



دانشگاه صنعتی شاهرود

دانشکده شیمی

گرایش شیمی تجزیه

پایان نامه کارشناسی ارشد

مدل سازی QSPR اندیس بازداری هیدروکربن های  
آروماتیک چندحلقه ای با روش رگرسیون خطی چند گانه  
و شبکه عصبی مصنوعی

فاطمه یامی

استاد راهنما:

دکتر ناصر گودرزی

استاد مشاور:

دکتر منصور عرب چم جنگلی

تیر ماه ۱۳۸۸

## فهرست مطالب

صفحه	عنوان
۱	چکیده
	<b>فصل اول</b>
	مقدمه
۲	۱-۱- معرفی هیدروکربنهای آروماتیک چندحلقه ای
۳	۱-۱-۲- منابع PAHs
۳	۱-۱-۲-۱- منابع خانگی
۳	۱-۱-۲-۲- منابع متحرک
۴	۱-۱-۲-۳- منابع کشاورزی
۴	۱-۱-۲-۴- منابع طبیعی
۴	۱-۱-۲-۵- منابع صنعتی
۵	۱-۱-۳- اثرات مخرب ترکیبات PAHs بر محیط زنده
۵	۱-۱-۳-۱- تنفس هوای آلوده
۶	۱-۱-۳-۲- خوردن و آشامیدن
۶	۱-۱-۳-۳- لمس کردن و تماس پوستی
۶	۱-۱-۴- جداسازی و آنالیز کمی و کیفی PAHs
۷	۱-۱-۴-۱- کروماتوگرافی مایع با عملکرد بالا
۷	۱-۱-۴-۲- کروماتوگرافی گازی با آشکارساز طیف سنج جرمی
۷	۱-۱-۴-۳- اساس کروماتوگرافی گازی
۸	۱-۱-۴-۲-۲- آشکارساز
۹	۱-۱-۴-۳-۱- اندیس بازداری

۱۰	۱-۱-۴-۴-مشکلات موجود در اندازه گیری اندیس بازداری ترکیبات PAHs
۱۲	۱-۲-هدف تحقیق
۱۳	۱-۳-مراحل تحقیق
۱۴	۱-۴-مروری بر کارهای انجام شده

## فصل دوم

### کمومتریکس

۱۶	۲-۱-مقدمه
۱۷	۲-۲-ارتباط کمی ساختار-خاصیت
۱۸	۲-۲-۱-جمع آوری سری داده ها
۱۹	۲-۲-۲-بهبود کردن ساختار هندسی ترکیبات
۲۰	۲-۲-۳-محاسبه توصیف کننده ها
۲۱	۲-۲-۳-۱-توصیف کننده های صفر بعدی
۲۱	۲-۲-۳-۲-توصیف کننده های یک بعدی
۲۲	۲-۲-۳-۱-گروه های عاملی
۲۲	۲-۲-۳-۲-قطعات اتم مرکزی
۲۲	۲-۲-۳-۳-توصیف کننده های تجربی
۲۳	۲-۲-۳-۴-خصوصیات مولکولی
۲۳	۲-۳-۳-۲-توصیف کننده های دوبعدی
۲۳	۲-۳-۳-۱-توصیف کننده های توپولوژی
۲۴	۲-۳-۳-۲-توصیف کننده های BCUT
۲۴	۲-۳-۴-توصیف کننده های سه بعدی
۲۵	۲-۳-۴-۱-اندیس آروماتیکی

۲۵	.....۲-۲-۳-۴-۲-خصوصیات مولکولی راندیک
۲۶	.....۲-۲-۳-۴-۳-توصیف کننده های هندسی
۲۷	.....۲-۲-۳-۴-۴-RDF توصیف کننده های
۲۷	.....۲-۲-۳-۴-۵-3DMORSE توصیف کننده های
۲۷	.....۲-۲-۳-۴-۶-WHIM توصیف کننده های
۲۸	.....۲-۲-۳-۴-۷-GETAWAY توصیف کننده های
۲۹	.....۲-۲-۴-انتخاب بهترین توصیف کننده ها
۲۹	.....۲-۲-۵-مدل سازی
۳۰	.....۲-۲-۶-رگرسیون خطی چند گانه
۳۲	.....۲-۲-۶-۱-آنالیز مدل های آماری و انتخاب مدل
۳۴	.....۲-۲-۷-شبکه های عصبی مصنوعی
۳۴	.....۲-۲-۷-۱-مقدمه
۳۵	.....۲-۲-۷-۲-ماهیت شبکه عصبی مصنوعی
۳۶	.....۲-۲-۷-۳-دلیل استفاده از شبکه عصبی
۳۶	.....۲-۲-۷-۴-نورون بیولوژیکی
۴۰	.....۲-۲-۷-۵-تابع انتقال
۴۳	.....۲-۲-۷-۶-ساختار شبکه ی عصبی
۴۳	.....۲-۲-۷-۶-۱-شبکه ی تک لایه
۴۴	.....۲-۲-۷-۶-۲-شبکه ی چند لایه
۴۵	.....۲-۲-۷-۷-انواع شبکه
۴۵	.....۲-۲-۷-۷-۱-شبکه های پیش خور
۴۶	.....۲-۲-۷-۷-۲-شبکه های برگشتی
۴۷	.....۲-۲-۷-۸-طراحی شبکه عصبی

