظور، ابتدا توصیف کننده هایی که برای تمام مولکول ها، مقادیر ثابت و یا تقریباً ثابت داشتند، حذف شدند. سپس به وسیله ی برنامه ی نوشته شده در نرم افزار MATLAB توصیف کننده هایی که همبستگی بزرگتر از ۰/۹ داشتند، نیز حذف گردید. در پایان این مرحله ۱۱۸ توصیف کننده از مجموع ۳۲۴۳ توصیف کننده ی تئوری باقی ماندند. در این مرحله دو متغیر تجربی دما، فشار و یک توصیف کننده جرم مولی نیز به ماتریس ورودی اضافه شدند.

هدف بعدی انتخاب بهترین توصیف کننده ها از بین توصیف کننده های باقیمانده برای ساخت مدل است. در این پایان نامه از روش غیرخطی الگوریتم ژنتیک برای انتخاب توصیف کننده های مناسب استفاده شد که در ادامه در مورد آن توضیح داده خواهد شد.

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

به منظور اجرای برنامه‌ی الگوریتم ژنتیک، تمام ۱۱۸ توصیف‌کننده باقی مانده به همراه سه توصیف‌کننده دما، فشار و جرم مولی 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 بیش از ۵۰٪ در این الگوریتم تکرار شده‌اند به همراه تعداد دفعات تکرار در جدول (۳-۱۰) نشان داده شده‌اند. طبقه و نام کامل آن‌ها نیز در جدول (۳-۱۱) آورده شده است. همچنین ماتریس همبستگی بین این توصیف‌کننده‌ها در جدول (۳-۱۲) ارائه شده که این ماتریس عدم همبستگی بین توصیف‌کننده‌ها را نشان می‌دهد.

جدول (۳-۱۰) - تعداد دفعات تکرار توصیف‌کننده‌ها در روش انتخابی الگوریتم ژنتیک پس از ۱۶ بار اجرا

توصیف کننده	temperature	pressure	mass	Ms	R1m	PW4	RTm	SHP2	X0Av
تعداد تکرار	۱۶	۱۶	۱۶	۱۴	۱۶	۱۶	۱۵	۱۴	۱۴
توصیف کننده	PW3	R7u	GGI3	SAG	G2s	AMW	HF		
تعداد تکرار	۱۳	۱۶	۱۶	۱۶	۱۶	۱۶	۱۶		

جدول (۳-۱۱) - طبقه و نام کامل توصیف‌کننده‌های تئوری انتخابی توسط الگوریتم ژنتیک

شماره	نشان	طبقه	نام کامل
۱	R1m	GETAWAY descriptors	R autocorrelation of lag 1 / weighted by atomic masses
۲	PW4	Topological descriptors	Path/walk 4-Randic shape index
۳	Ms	Constitutional descriptors	Mean electrotopological state
۴	RTm	GETAWAY descriptors	R total index / weighted by atomic masses
۵	SHP2	Randic molecular profiles	Average shape profile index of order 2
۶	X0Av	Connectivity indices	Average valance connectivity index chi-0
۷	PW3	Topological descriptors	Path/walk 3-Randic shape index
۸	R7u	GETAWAY descriptors	R autocorrelation of lag 7 / unweighted
۹	GGI3	Topological charge indices	Topological charge index of order 3
۱۰	SAG	Electronic descriptors	Surface Area (Approx)
۱۱	G2s	WHIM descriptors	2rd component symmetry directional WHIM index / weighted by atomic electrotopological states
۱۲	AMW	Constitutional descriptors	Average Molecular weight
۱۳	HF	Electronic descriptors	Heat of formation

جدول (۳-۱۲) - همبستگی بین توصیف‌کننده‌های انتخابی الگوریتم ژنتیک

	Ms	R1m	PW4	RTm	SHP2	X0Av	PW3	R7u	GGI3	SAG	G2s	AMW	HF
Ms	۱												
R1m	۰/۱۲۳	۱											
PW4	-۰/۷۰۷	۰/۲۳۳	۱										
RTm	۰/۳۹۳	-۰/۸۵۲	۰/۰۳۱	۱									
SHP2	۰/۵۶۰	۰/۰۲۸	-۰/۶۵۱	۰/۲۹۵	۱								
X0Av	۰/۱۴۶	-۰/۴۸۲	-۰/۶۹۷	-۰/۳۶۰	۰/۶۰۹	۱							
PW3	-۰/۶۱۸	۰/۵۰۰	۰/۷۵۷	۰/۱۱۸	-۰/۴۶۹	-۰/۵۷۳	۱						
R7u	-۰/۲۴۲	۰/۲۹۷	۰/۶۰۳	۰/۲۰۵	-۰/۲۷۴	-۰/۵۴۷	۰/۴۵۱	۱					
GGI3	-۰/۲۳۹	۰/۱۸۴	۰/۵۶۳	۰/۱۱۴	-۰/۰۹۷	-۰/۲۱۶	۰/۴۴۹	۰/۲۸۹	۱				
SAG	-۰/۰۶۱	-۰/۲۲۱	-۰/۳۷۴	-۰/۳۹۲	۰/۱۶۰	۰/۵۳۷	-۰/۱۳۵	-۰/۴۳۲	-۰/۲۳۰	۱			
G2s	۰/۱۸۱	-۰/۴۲۰	-۰/۲۰۱	-۰/۲۸۳	۰/۱۴۰	۰/۰۶۷	-۰/۲۵۸	-۰/۲۳۶	-۰/۲۵۰	۰/۰۱۹	۱		
AMW	۰/۲۲۱	۰/۵۱۷	۰/۴۰۹	۰/۵۵۵	-۰/۱۶۷	-۰/۸۰۳	۰/۳۷۷	۰/۴۶۶	۰/۳۴۳	-۰/۶۲۳	۰/۰۴۴	۱	
HF	-۰/۰۰۲	۰/۱۲۵	-۰/۰۳۵	۰/۱۰۶	-۰/۱۱۷	-۰/۲۵۸	۰/۰۸۳	۰/۱۳۶	-۰/۶۵۲	-۰/۱۱۸	۰/۱۱۲	۰/۱۱۶	۱

### ۳-۲-۲-۳- بهینه‌سازی شبکه عصبی مصنوعی با توصیف‌کننده‌های الگوریتم

#### ژنتیک

شبکه عصبی مصنوعی مورد مطالعه در این بخش، شبکه‌ای سه لایه است. در این مرحله شبکه‌هایی با 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 هر یک از شبکه‌های طراحی شده پارامترهای شبکه بهینه‌سازی گردیدند. برای این منظور شبکه‌هایی با مقدار  $\mu$  ثابت و تعداد ورودی‌های از ۴ تا ۱۶ توصیف‌کننده با گام ۱، تعداد گره لایه پنهان از ۶ تا ۲۴ با گام ۱ و تعداد دور آموزشی از ۶۰ تا ۶۰۰ با گام ۱۰ در نظر گرفته شدند و به ازای همه ترکیب‌های ممکن از این سه پارامتر، چهار شبکه‌ی عصبی مصنوعی طراحی شده، بهینه‌سازی گردیدند. لازم به ذکر است که در اینجا نیز توصیف‌کننده‌های انتخاب شده توسط روش GA که در جدول (۳-۱۰) نمایش داده شده‌اند، به روش رتبه‌بندی همبستگی، CR، به شبکه‌ی عصبی وارد شدند. در جدول (۳-۱۳) مقادیر همبستگی توصیف‌کننده‌های تئوری انتخاب شده توسط روش GA با متغیر وابسته، یعنی دانسیته، گزارش شده است.

جدول (۳-۱۳)- مقادیر همبستگی توصیف‌کننده‌های تئوری انتخابی با روش GA با دانسیته

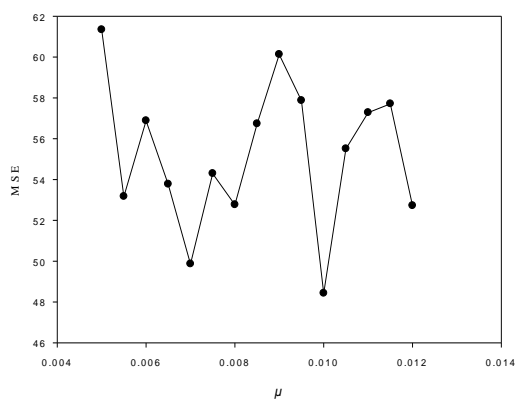
توصیف کننده	Ms	R1m	PW4	RTm	SHP2	X0Av	PW3	R7u	GGI3
مقدار همبستگی	-۰/۶۱۲	۰/۲۷۱	۰/۸۳۸	۰/۰۸۱	-۰/۵۳۱	-۰/۵۸۸	۰/۷۰۶	۰/۵۷۴	۰/۴۸۰
توصیف کننده	SAG	G2s	AMW	HF					
مقدار همبستگی	-۰/۳۴۰	-۰/۲۷۵	۰/۴۰۵	۰/۰۱۹					