۴۷	۲-۷-۸-۱- ایجاد شبکه
۴۷	۲-۷-۸-۲- تعیین مقدار اولیه ی وزن و بایاس
۴۷	۲-۷-۸-۳- شبیه سازی
۴۸	۲-۷-۹- آموزش شبکه
۴۸	۲-۷-۱۰- قواعد یادگیری
۴۸	۲-۷-۱۰-۱- الگوریتم پس انتشار
۵۲	۲-۷-۱۰-۲- الگوریتم لونبرگ مارکواردت

## فصل سوم

### مدل سازی اندیس بازدارى ترکیبات PAHs

۵۵	۳-۱- مقدمه
۵۵	۳-۲- انتخاب سری داده ها
۵۶	۳-۲-۱- نتایج
۶۳	۳-۳- محاسبه ی توصیف کننده ها
۶۵	۳-۴- انتخاب توصیف کننده ی مناسب
۶۵	۳-۵- آنالیز رگرسیون خطی چند گانه
۷۱	۳-۶- مدل سازی با شبکه ی عصبی مصنوعی
۷۱	۳-۶-۱- مقدمه
۷۲	۳-۶-۲- انتخاب ورودی شبکه
۷۳	۳-۶-۳- انتخاب مقادیر اولیه ی وزن و بایاس
۷۳	۳-۶-۴- بهینه سازی تابع انتقال و گر لایه ی پنهان
۷۶	۳-۶-۵- بهینه سازی تعداد چرخه های آموزش

## فصل چهارم

### بحث و نتیجه گیری

- ۱-۴- مقایسه روش های رگرسیون خطی و شبکه عصبی با استفاده از سری ارزیابی و تست ..... ۷۸
- ۲-۴- ارزیابی مدل خطی و شبکه عصبی با روش رد مرحله ای تک تک ..... ۸۵
- ۳-۴- بررسی ارتباط توصیف کننده های انتخاب شده در مدل واندیس بازداري ترکیبات ..... ۱۰۰
- آینده نگری ..... ۱۰۴
- فهرست منابع لاتین ..... ۱۰۵
- فهرست منابع فارسی ..... ۱۱۰
- چکیده لاتین ..... ۱۱۱

## فهرست اشکال

صفحه	عنوان
۳۶	شکل ۲-۱- شبکه ی عصبی مصنوعی
۳۸	شکل ۲-۲- نرون عصبی
۳۹	شکل ۲-۳- مقایسه ی نرون عصبی و مصنوعی
۴۰	شکل ۲-۴- توابع انتقال
۴۳	شکل ۲-۵- شبکه ی تک لایه
۴۵	شکل ۲-۶- شبکه ی چند لایه
۴۶	شکل ۲-۷- شبکه ی پیش خور
۴۶	شکل ۲-۸- شبکه ی برگشتی
۵۱	شکل ۲-۹- مراحل تصحیح وزن در پس انتشار خطا

### فصل سوم

۷۴	شکل ۳-۱- تغییرات گره و خطای تابع خطی
۷۵	شکل ۳-۲- تغییرات گره و خطای تابع تانژانتی
۷۵	شکل ۳-۳- تغییرات گره و خطای تابع لگاریتمی
۷۶	شکل ۳-۴- تغییرات خطای چرخه های آموزش

### فصل چهارم

۸۳	شکل ۴-۱- نمودار مقادیر محاسبه شده RI سری پیش بینی با MLR
۸۳	شکل ۴-۲- نمودار مقادیر محاسبه شده RI سری پیش بینی با ANN