مقدار پارامترهای بهینه شده هر یک از شبکه‌های GA-ANN طراحی شده در جدول (۳-۱۴) گزارش شده است. نتایج موجود در این جدول نشان می‌دهد که از میان چهار شبکه بهینه شده، شبکه‌ای با الگوریتم لونیبرگ-مارکوارت و تابع انتقال لگاریتم سیگموئید با ۱۱ توصیف‌کننده در لایه‌ی ورودی، ۱۲ گره در لایه‌ی پنهان و دور آموزشی ۵۱۰، کمترین میانگین مربع خطا را داشته و به عنوان شبکه بهینه انتخاب می‌شود. ۱۱ توصیف‌کننده انتخاب شده توسط شبکه عبارتند از دما، فشار، جرم مولی، .AMW, .GGI3, .SHP2, .PW3, .X0Av, .Ms, .R7u, .PW4

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

MSE	تعداد دور آموزش	تعداد نرون لایه‌ی پنهان	تعداد توصیف‌کننده	تابع آموزشی	تابع انتقال
۴۹/۵۴۰	۵۶۰	۱۸	۱۰	تنظیم بایزین	لگاریتم سیگموئید
۴۸/۵۰۰	۵۱۰	۱۲	۱۱	لونیبرگ-مارکوارت	لگاریتم سیگموئید
۵۰/۳۰۲	۱۳۰	۱۹	۱۰	تنظیم بایزین	تانژانت سیگموئید
۴۸/۸۳۰	۱۵۰	۱۵	۱۰	لونیبرگ-مارکوارت	تانژانت سیگموئید

در ادامه‌ی بهینه کردن شبکه مورد بررسی، پارامتر  $\mu$  نیز بایستی بهینه شود که بدین منظور، در ساختار شبکه‌ی بهینه با ۱۱ متغیر ورودی، ۱۲ گره در لایه‌ی پنهان، الگوریتم آموزشی لونیبرگ-مارکوارت، تابع انتقال لگاریتم سیگموئیدی و دور آموزشی ۵۱۰، مقدار  $\mu$  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 براساس نتایج محاسبات مقدار ۰/۰۱۰ به عنوان مقدار بهینه‌ی  $\mu$  انتخاب شد. نمودار مقدار میانگین مربعات خطا بر حسب  $\mu$  حول نقطه بهینه در شکل (۳-۴) رسم شده است.



شکل (۳-۴) - نمودار میانگین مربعات خطای حاصل از پیش‌بینی دانسیته سری ارزیابی بر حسب پارامتر  $\mu$  حول نقطه‌ی بهینه

توابع و مقادیر پارامترهای بهینه شده‌ی شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های منتخب الگوریتم ژنتیک (GA-ANN) در جدول (۳-۱۵) گزارش شده است.

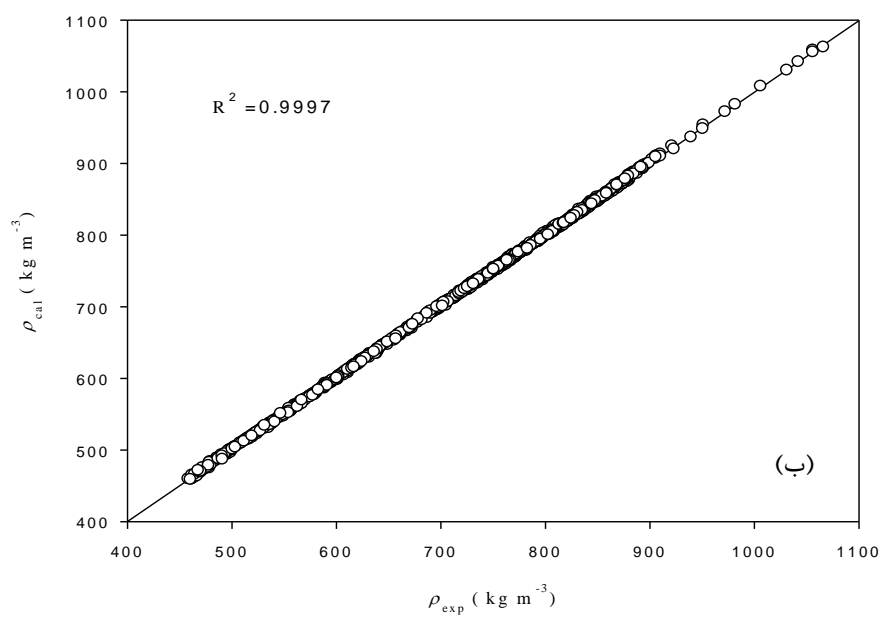
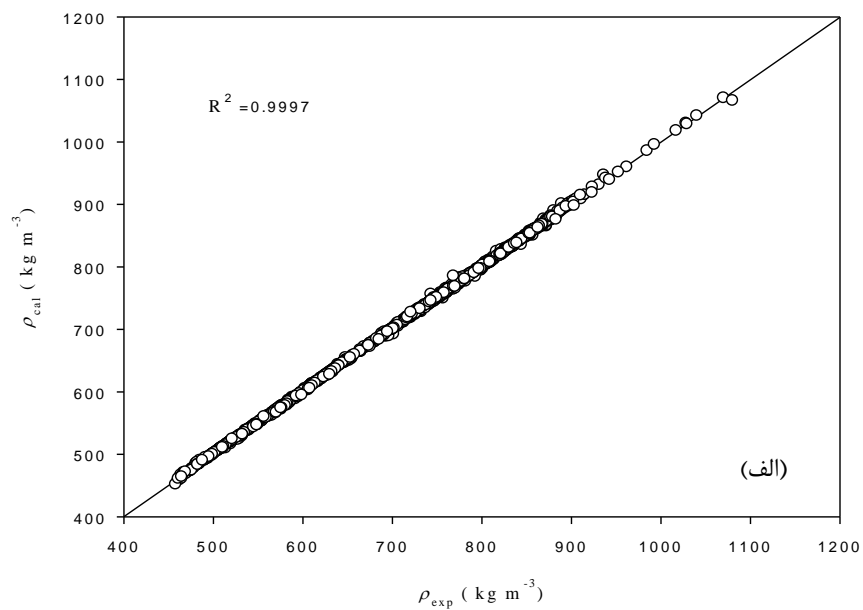
جدول (۳-۱۵) - توابع و پارامترهای بهینه شده شبکه‌ی عصبی با استفاده از توصیف‌کننده‌های تئوری انتخابی GA

Train lm	تابع آموزش
Log sig	تابع انتقال لایه پنهان
Pure line	تابع انتقال لایه خروجی
۱۲	تعداد نرون لایه پنهان
۱۱	تعداد متغیرهای ورودی
۵۱۰	تعداد دورهای آموزش
۰/۰۱۰	پارامتر $\mu$
۴۸/۴۲۵	مقدار MSE سری ارزیابی



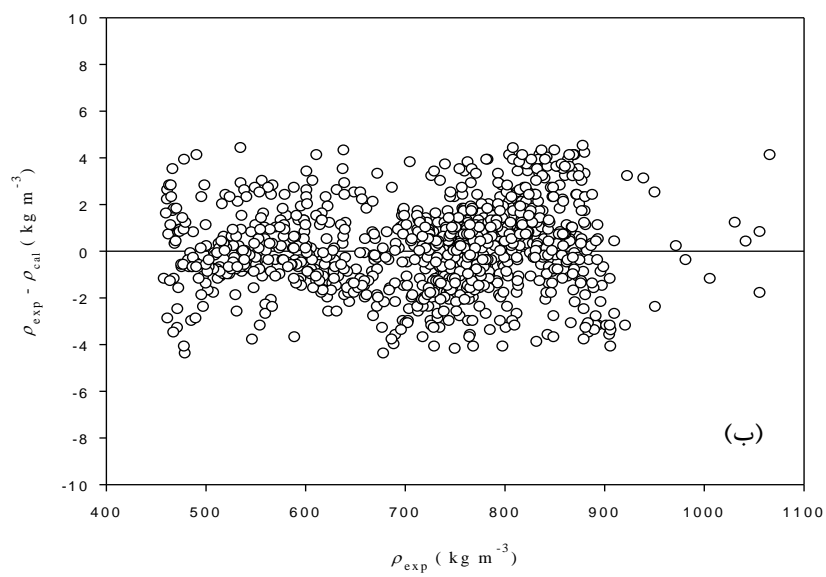
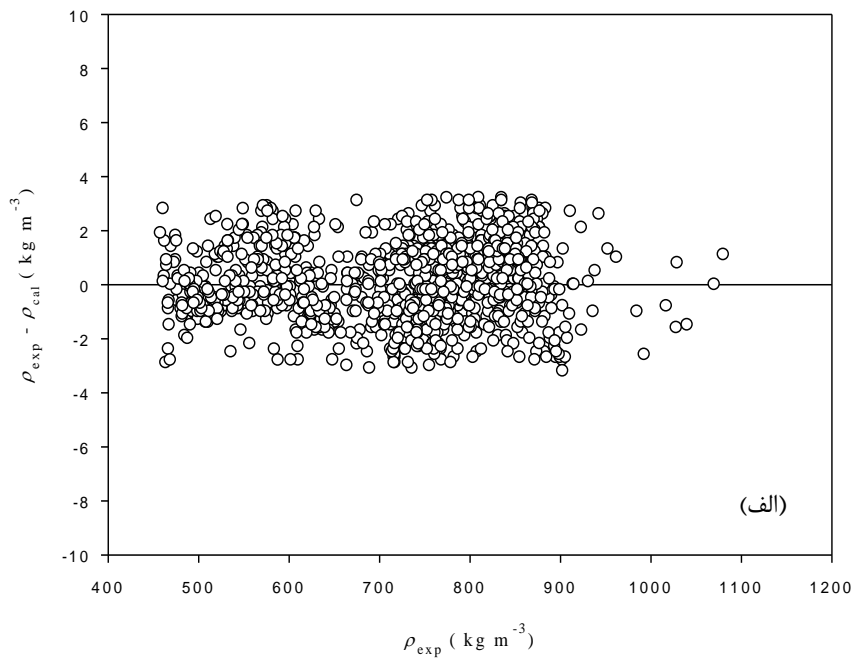
### ۳-۲-۴-ارزیابی شبکه‌ی بهینه شده‌ی GA-ANN

برای ارزیابی شبکه‌ی بهینه شده‌ی GA-ANN، نمودارهای برگشتی و خطای باقیمانده برای دو سری ارزیابی و تست داخلی رسم شدند. این نمودارها در شکل‌های (۳-۵) و (۳-۶) نشان داده شده‌اند.



شکل (۳-۵) - نمودار برگشتی مقادیر پیش‌بینی شده‌ی دانسیته‌ی هیدروکربن‌ها بر حسب مقادیر تجربی

برای (الف) سری ارزیابی و (ب) سری تست داخلی

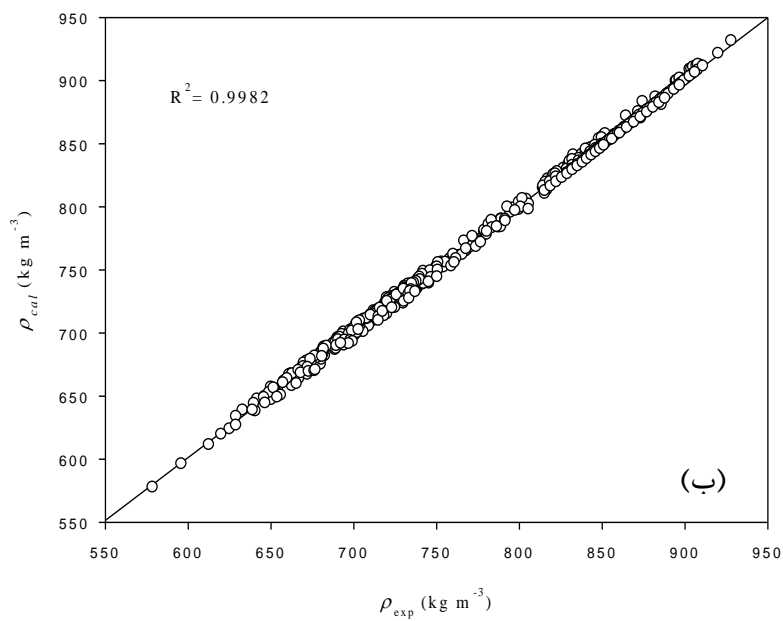
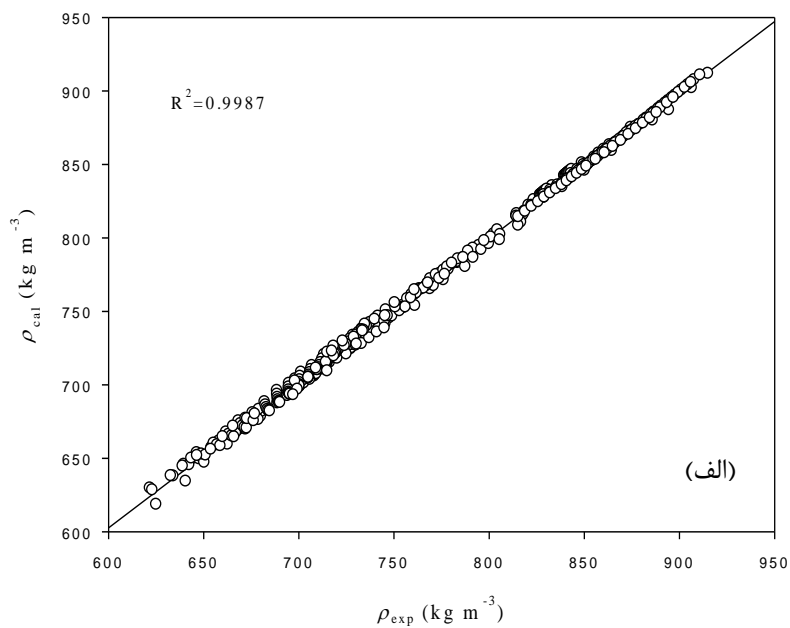


شکل (۳-۶)- نمودار خطای باقیمانده برای پیش‌بینی دانسیته‌ی هیدروکربن‌ها بر حسب مقادیر تجربی دانسیته

برای (الف) سری ارزیابی و (ب) سری تست داخلی

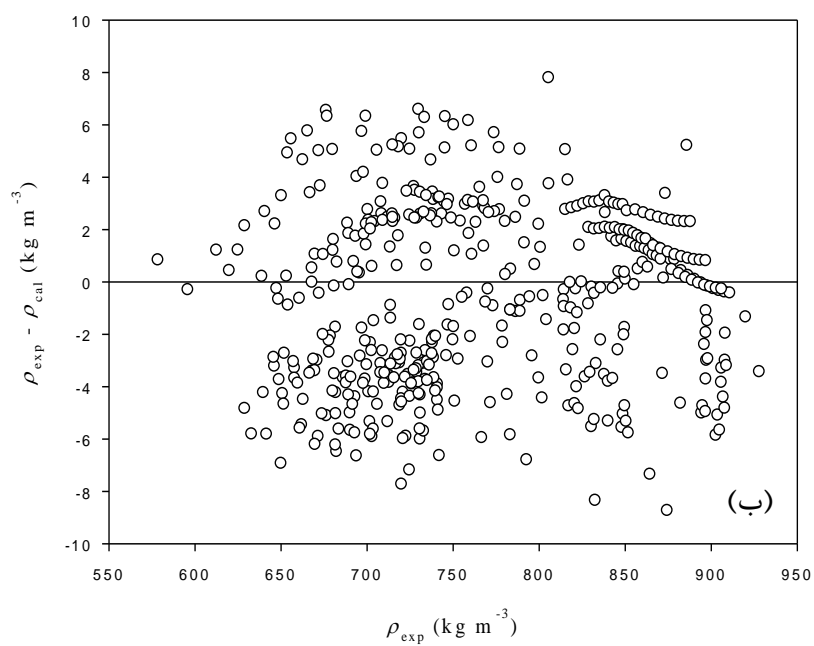
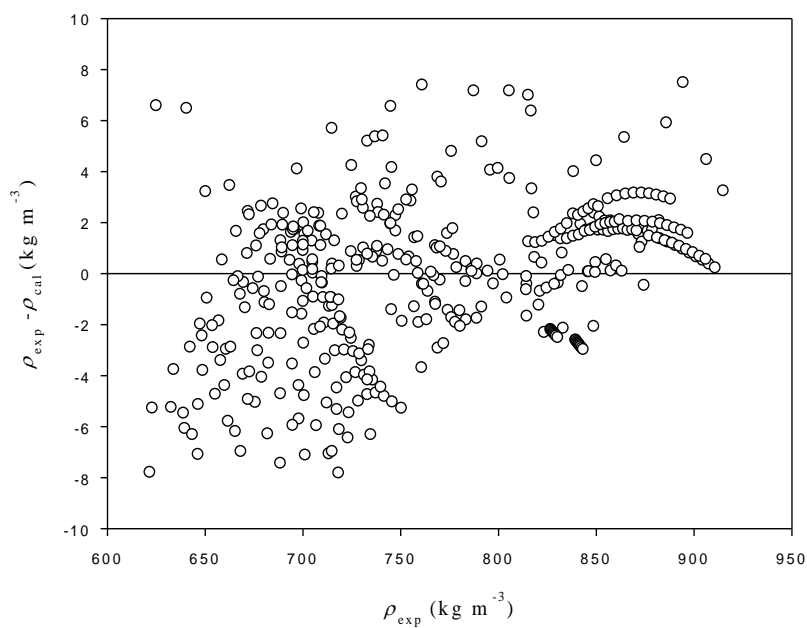
### ۳-۳- ارزیابی قدرت پیش‌بینی مدل‌های غیرخطی بهینه شده برای سری تست خارجی

در این مرحله قدرت پیش‌بینی مدل‌های به دست 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 خطای باقیمانده و پارامترهای آماری در جدول (۳-۱۶) استفاده شده است. در شکل (۳-۷) نمودارهای برگشتی مدل‌های بهینه شده‌ی GCM-ANN و GA-ANN برای سری تست خارجی رسم شده‌اند. نتایج موجود در این شکل‌ها نشان می‌دهد که اگر چه هر دو روش کارایی خوبی دارند اما مدل GCM-ANN برای پیش‌بینی سری تست خارجی از کارایی بالاتر و بهتری برخوردار است. این موضوع در شکل (۳-۸) 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 کمتر از مقدار خطای باقیمانده در مدل GA-ANN بهینه شده است.