۸۴	شکل ۳-۴-مقادیر محاسبه شده اندیس بازداری سری تست با MLR
۸۴	شکل ۴-۴- مقادیر محاسبه شده اندیس بازداری سری تست با ANN
۹۵	شکل ۵-۴-مقادیر پیش بینی شده سری داده ها با MLR
۹۵	شکل ۶-۴- مقادیر پیش بینی شده سری داده ها با ANN
۹۶	شکل ۷-۴-نمودار باقی مانده ها با MLR سری پیش بینی و تست
۹۶	شکل ۸-۴- نمودار باقی مانده ها با ANN سری پیش بینی و تست
۹۷	شکل ۹-۴-نمودار باقی مانده ها برای سری داده ها با MLR
۹۷	شکل ۱۰-۴- نمودار باقی مانده ها برای سری داده ها با ANN
۱۰۱	شکل ۱۱-۴-شکل ۱۱-۴-شکل چند آروماتیک چند حلقه ای
۱۰۳	شکل ۱۲-۴-شکل چند آروماتیک چند حلقه ای



## فهرست جداول

صفحه	عنوان
۵۷	جدول ۱-۳- شکل و اندیس بازداري سری داده ها
۶۴	جدول ۲-۳- نام توصيف کننده ها
۶۶	جدول ۳-۳- آماره های مدل های خطی
۶۶	جدول ۴-۳- ضرایب رگرسیون و اثر متوسط توصيف کننده ها
۶۷	جدول ۵-۳- ماتریس همبستگی میان توصيف کننده ها
۶۸	جدول ۶-۳- مقادير توصيف کننده های وارد شده در مدل خطی
۷۴	جدول ۷-۳- مقادير گره و خطای توابع انتقال
۷۷	جدول ۸-۳- شرایط بهینه ی شبکه

### فصل چهارم

۷۹	جدول ۱-۴- مقادير محاسبه شده اندیس بازداري سری پیش بینی با MLR
۸۰	جدول ۲-۴- مقادير محاسبه شده اندیس بازداري سری پیش بینی با ANN
۸۱	جدول ۳-۴- مقادير پیش بینی شده RI سری تست با MLR
۸۲	جدول ۴-۴- مقادير پیش بینی شده RI سری تست با ANN
۸۶	جدول ۵-۴- مقادير پیش بینی شده RI توسط MLR با روش رد مرحله ای تک تک
۹۰	جدول ۶-۴- مقادير پیش بینی شده RI توسط ANN با روش رد مرحله ای تک تک
۹۸	جدول ۷-۴- مقایسه ی پارامتر های آماری بدست آمده از مدل MLR و ANN
۱۰۱	جدول ۸-۴- مقادير عددی توصيف کننده های SNar و Mor06m
۱۰۳	جدول ۹-۴- مقادير عددی توصيف کننده های RPCG و GATS1e

## چکیده

در این تحقیق از مطالعات کمی ساختار-خاصیت جهت مدل سازی و پیش بینی اندیس بازداری یک سری از هیدروکربن های آروماتیک چند حلقه ای (PAHs) استفاده شده است. این ترکیبات به دلیل خاصیت سمی و سرطان زا بودنشان، اثرات مضر بر روی سلامت انسانها و محیط پیرامونشان دارند. به همین دلیل از دیر باز مورد توجه بسیاری از دانش مندان بوده اند. منابع مختلف انسانی و طبیعی از جمله احتراق ناقص سوخت های فسیلی و مواد آلی منجر به تولید دسته ی وسیعی از این هیدروکربن ها می شوند. اندیس بازداری (RI) پارامتری است که برای جداسازی، شناسایی کمی و کیفی بسیاری از هیدروکربن های آروماتیک چند حلقه ای استفاده می شود. این خاصیت با روش هایی چون کروماتوگرافی مایع با عملکرد بالا و کروماتوگرافی گازی با آشکار ساز طیف سنج جرمی قابل اندازه گیری می باشد. اندیس بازداری بازتابی از برهمکنش مولکول با فاز ساکن ستون کروماتوگرافی است. این برهمکنش وابسته به ویژه گی های ساختاری مولکول می باشد. مطالعات کمی ساختار-خاصیت (QSPR) به خوبی ارتباط بین ویژه گی های ساختاری و اندیس بازداری هیدروکربن های آروماتیک چند حلقه ای را توجیه کرده و این پارامتر را پیش بینی می کند. به همین منظور ابتدا ساختار ترکیبات با استفاده از نرم افزار HyperChem رسم شد. سپس توصیف کننده های ساختاری با استفاده از نرم افزار Dragon محاسبه شدند. سرانجام با استفاده از مقادیر عددی توصیف کننده ها و اندیس بازداری ۸۳ هیدروکربن آروماتیک چند حلقه ای، ابتدا با رگرسیون خطی (MLR) مدل خطی ساخته شد. سپس با استفاده از شبکه ی عصبی مصنوعی (ANN) اندیس بازداری برای ترکیبات PAH پیش بینی شد. نتایج، برتری مدل غیر خطی را در پیش بینی اندیس بازداری نشان دادند.