شکل (۷-۳) - نمودارهای برگشتی مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری تست

خارجی برای مدل‌های بهینه شده‌ی (الف) GCM-ANN و (ب) GA-ANN



شکل (۳-۸) - نمودارهای خطای باقیمانده مقادیر پیش‌بینی شده دانسیته بر حسب مقادیر تجربی برای سری

تست خارجی برای مدل‌های بهینه شده‌ی (الف) GCM-ANN و (ب) GA-ANN

### ۳-۴- ارزیابی پارامترهای آماری

پارامترهای آماری مقادیر تعیین‌کننده‌ای برای ارزیابی مدل‌های بهینه شده هستند. این مقادیر برای ارزیابی مدل‌های بهینه شده با روش‌های غیرخطی GCM-ANN و GA-ANN محاسبه شده که نتایج این محاسبات در جدول (۳-۱۶) نشان داده شده است. اگر به اعداد جدول برای سری تست در دو مدل بهینه شده دقت کنید می‌بینید مقدار  $R^2$  برای مدل GCM-ANN بیشتر از روش GA-ANN است همچنین سایر پارامترهای آماری روش GCM-ANN مثل MSE کمتر از روش‌های GA-ANN است. با توجه به این مقادیر می‌توان نتیجه گرفت روش سهم‌گروه برای پیش‌بینی دانسیته‌ی سیالات خالص روش کارآمدتری است. از طرفی روش سهم‌گروه یک روش ساده و قابل دسترس است و نیاز به نرم افزارهای متعدد برای محاسبه توصیف‌کننده‌ها ندارد.

جدول (۳-۱۶) - پارامترهای آماری مدل‌های بهینه شده‌ی GCM-ANN و GA-ANN برای دو سری ارزیابی و تست

پارامتر آماری	سری داده	GCM-ANN	GA-ANN
R <sup>2</sup>	ارزیابی	۰/۹۹۹۸	۰/۹۹۹۷
	تست داخلی	۰/۹۹۹۸	۰/۹۹۹۷
	تست خارجی	۰/۹۹۸۷	۰/۹۹۸۲
MSE	ارزیابی	۶/۸۲۴	۴۸/۳۲۵
	تست داخلی	۷/۱۹۰	۶۵/۷۲۱
	تست خارجی	۹/۵۵۶۰	۷۰/۱۶۱۲
PRESS	ارزیابی	۲۸۰/۲۵	۲۲۰۹/۹۷
	تست داخلی	۴۳۳/۸۸۴۴	۳۱۸۶/۱۱
	تست خارجی	۳۵۲۵/۰۹	۵۴۶۷/۳۰۲
MAE	ارزیابی	۰/۳۰۶۸۹۳	۱/۱۳۹۴۶
	تست داخلی	۰/۳۵۳۰۷۱	۱/۳۷۰۰۹۶
	تست خارجی	۲/۳۱۵۴۲۷	۲/۸۸۷۷۹۳
MPE	ارزیابی	۰/۰۰۴۲۳۸	۰/۰۰۵۹۱۷
	تست داخلی	۰/۰۱۰۷۳۲	۰/۰۱۶۶۰۹
	تست خارجی	-۰/۰۲۴۷۴	-۰/۰۸۹۶
MAPE	ارزیابی	۰/۰۳۹۵۸۵	۰/۱۶۰۹۴۳
	تست داخلی	۰/۰۳۵۷۳۳	۰/۲۹۳۰۹۸
	تست خارجی	۰/۳۱۰۵۴۲	۰/۳۸۶۳۹۶
D <sub>max</sub>	ارزیابی	۰/۴۲۸۵۷۱	۰/۶۲۵۲۷
	تست داخلی	۰/۵۲۹۶۳۹	۰/۷۹۱۲
	تست خارجی	۱/۲۵۵۸۱۵	۲/۰۷۴۶۸۱



### ۳-۵- بررسی ارتباط توصیف کننده‌های انتخاب شده در مدل GA-ANN با دانسیته‌ی ترکیبات

در این قسمت به طور خلاصه ارتباط بین توصیف کننده‌های تئوری انتخاب شده در مدل GA-ANN و دانسیته سیالات خالص مورد بررسی قرار خواهد گرفت. این توصیف کننده‌ها شامل PW4، AMW، GGI3، SHP2، PW3، X0Av، Ms، R7u.

#### ۳-۵-۱- توصیف کننده‌ی X0Av

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

این اندیس‌ها اولین بار به وسیله راندیک<sup>۱</sup> در سال ۱۹۷۵ ارائه شدند و سپس در سال ۱۹۷۶ توسط هال<sup>۲</sup> اصلاح گردیدند. این توصیف کننده‌ها با رابطه‌ی (۳-۱) قابل محاسبه هستند:

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

که در این رابطه  $\delta_k^v$ ، درجه رأس ظرفیت<sup>۳</sup> اتم kام در گراف مولکولی است و به شکل رابطه‌ی زیر تعریف می‌شود:

$$\delta_k^v = \frac{(Z_K^v - h_k)}{(Z - Z_k^v - 1)} \quad (2-3)$$

---

1- Randic  
2- Hall  
3- Valence vertex degree

که در آن  $Z$  عدد اتمی اتم  $k$ ،  $Z_k^v$  تعداد الکترون‌های لایه‌ی ظرفیت اتم  $k$  و  $h_k$  تعداد اتم‌های هیدروژن متصل به اتم غیر هیدروژنی  $k$  است [۸۳]. در جدول (۳-۱۷) چگونگی ارتباط این توصیف‌کننده با دانسیته ۶ ترکیب مختلف در دما و فشار ثابت به عنوان مثال آورده شده است. بررسی اثر این توصیف‌کننده نشان می‌دهد که با افزایش دانسیته مقدار توصیف‌کننده X0Av کاهش می‌یابد.

جدول (۳-۱۷) - مثال‌هایی از مقدار توصیف‌کننده X0Av در تعدادی از ترکیبات

شماره ترکیب	نام ترکیب	$T$ (K)	$p$ (MPa)	مقدار محاسبه شده توصیف‌کننده X0Av	Density ( $\text{kg m}^{-3}$ )
۱	<i>n</i> -Hexane	۲۹۸/۱۵	۰/۱	۰/۸۰۵	۶۵۵
۲	<i>n</i> -Heptane			۰/۷۹۱	۶۷۹/۶
۳	<i>n</i> -Octane			۰/۷۸۰	۶۹۸/۷
۴	<i>n</i> -Decane			۰/۷۶۶	۷۲۷/۲
۵	<i>n</i> -Dodecane			۰/۷۵۶	۷۴۶
۶	<i>n</i> -Tetradecane			۰/۷۴۹	۷۶۲

### ۳-۵-۲ - توصیف‌کننده‌های Constitutional

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

توصیف‌کننده AMW با رابطه‌ی زیر قابل محاسبه است:

$$AMW = MW / nAT \quad (3-3)$$

که در این رابطه  $MW$  وزن مولکولی و  $nAT$  تعداد اتم‌ها در ترکیب است.

توصیف‌کننده  $Ms$  نیز طبق رابطه‌ی زیر محاسبه می‌شود:

$$Ms = \sum S_i / nSK \quad (4-3)$$

که در آن  $S_i$  حالت الکتروتوپولوژیکی<sup>۱</sup> و  $nSK$  تعداد اتم‌های غیرهیدروژنی در ترکیب است. مقدار  $S_i$  عبارتست از:

$$S_i = I_i + \Delta I_i \quad (5-3)$$

که در این رابطه  $I_i$  مقدار حالت ذاتی اتم<sup>۲</sup> و  $\Delta I_i$  از روابط زیر به دست می‌آیند:

$$I_i = \frac{Z^v - h + 1}{\delta - h} \quad (6-3)$$

$$\Delta I_i = \sum (I_i - I_j) / r_{ij}^2 \quad (7-3)$$

در روابط بالا،  $Z^v$  تعداد الکترون‌های لایه‌ی ظرفیت اتم،  $\delta$  تعداد الکترون‌های اوربیتال سیگما و  $r_{ij}$  فاصله دو اتم  $i$  و  $j$  در گراف مولکولی است [۸۴].

و ارتباط بین این دو توصیف‌کننده با دانسیته برای چند مولکول از سری داده‌ها در دما و فشار معین در جدول (۳-۱۸) گزارش شده است. همانطور که مشاهده می‌کنید توصیف‌کننده  $Ms$  دارای اثر منفی است یعنی با افزایش مقدار این توصیف‌کننده دانسیته کاهش می‌یابد و برعکس توصیف‌کننده  $AMW$  دارای اثر مثبت بوده و با افزایش مقدار این توصیف‌کننده دانسیته نیز افزایش می‌یابد.

---

1- Electrotopological state  
2- Intrinsic state values

جدول (۳-۱۸) - مثال‌هایی از مقدار توصیف‌کننده‌های AMW و Ms در تعدادی از ترکیبات

شماره ترکیب	نام ترکیب	$T$ (K)	$p$ (MPa)	مقدار محاسبه شده توصیف‌کننده Ms	مقدار محاسبه شده توصیف‌کننده AMW	Density ( $\text{kg m}^{-3}$ )
۱	<i>n</i> -Heptane	۳۱۳/۱۵	۰/۱	۱/۶۴	۴/۳۶	۵۵۸/۷
۲	<i>n</i> -Dodecane			۱/۶۱	۴/۴۸	۷۰۲/۰۲
۳	<i>n</i> -Pentadecane			۱/۵۷	۴/۵۲	۷۵۴/۵۲
۴	<i>n</i> -Octadecane			۱/۵۶	۴/۵۵	۷۶۸/۳۸
۵	<i>n</i> -Nonadecane			۱/۵۵	۴/۵۶	۷۷۲/۰۲

### ۳-۵-۳ - توصیف‌کننده‌های GETAWAY<sup>۱</sup>

این توصیف‌کننده‌ها که با توجه به مختصات فضایی اتم‌ها در یک مولکول به راحتی قابل محاسبه‌اند با استفاده از ماتریس قدرت نفوذ<sup>۲</sup> تعریف می‌شوند و بیان‌کننده‌ی ویژگی‌های هندسی و توپولوژیکی مولکول‌ها هستند. ماتریس قدرت نفوذ یا ماتریس تاثیر مولکول<sup>۳</sup> (MIM) با رابطه زیر بیان می‌شود:

$$H = M.(M^T.M)^{-1}.M^T \quad (۸-۳)$$

که  $M$  ماتریس مختصات اتمی،  $T$  به معنای ماتریس ترانهاده 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 هر یک بیانگر اثر یک اتم در ایجاد شکل کلی یک مولکول است. بالتبع اتم‌های سطحی اعداد لوریج بزرگتری نسبت اتم‌های مرکزی دارند. همچنین مولکول‌های کروی اعداد لوریج پایین‌تری نسبت به مولکول‌های کشیده دارند.

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

توصیف‌کننده‌های GETAWAT به دو زیر مجموعه‌ی H و R تقسیم می‌شوند. توصیف‌کننده‌های H-GETAWAY به اندازه و شکل مولکول وابسته بوده به‌گونه‌ای 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.

$$[R]_{ij} = \left[ \frac{\sqrt{h_{ii} \cdot h_{ij}}}{r_{ij}} \right] \quad i \neq j \quad (9-3)$$

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 [۸۳].

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.

جدول (۳-۱۹) چگونگی ارتباط این توصیف‌کننده با دانسیته‌ی شش ترکیب در دما و فشار معین به عنوان مثال آورده شده است. بررسی اثر این توصیف‌کننده نشان می‌دهد که با افزایش مقدار این توصیف‌کننده مقدار دانسیته کاهش می‌یابد.

جدول (۳-۱۹) - مثال‌هایی از مقادیر توصیف‌کننده R7u در تعدادی از ترکیبات

شماره ترکیب	نام ترکیب	$T$ (K)	$p$ (MPa)	مقدار محاسبه‌شده توصیف‌کننده R7u	Density ( $\text{kg m}^{-3}$ )
۱	<i>n</i> -Heptane	۳۱۳/۱۵	۰/۱	۰/۰۳۰	۶۶۸/۷
۲	<i>n</i> -Dodecane			۰/۰۱۶	۷۲۷/۴
۳	<i>n</i> -Tetradecane			۰/۰۱۴	۷۴۹
۴	<i>n</i> -Pentadecane			۰/۰۱۳	۷۵۴/۵
۵	<i>n</i> -Heptadecane			۰/۰۱۲	۷۶۴/۲
۶	<i>n</i> -Nonadecane			۰/۰۱۱	۷۷۲/۴

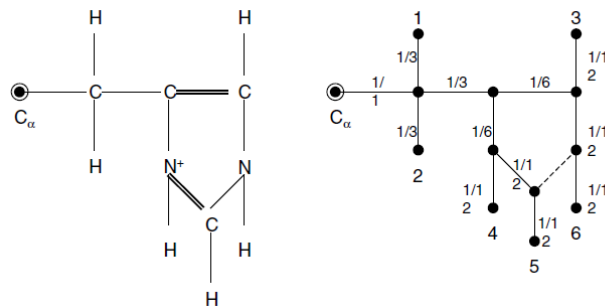
### ۳-۵-۴- توصیف‌کننده‌های توپولوژیکی

از شاخص‌های توپولوژیکی در توسعه کمی روابط ساختار-خاصیت (QSPR) که در آن فعالیت بیولوژیکی یا سایر خواص مولکول‌ها با ساختار شیمیایی آن‌ها ارتباط دارد استفاده می‌گردد. توصیف‌کننده‌های توپولوژیکی از روی گراف‌های مولکولی محاسبه می‌شوند و تنها به نوع اتم، نوع پیوندها و نحوه ارتباط اتم‌ها به یکدیگر وابسته هستند. برای مشخص کردن زنجیره‌ها و شکل زنجیره‌های جانبی از توصیف‌کننده‌های توپولوژیکی بر اساس نظریه گراف که به گراف‌های مولکولی وزن‌دار شده معروف هستند استفاده می‌گردد. هر رأس زنجیره بجز ریشه (اتم کربن  $C_\alpha$ ) وزن‌دار شده و تمام مسیرهای کوتاه وزن‌دار شده بین کربن ریشه ( $C_\alpha$ ) و رأس پایانی نام قرار دارند. وزن هر مسیر به وسیله‌ی مجموع وزن‌های اتمی همه‌ی اتم‌ها تعیین می‌شود. علاوه بر این اگر بیش از یک مسیر کوتاه بین دو رأس وجود داشته باشد مسیری که دارای کم‌ترین مقدار وزن رأس است انتخاب می‌شود.

مقدار احتمالی  $p_i$  به وسیله مسیری که رأس ریشه را به هر رشته‌ی انتهایی نام متصل می‌کند به صورت زیر محاسبه می‌شود:

$$p_i = (\delta'_1 \cdot \dots \cdot \delta'_{i-1})^{-1} \quad (10-3)$$

که  $\delta'$  تعداد پیوندهای هر اتم در مسیر بدون در نظر گرفتن آن‌هایی که در مرحله قبل شمارش شدند.



شکل (۹-۳) - مولکول هیستیدین و گراف مولکولی آن. مقادیر احتمالی محاسبه شده مربوط به هر پیوند از رأس  $C_\alpha$

با استفاده از مقادیر احتمالی محاسبه شده مقدار  $p_{i,p}$  به صورت زیر محاسبه می‌شود:

$$p_{i,p} = p_i \sum_k w_k \quad (11-3)$$

که  $p_{i,p}$  مجموع تمام رأس‌های بین اتم ریشه و رأس انتهایی نام و  $w$  نشان‌دهنده‌ی وزن اتم‌ها در مسیر است.

توصیف‌کننده‌های PW3، PW4 و GGI3 جزو توصیف‌کننده‌های توپولوژیکی هستند. جدول (۲۰-۳) چگونگی ارتباط این توصیف‌کننده‌ها با دانسیته‌ی چند ترکیب در دما و فشار معین نشان می‌دهد. مقادیر موجود در جدول (۲۰-۳) نشان می‌دهد که با افزایش مقدار این توصیف‌کننده‌ها مقدار دانسیته افزایش می‌یابد.

جدول (۳-۲۰) - مثال‌هایی از مقادیر توصیف‌کننده PW3، PW4 و GGI3 در تعدادی از ترکیبات

شماره ترکیب	نام ترکیب	$T$ (K)	$p$ (MPa)	مقدار محاسبه شده توصیف‌کننده PW3	مقدار محاسبه شده توصیف‌کننده PW4	مقدار محاسبه شده توصیف‌کننده GGI3	Density ( $\text{kg m}^{-3}$ )
۱	<i>n</i> -Heptane	۳۱۳/۱۵	۰/۱	۰/۲۱۹	۰/۰۹۷	۰/۱۲۵	۵۵۸/۷
۲	<i>n</i> -Dodecane			۰/۲۳۲	۰/۱۰۹	۰/۱۲۶	۷۰۲/۰۲
۳	<i>n</i> -Pentadecane			۰/۲۳۶	۰/۱۱۲	۰/۱۲۷	۷۵۴/۵۲
۴	<i>n</i> -Octadecane			۰/۲۳۸	۰/۱۱۴	۰/۱۲۹	۷۶۸/۳۸
۵	<i>n</i> -Nonadecane			۰/۲۳۹	۰/۱۱۵	۰/۱۲۹	۷۷۲/۰۲

### ۳-۶- بررسی میزان مشارکت توصیف‌کننده‌های منتخب شبکه‌ی عصبی

#### مصنوعی

میزان مشارکت توصیف‌کننده‌های انتخاب شده توسط شبکه‌ی بهینه‌ی GA-ANN به صورت زیر

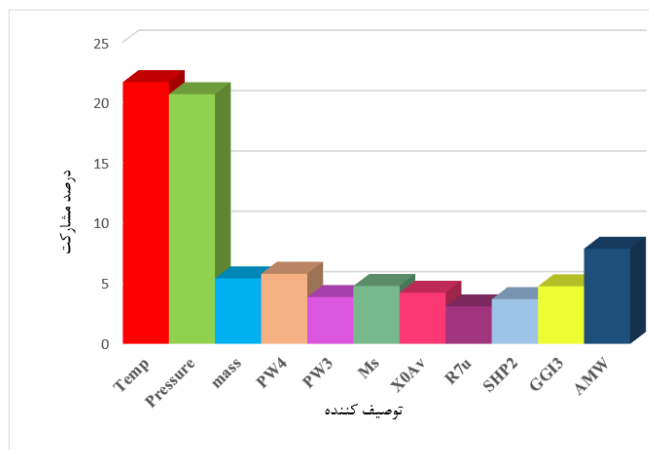
تعیین شد:

- ۱- ابتدا وزن مربوط به توصیف‌کننده‌ی مورد نظر از شبکه بهینه حذف شد.
- ۲- مقدار متغیر وابسته با استفاده از توصیف‌کننده‌های باقی‌مانده برای هر یک از ترکیبات سری ارزیابی پیش‌بینی گردید.
- ۳- میانگین خطای مطلق (MAE) حاصل از ترکیبات سری ارزیابی دز غیاب توصیف‌کننده‌ی مورد نظر محاسبه شد.
- ۴- مراحل ۱ تا ۳ برای دیگر توصیف‌کننده‌ها نیز تکرار گردید.
- ۵- در نهایت درصد مشارکت هر توصیف‌کننده توسط رابطه (۳-۱۲) محاسبه شد:



$$C_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad (12-3)$$

در این رابطه  $C_i$  درصد مشارکت توصیف‌کننده‌ی حذف شده نام،  $N$  تعداد توصیف‌کننده‌ها در مدل و  $\Delta m_i$  میانگین خطای مطلق حاصل از سری ارزیابی در غیاب توصیف‌کننده‌ی نام 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 (۱۰-۳) نشان داده شده است.



شکل (۱۰-۳)-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 جرم کل است که به خوبی وابستگی شدید دانسیته به دما و فشار را نشان می‌دهد.

### ۳-۷- نتیجه گیری

خواص حجمی هیدروکربن‌ها در صنایع شیمیایی وابسته به استخراج و پالایش نفت کاربرد گسترده‌ای دارد. همچنین دانستن دانسیته‌ی این ترکیبات در دما و فشارهای بالا برای مطالعات مهندسی نفت و مخازن زیرزمینی از ضروریات است. از آن جایی که اندازه‌گیری دانسیته به صورت تجربی به صرف هزینه، زمان و پیچیدگی‌های زیادی همراه است لذا استفاده از روش‌های پیش‌بینی از اهمیت زیادی برخوردار است. به همین دلیل در این پایان‌نامه پیش‌بینی دانسیته‌ی طیف وسیعی از هیدروکربن‌ها در گستره‌ی وسیعی از دما و فشار مورد بررسی قرار گرفت. برای این منظور از روش مدل‌سازی شبکه‌ی عصبی مصنوعی به عنوان روشی قدرتمند و کارآمد برای پیش‌بینی دانسیته‌ی هیدروکربن‌ها استفاده شد. توصیف‌کننده‌های مناسب در این تحقیق با استفاده از دو روش به دست آمدند: توصیف‌کننده‌های برمبنای سهم‌گروه‌ها (GCM) و توصیف‌کننده‌های تئوری که با استفاده از الگوریتم ژنتیک (GA) انتخاب شده بودند. با مقایسه‌ی پارامترهای آماری موجود در جدول (۳-۱۶) برای مدل‌های بهینه‌ی GCM-ANN و GA-ANN مشخص شد که، روش مدل‌سازی GCM-ANN نتایج بهتری را نسبت به GA-ANN برای پیش‌بینی دانسیته این ترکیبات ارائه می‌دهد. میانگین مربعات خطا (MSE) و متوسط درصد خطای نسبی (MPE) برای مدل بهینه شده‌ی GCM-ANN به ترتیب ۹/۵۵۶ و ۰/۰۲۴۷۴- و برای مدل بهینه شده‌ی GA-ANN به ترتیب ۷۰/۱۶۱۲ و ۰/۰۸۹۶- بدست آمدند. بررسی نمودارهای خطای باقیمانده نیز نشان داد که گستره‌ی خطاهای باقیمانده سری‌های ارزیابی و تست در مدل حاصل از روش سهم‌گروه به ترتیب  $\pm 3 \text{ Kg m}^{-3}$  و  $\pm 3 \text{ Kg m}^{-3}$  و در مدل حاصل از روش GA به ترتیب  $\pm 4 \text{ Kg m}^{-3}$  و  $\pm 5 \text{ Kg m}^{-3}$  است که این نیز تایید می‌کند روش سهم‌گروه نسبت به الگوریتم ژنتیک عملکرد بهتری دارد. با توجه به این که روش انتخاب توصیف‌کننده‌ها به کمک روش سهم‌گروه بسیار ساده بوده و نیاز به نرم‌افزار خاصی ندارد این موضوع اهمیت مدل بهینه‌ی GCM-ANN را دو چندان می‌کند. همچنین نتایج حاصل از بررسی عملکرد این مدل در پیش‌بینی

دانسیتته‌ی چند هیدروکربن انتخابی توسط مدل‌های بهینه شده GCM-ANN و GA-ANN در این تحقیق با مدل ارائه شده توسط موسوی و همکارانش برای پیش‌بینی دانسیته‌ی هیدروکربن‌ها [۱۸] نشان داد که مدل ارائه شده در این تحقیق متوسط درصد خطای نسبی کمتری داشته و از توانایی بالاتری در پیش‌بینی دانسیته‌ی هیدروکربن‌ها برخوردار است.

### ۳-۸- آینده‌نگری

از روش ارائه شده می‌توان برای پیش‌بینی سایر خواص ترمودینامیکی هیدروکربن‌ها یا خواص انتقالی آن‌ها نیز استفاده نمود.

همچنین می‌توان به جای استفاده از روش شبکه عصبی مصنوعی از روش‌های غیرخطی دیگر مثل ماشین بردار پشتیبان (SVM) برای مدل‌سازی استفاده نمود و نتایج را با این کار مقایسه کرد.

## منابع

- [1] Yen, L. C. and Woods, S. S. (1966) 'A generalized equation for computer calculation of liquid densities', *American Institute of Chemical Engineers*, Vol. 12, pp. 95-97, 1966.
- [2] Chueh, P. L. and Prausnitz, J. M. (1967) 'VLE at high pressure: Calculation of partial molar volumes in nonpolar liquid mixtures', *American Institute of Chemical Engineers*, Vol. 13, pp. 1099-1104, 1967.
- [۳] مورتیمر چ، (۱۹۸۳)، 'کتاب شیمی عمومی ۱'، مترجم عیسی یآوری، ویرایش ششم، ۱۳۷۵  
نشر علوم دانشگاهی تهران.
- [4] Parsafar, G. and Mason, E. A. (1993) 'Linear isotherms for dense fluids: a new regularity', *The Journal of Physical Chemistry*, 97(35), pp. 9048–9053.
- [5] Parsafar, G. and Kalantar, Z. (2003) 'Extension of Linear Isotherm Regularity to Long Chain Alkanes', *Iranian Journal of Chemistry and Chemical Engineering*, 22(2), pp. 1–8.
- [6] Parsafar, G. A. and Kalantar, Z. (2005) 'Extension of linear isotherm regularity to long chain primary, secondary and tertiary alcohols, ketones and 1-carboxylic acids by group contribution method', *Fluid Phase Equilibria*, 234(1–2), pp. 11–21.
- [7] Kalantar, Z., Nikoofard, H. and Javadi, F. (2013) 'Extension of LIR Equation of State to Alkylamines Using Group Contribution Method', *ISRN Physical Chemistry*, 2, pp. 1-9.
- [۸] مرتضوی فر آ، (۱۳۹۰)، پایان نامه کارشناسی ارشد: "پیش بینی خواص حجمی استرها با استفاده از معادله ی حالت MLIR و روش سهم گروه"، دانشکده شیمی، دانشگاه شاهرود.
- [9] Goharshadi, E. K., Morsali, A. and Abbaspour, M. (2005) 'New regularities and an equation of state for liquids', *Fluid Phase Equilibria*, 230(1–2), pp. 170–175.
- [10] Goharshadi, E. K. and Moosavi, F. (2005) 'Prediction of thermodynamic properties of some hydrofluoroether refrigerants using a new equation of state', *Fluid Phase Equilibria*, 238(1), pp. 112–119.

[11] Goharshadi, E. K. and Berenji, A. R. (2006) 'A new equation of state for predicting the thermodynamic properties of liquid alkali metals', *Journal of Nuclear Materials*, 348(1-2), pp. 40-44.

[12] Goharshadi, E. K. and Abareshi, M. (2008) 'Prediction of volumetric and thermodynamic properties of two aromatic-alcohol mixtures using GMA equation of state', *Fluid Phase Equilibria*, 268(1-2), pp. 61-67.

[13] Goharshadi, E. K. and Moosavi, M. (2006) 'Application of a new equation of state to liquid refrigerant mixtures', *Thermochimica Acta*, 447(1), pp. 64-68.

[14] Moosavi, M. (2011) '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 Equilibria*. Elsevier B.V., 310(1-2), pp. 63-73.

[15] Soave, G. (1972) 'Equilibrium constants from a modified Redlich-Kwong equation of state', *Chemical Engineering Science*, 27(6), pp. 1197-1203.