کلمات کلیدی : Polycyclic Aromatic Hydrocarbons (PAHs), Artificial Neural Network (ANN), Multi Linear Regression (MLR), Retention Index (RI), Quantitative Structure Property Relationship (QSPR).

فصل اول

مقدمه

## ۱-۱- معرفی هیدروکربن های آروماتیک چند حلقه ای

هیدروکربن های آروماتیک چند حلقه ای گروهی از ترکیبات هستند که از دو یا چند حلقه آروماتیک تشکیل شده اند. این ترکیبات خواص فیزیکی و شیمیایی مشابهی دارند. از مهمترین آلوده کننده های محیطی بوده و دارای منابع مختلف طبیعی و بشری هستند. ترکیبات PAHs سمی و سرطان زا هستند و به شکل های مختلف بوده، از روش های مختلفی از جمله تغذیه و تنفس وارد بدن انسان و جانوران می شوند و بر روی سلامت آنها اثر می گذارند. این هیدروکربن ها اکثراً ایزومر بوده و اثرات آنها به عوامل زیر بستگی دارد [۱]:

۱- شرایط محیطی مثل دما، گردش و تجمع هوا، اجزاء خاک و مواد موجود در آب و خاک و رسوبات، میزان بارش برف و باران، تابش خورشید و... که بر روی میزان غلظت آنها در محیط اثر می گذارد. بررسی ها نشان داده است که غلظت آنها در اتمسفر در زمستان بیشتر از بهار و تابستان است. در زمستان لایه های اتمسفر به دلیل کم بودن تابش خورشید پایین تر هستند و لذا تخریب فوتو شیمیایی این ترکیب توسط اوزن در زمستان کمتر از فصل گرم است.

۲- خواص فیزیکی و شیمیایی این ترکیبات که از طریق این خواص بر روی محیط اثر می گذارند. این ترکیبات هم به شکل ذرات ریز و هم به شکل بخار هستند و وزن مولکولی این ترکیبات تعیین کننده فاز آنها می باشد. ترکیبات با وزن مولکولی کم در فاز بخار و ترکیبات با وزن مولکولی بالا بیشتر به شکل معلق می باشند. همچنین برخی از این ترکیبات غیر فرار بوده، فشار بخار کمی دارند و در مقابل فرایندهای شیمیایی بی اثر هستند. دمای جوش و ذوب بالایی دارند، حلالیت آنها در آب کم، ویسکوزیته ی بالا و کشش سطحی کمی دارند [۲].

میزان آلودگی خاک، آب و بافت زنده بستگی به خاصیت آبدوستی آنها دارد. ترکیبات با خاصیت آبدوستی بالا تمایل به جذب در اکوسیستم ها و بافت های زنده را دارند. با افزایش تراکم حلقه ها در این ترکیبات خاصیت سمی و سرطان زایی آنها افزایش یافته و در مقابل آنزیم های تخریب کننده مقاوم تر هستند [۳].

## ۱-۱-۲- منابع PAHs

### ۱-۱-۲-۱- منابع خانگی<sup>۱</sup>

این منابع باعث آلودگی اتمسفر با PAHs می شوند. فرآیندهای مختلف تولید گرما و پخت و پز چه در صنعت و چه در مصارف خانگی از جمله راه های تولید PAHs می باشد. احتراق انواع سوخت ها جهت تأمین گرما مثل گاز طبیعی، LPG، چوب، زغال، زغال سنگ، و غیره بویژه تحت شرایطی که اکسیژن کافی نباشد باعث تولید ترکیبات متنوعی از PAHs می شوند.

### ۱-۱-۲-۱-۱- منابع متحرک<sup>۲</sup>

در حین انتقال سوخت وسایلی مثل هواپیما، قطار، کشتی، زیردریایی، و غیره... ترکیبات PAH که جزء تشکیل دهنده سوخت هستند به شکل ذرات ریز و بخار وارد فضا، اتمسفر و آب شده و باعث آلودگی می شوند.

---

<sup>۱</sup> -Domestic source

<sup>۲</sup> -Mobile source

### ۱-۱-۲-۳- منابع کشاورزی<sup>۱</sup>

در اثر سوزاندن زمین‌های بائر جهت استفاده مجدد، سوزاندن خار و خاشاک، کاه و علف‌های بیابانی تحت شرایط نامناسب از جمله کمبود اکسیژن، انواع هیدروکربن‌های آروماتیک چند حلقه‌ای حاصل می‌شود.