[16] Twu, C. H., Coon, J. E. and Cunningham, J. R. (1995) 'A new generalized alpha function for a cubic equation of state Part 1. Peng-Robinson equation', *Fluid Phase Equilibria*, 105(1), pp. 49-59.

[17] شعبانی س، (۱۳۸۸)، پایان نامه کارشناسی ارشد: "مدل سازی شبکه عصبی موجک با استفاده از روش سهم گروه ها برای پیش بینی دانسیته ی آلکان ها و سیکلو آلکان ها در محدوده ی وسیعی از دما و فشار"، دانشکده شیمی، دانشگاه شاهرود.

[18] Moosavi, M. and Soltani, N. (2013) 'Prediction of hydrocarbon densities using an artificial neural network-group contribution method up to high temperatures and pressures', *Thermochimica Acta*. Elsevier B.V., 556, pp. 89-96.

[19] پایین محلی ح، (۱۳۸۸)، پایان نامه کارشناسی ارشد: "مطالعه اثر بازدارنگی مشتقات 5- هیدروکسی-3-(2H)-پریدارینون بر روی ویروس هیپاتیت نوع C با استفاده از روش های کمومتریکس"، دانشکده شیمی، دانشگاه صنعتی شریف.

[20] Manssnat, D. L., Vandeginste, B. G., Deming, S. N. and Kaufman, L. (1998) 'Chemometrics, A Text Book' Elsevier, Amesterdom.

[۲۱] 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.

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

[۲۳] 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.

[۲۴] کریمی س، (۱۳۸۵)، پایان نامه 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 ی عصبی مصنوعی"، دانشکده شیمی، دانشگاه صنعتی شریف.

[25] [www.iasbs.ac.ir/chemistry/chemometrics/.../8th/qsar\\_introduction](http://www.iasbs.ac.ir/chemistry/chemometrics/.../8th/qsar_introduction).

[26] Arab Chamjangali, M. (2009) 'Modelling of cytotoxicity data (CC50) of anti-HIV 1-[5-chlorophenyl) sulfonyl]-1H-pyrrole derivatives using calculated molecular descriptors and levenberg-marquardt artificial neural network', *Chemical Biology and Drug Design*, 73(4), pp. 456–465.

[27] Holland, J. H. (1975) 'Adoptation in natural and artificial systems' University of mishigan press.

[28] Zhuo, L., Zheng, J., Li, X., Wang, F., Ai, B., Qian, J. (2008) 'A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using

support vector machine', *Proceedings of SPIE - The International Society for Optical Engineering*, 7147, p. 71471J–71471J–9.

[29] Melanie M. (1999), "An introduction to genetic algorithms", Vol. 3, Cambridge, Massachusetts London, England, pp. 1-144.

[۳۰] باهر ا، (۱۳۸۸)، پایان‌نامه دکتری: "پیش‌بینی و مطالعه QSPR بازداري ترکیبات آلي، دارویی و آلاینده‌ها با استفاده از رگرسیون بردارهای پشتیبان"، دانشکده شیمی، دانشگاه مازندران.

[۳۱] کلانتر ز، (۱۳۸۵)، پایان‌نامه دکتری: "پیش‌بینی خواص ترمودینامیکی و انتقال سیالات آلي چگال با استفاده از روش سهم‌گروه‌ها و شبکه‌ی عصبی موجک"، دانشکده شیمی، دانشگاه صنعتی اصفهان.

[32] Pan, Y., Jiang, J. and Wang, Z. (2007) 'Quantitative structure-property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network', *Journal of Hazardous Materials*, 147(1–2), pp. 424–430.

[33] Valderrama, J. O., Reátegui, A. and Rojas, R. E. (2009) 'Density of Ionic Liquids Using Group Contribution and Artificial Neural Networks', *Industrial & Engineering Chemistry Research*, 48(6), pp. 3254–3259.

[34] Gharagheizi, F., Eslamimanesh, A., Mohammadi, H. A. (2011) 'Determination of critical properties and acentric factors of pure compounds using the artificial neural network group contribution algorithm', *Journal of Chemical and Engineering Data*, 56(5), pp. 2460–2476.

[35] Matsuda, H., Yamamoto, H., Kurihara, K., Tochigi, K. (2007) 'Computer-aided reverse design for ionic liquids by QSPR using descriptors of group contribution type for ionic conductivities and viscosities', *Fluid Phase Equilibria*, 261(1–2), pp. 434–443.

[۳۶] حسین کوچک م، (۱۳۸۴)، پایان‌نامه کارشناسی ارشد، "بهبود متدولوژی سطح پاسخ با استفاده از شبکه‌های عصبی و الگوریتم"، دانشکده مهندسی صنایع، دانشگاه صنعتی شریف.



[۳۷] طاهر زاده 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.

[38] Hebb, D. O. (1949) 'the organization of behavior', Weily, New York.

[39] Hegan, M. T., Demuth B., Beale M. (1995) 'Neural network design', PWS.

[40] HOPFIELD, J. (1982) 'Neural networks and physical systems with emergent collective computational abilities' *Proc. Natl. Acad. Sci. USA*, 79, pp. 2554-2558.

[41] Zupan, J. and Gasteiger J. (1999) 'Neural Network for Chemistry and Drug Design' Weily-VCH Publishers, Weinheim(Germany).

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

[۴۳] 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.

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

[45] Iglesias-silva, G. A, Estrada-baltazar, A. and Hall, K. R. (1999) 'Experimental Liquid Viscosity of Pentane + Octane + Decane Mixtures from 298 . 15 to 373 . 15 K up to 25 MPa', *J. Chem. Eng. Data*, 9568(97), pp. 1304–1309.

[46] Dymond, J. H., Young, K. J. and Isdale, J. D. (1979) 'behaviour for n-hexane + n-hexadecane in the range 298 to 373 K and 0.1 to 500 MPa', *J. Chem. Thermodynamic*, 11(2), pp. 887–895.

[47] Doolittle, A. K. (1964) 'Specific Volumes of n -Alkanes.', *Journal of Chemical & Engineering Data*, 9(2), pp. 275–279.

[48] Caudwell, D. R. *et al.* (2009) 'Viscosity and Density of Five Hydrocarbon Liquids

at Pressures up to 200 MPa and Temperatures up to 473 K Viscosity and Density of Five Hydrocarbon Liquids at Pressures up to 200 MPa and Temperatures up to 473 K ', *Journal of Chemical & Engineering Data*, 54, pp. 359–366.

[49] Caudwell, D. R. *et al.* (2004) 'The viscosity and density of n-dodecane and n-octadecane at pressures up to 200 MPa and temperatures up to 473 K', *International Journal of Thermophysics*, 25(5), pp. 1339–1352.

[50] Khasanshin, T. S. and Shchemelev, A. P. (2002) 'The thermodynamic properties of n-tetradecane in liquid state', *High Temperature*, 40(2), pp. 207–211.

[51] Daridon, J. L., Carrier, H. and Lagourette, B. (2002) 'Pressure Dependence of the Thermophysical Properties of n-Pentadecane and n-Heptadecane', *International Journal of Thermophysics*, 23(3), pp. 697–708.

[52] Dutour, S., Daridon, J. L. and Lagourette, B. (2000) 'Pressure and temperature dependence of the speed of sound and related properties in normal octadecane and nonadecane', *International Journal of Thermophysics*, 21(1), pp. 173–184.

[53] Dutour, S., Lagourette, B. and Daridon, J. L. (2002) 'High-pressure speed of sound, density and compressibility of heavy normal paraffins: C<sub>28</sub>H<sub>58</sub> and C<sub>36</sub>H<sub>74</sub>', *The Journal of Chemical Thermodynamics*, 34, pp. 475–484.

[54] <https://www.nist.gov/chemistry> webbook.

[55] Dawson, P. P., Silberberg, I. H. and McKetta, J. J. (1973) 'Volumetric Behavior, Vapor Pressures, and Critical Properties of Neopentane', *Journal of Chemical and Engineering Data*, 18(1), pp. 7–15.

[56] Walter, T., Bardelmeier, U. and Würflinger, A. (1992). 'Ultrasonic Studies of Rotational Isomerism and P, V, T Data of 2-Methylbutane at Elevated Pressures'. *Berichte der Bunsengesellschaft für physikalische Chemie*, 96(5), pp. 717-721.

[57] Kelso, E. A., and Felsing, W. A. (1942). 'PVT Relations and Derived Quantities for Hexanes'. *Industrial & Engineering Chemistry*, 34(2), pp. 161-163.

- [58] Plantier, F. and Daridon, J. L. (2005) ‘Speed of Sound of 2-Methylpentane, 2, 3-Dimethylpentane, and 2, 2, 4-Trimethylpentane from (293.15 to 373.15) K and up to 150 MPa’. *Journal of Chemical & Engineering Data*, 50(6), pp. 2077-2081.
- [59] Takagi, T., Sakura, T. and Guedes, H. J. R. (2002) ‘Speed of sound in liquid cyclic alkanes at temperatures between (283 and 343) K and pressures up to 20 MPa’, *Journal of Chemical Thermodynamics*, 34(12), pp. 1943–1957.
- [60] Voss, S. F. and Sloan, E. D. (1989) ‘Thermal conductivity and heat capacity of synthetic fuel components’. *International Journal of Thermophysics*, 10(5), pp. 1029-1040.
- [61] Tapriyal, D., Enick, R., McHugh, M., Gamwo, I. and Morreale, B. (2012) ‘High Temperature, high pressure equation of state density correlations and viscosity correlations (No. NETL-TRS-1-2012)’. *National Energy Technology Laboratory (NETL)*, (July), pp. 1–98.
- [62] Baylaucq, A., Zeberg-Mikkelsen, C. K., Daugé, P. and Boned, C. (2002) ‘Dynamic viscosity and density of heptylbenzene and heptylcyclohexane up to 100 MPa’, *Journal of Chemical and Engineering Data*, 47(4), pp. 997–1002.
- [63] Dutour, S., Carrier, H. and Daridon, J. L. (2003) ‘Compressibilities of liquid pentadecylcyclohexane and nonadecylcyclohexane from high pressure speed of sound and density measurements’, *Journal of Chemical Thermodynamics*, 35(10), pp. 1613–1622.
- [64] Jahangiri, M., Jacobsen, R. T., Stewart, R. B. and McCarty, R. D. (1986) ‘Thermodynamic properties of ethylene from the freezeline to 450K at pressures to 260 MPa’, *Journal of Physical and Chemical Reference Data*, 15(2), pp. 593–734.
- [65] Ihmels, E. C., Fischer, K. and Gmehling, J. (2005) ‘Thermodynamic properties of the butenes: Part I. Experimental densities, vapor pressures, and critical points’, *Fluid Phase Equilibria*, 228–229, pp. 155–171.
- [66] Vargaftik, N. B. (1975) ‘Handbook of Physical Properties of Liquids and Gases’. Second Edition, Hemisphere Publication Corporation, New York, pp 323.

- [67] Sagdeev, D. I., Fomina, M. G., Mukhamedzyanov, G. K. and Abdulagatov, I.M. (2014) 'Experimental study and correlation models of the density and viscosity of 1-hexene and 1-heptene at temperatures from (298 to 473) K and pressures up to 245 MPa', *Journal of Chemical and Engineering Data*, 59(4), pp. 1105–1119.
- [68] Dymond, J. H., Malhotra, R., Isdale, J. D. and Glen, N. F. (1988) '(p,  $\rho$ , T) of n-heptane, toluene, and oct-1-ene in the range 298 to 373 K and 0.1 to 400 MPa and representation by the Tait equation', *The Journal of Chemical Thermodynamics*, 20(5), pp. 603–614.
- [69] Khasanshin, T. S. and Shchemelev, A. P. (2002) 'The thermodynamic properties of n-tetradecane in liquid state', *High Temperature*, 40(2), pp. 207–211.
- [70] Khasanshin, T. S., Poddubskij, O. G., Shchamialiou, A.P. and Samuilov, V.S. (2006) 'The thermodynamic properties of 1-alkenes in the liquid state: 1-Hexadecene', *Fluid Phase Equilibria*, 245(1), pp. 26–31.
- [71] Letcher, T. M. and Marsicano, F. (1974) 'Vapour pressures and densities of some unsaturated C6 acyclic and cyclic hydrocarbons between 300 and 320 K', *The Journal of Chemical Thermodynamics*, 6(5), pp. 509–514.
- [72] Deul, R., Rosenzweig, S. and Franck, E. U. (1991) 'The Dielectric Constant and Density of Benzene to 400 °C and 3000 bar', *Berichte der Bunsengesellschaft für physikalische Chemie*, 95(4), pp. 515–519.
- [73] Segovia, J. J., Fandiño, O., López, E. R., Lugo, L., Martín, M. C. and Fernández, J. (2009) 'Automated densimetric system: Measurements and uncertainties for compressed fluids', *Journal of Chemical Thermodynamics*. Elsevier Ltd, 41(5), pp. 632–638.
- [74] Garg, S. K., Banipal, T. S. and Ahluwalia, J. C. (1993) 'Heat capacities and densities of liquid o-xylene, m-xylene, p-xylene, and ethylbenzene, at temperatures from 318.15 K to 373.15 K and at pressures up to 10 MPa', *The Journal of Chemical Thermodynamics*, 25(1), pp. 57-62.

- [75] Akhundov, T. S., Imanov, Sh.Yu. (1969) ‘Experimental observation of specific volumes of p-xylene at temperatures 50-400°C and pressures up to 500 bar’, *Izv.Vyssh.Uchebn.Zaved.,Nef. Gaz.*, 12, pp. 66-68.
- [76] Zhou, H., Lagourette, B., Alliez, J., Xans, P. and Montel, F. (1990) ‘Extension of the application of the Simha equation of state to the calculation of the density of n-alkanes-benzene and alkylbenzenes mixtures’, *Fluid Phase Equilibria*, 59(3), pp. 309–328.
- [77] Chyliński, K. and Gregorowicz, J. (1991) ‘Liquid-phase PVT data of alkylbenzenes’, *Fluid Phase Equilibria*, 64(C), pp. 237–249.
- [78] Chang, J. S., Lee, M. J. and Lin, H. (1996) ‘Densities of m-Xylene+ Quinoline and m-Xylene+ Tetralin from (333 to 413) K and up to 30 MPa’, *Journal of Chemical and Engineering Data*, 41(5), pp. 1117–1120.
- [79] Milhet, M., Baylaucq, A. and Boned, C. (2005) ‘Volumetric properties of 1-phenyldecane and 1-phenylundecane at pressures to 65 MPa and temperature between 293.15 and 353.15 K’, *Journal of Chemical & Engineering Data*, 50(4), pp. 1430-1433.
- [80] Dutour, S., Carrier, H., Daridon, J. L., Lagourette, B. and Gao, G.H. (2002) ‘Speed of sound, density, and compressibility of alkylbenzenes as a function of pressure and temperature: tridecylbenzene and pentadecylbenzene’, *Journal of Chemical & Engineering Data*, 47(6), pp. 1532-1536.
- [81] Dutour, S., Carrier, H., Lagourette, B., Gao, G. H. and Daridon, J.L. (2004) ‘Speed of sound, density, and compressibility of alkyl-benzenes as a function of pressure and temperature: Heptadecylbenzene and octadecylbenzene’, *Journal of Chemical and Engineering Data*, 49(4), pp. 983–987.
- [82] Russell, F. R. and Hottel, H. C. (1938) ‘Compressibility of Liquid Naphthalene’, *Industrial and Engineering Chemistry*, 30(3), pp. 343–345.
- [83] Todeschini, R., and Consonni, V. (2009) ‘Molecular descriptors for chemoinformatics’. John Wiley & Sons
- [84] Hall, L. H. and Mohney, B. (1991) ‘The Electrotopological State : An Atom Index for QSAR ’, 51, pp. 43–51



## Abstract

In the present work, a QSPR study was performed to predict the density of saturated and unsaturated aliphatic hydrocarbons containing linear and branched alkanes, cycloalkanes, alkenes and cycloalkenes as well as aromatic hydrocarbons over a wide range of temperatures and pressures. The first selected descriptors were temperature ( $T$ ), pressure ( $p$ ) and molecular mass ( $M$ ) because the density is related to them. The other descriptors have been obtained from chemical structure of compounds using two different approaches. In the first approach, the best descriptors were selected based on group contribution method. After analyzing the chemical structures of 79 studied hydrocarbons, There were found 15 useful functional groups. In the second method, a large number of molecular descriptors (3224 descriptors) were calculated using Dragon software from chemical structure of compounds. Then, a subset of calculated descriptors was selected from 22 classes of Dragon descriptors with a genetic algorithm based on partial least square (GA-PLS) method as a feature selection technique. Only 11 descriptors were obtained by genetic algorithm (GA) as the most feasible descriptors. The selected descriptors with two feature selection methods and 3 variable ( $T$ ,  $P$  and  $M$ ) were used as input nodes for constructing different artificial neural networks. 4 feed-forward artificial neural networks were designed by combination of Levenberg-Marquardt (LM) and Bayesian regularized (BR) algorithms with two transfer function: logarithm-sigmoid and tangent-sigmoid. Different parameters for designed networks such as number of neurons in input and hidden layers and the number of iteration (epoch) were optimized. After optimization of parameters for generated artificial neural networks using two types of descriptors, the performance of the models was investigated by the external test set. The mean square error ( $MSE$ ) and determination coefficient ( $R^2$ ) for the external test set using optimized GCM-ANN model are 9.556 and 0.9987, and using optimized GA-ANN model are 70.1612 and 0.9982 respectively. The result showed that the Levenberg-Marquardt artificial neural network with logarithm-sigmoid transfer function and GCM based descriptors, may be simulated the relationship between the structural descriptors and density of the desired molecules accurately.

**Keywords:** quantitative structure–property relationship (QSPR), artificial neural network (ANN), hydrocarbons, density, group contribution method (GCM), Genetic algorithm (GA).



Shahrood University of Tchnology

Faculty of Chemistry

M.Sc. Thesis in Physical Chemistry

**Quantitative structure-property relationship study of density  
for hydrocarbons over a wide range of temperature and  
pressure**

Mansoureh Motevalli

: Supervisors

**Dr. Zahra Kalantar**

**Dr. Mohsen Sargolzaei**

January 2018