### ۱-۱-۲-۴- منابع طبیعی<sup>۲</sup>

سوختن اتفاقی جنگل‌ها، درخت‌زارها، علف‌زارها و مرداب‌ها از جمله منابع طبیعی ترکیبات PAHs می‌باشند.

### ۱-۱-۲-۵- منابع صنعتی<sup>۳</sup>

مهمترین منابع صنعتی که باعث تولید PAHs می‌شوند، عبارتند از:

۱- تولید آلومینیوم

۲- تولید زغال (قسمتی از مراحل تولید آهن و استیل)

۳- صنعت محافظت از چوب

۴- تولید سیمان، بتن و آسفالت

۵- صنعت پتروشیمی و فراورده‌های مربوطه

۶- تولید لاستیک

---

<sup>۱</sup> -Agricultural source

<sup>۲</sup> -Natural source

<sup>۳</sup> -Industrial source

### ۱-۱-۳- اثرات مخرب ترکیبات PAHs بر روی محیط زنده

بسیاری از این آلوده کننده‌ها از جمله مشتقات نیتروژن، گوگرد و اکسیژن دار آنها که در اثر انجام بسیاری از فعالیت‌های انسانی وارد محیط می‌شوند، به راحتی از طریق تنفس وارد بدن انسان شده و صدمات زیادی به سلامت آنها وارد می‌کنند. این ترکیبات اثرات مخربی بر روی ساختار و اعمال ارگانسیم‌های زنده دارند. همچنین آنها می‌توانند ذرات بسیار ریزی با قطر 5 میکرومتر تشکیل داده و به دلیل کوچک بودن زمان زیادی در فضا معلق باشند و مسافت‌های طولانی را طی کنند. چندین نوع از این ترکیبات نیز در محصولات غذایی مثل انواع کرم و شامپو وجود دارند که با استفاده از این محصولات این ترکیبات از طریق پوست وارد بدن می‌شوند [۴]. انسانها به شیوه‌های مختلفی در معرض PAHs هستند که در ذیل به آنها اشاره شده است.

### ۱-۱-۳-۱- تنفس هوای آلوده

PAHs از طریق تنفس وارد بدن شده و صدمات زیادی به شش وارد می‌کنند. این ترکیبات باعث سرطانی شدن سلول‌های بافت شش می‌شوند. PAHs موجود در اتمسفر در اثر عواملی مثل دود سیگار، دود ناشی از کارخانه‌ها و موتورهای دیزلی، همچنین سوختن انواع سوخت‌های فسیلی تولید می‌شوند. افرادی که سیگاری هستند، همچنین افرادی که در نزدیکی مکان‌های صنعتی، شهرهای بزرگ و نزدیک بزرگراه‌ها زندگی می‌کنند، به شدت در معرض این ترکیبات هستند.

## ۱-۱-۳-۲- خوردن و آشامیدن

غذاهایی که روی زغال چوب طبخ می‌شوند یکی از منابع مهم PAHs هستند. ماهی‌هایی که در آب‌های آلوده زندگی می‌کنند نیز حاوی این ترکیبات هستند. آب رودخانه‌ها و دریاچه‌های موجود در نزدیکی مکان‌هایی که زباله و زوائد کارخانه‌ها حضور دارد نیز آلوده به این ترکیبات هستند. خوردن آب و غذاهای آلوده باعث وارد شدن این ترکیبات به خون شده و صدمات زیادی به کبد و کلیه وارد می‌کنند.

## ۱-۱-۳-۳- لمس کردن و تماس پوستی

PAH از طریق پوست نیز می‌توانند جذب بدن شوند و باعث ایجاد تاول و سرطانی شدن سلول‌های پوست شوند. راه‌های تماس پوستی با PAH شامل :

- الف-استحمام با آب آلوده
- ب-مصرف شامپو و کرم‌هایی که حاوی برخی از این ترکیبات هستند.
- ج-تماس با خاک آلوده [۶۵].

## ۱-۱-۴- جداسازی و آنالیز هیدروکربن‌های آروماتیک چند حلقه‌ای

از جمله تکنیک‌های مناسب برای آنالیز کمی و کیفی هیدروکربن‌های آروماتیک چندحلقه‌ای کروماتوگرافی مایع با عملکرد بالا (HPLC)<sup>۱</sup> و کروماتوگرافی گازی با آشکار ساز طیف سنج جرمی (GC/MS)<sup>۲</sup> می‌باشد.

<sup>۱</sup> -High Performance Liquid Chromatography

<sup>۲</sup> -Gas Chromatography/Mass Spectrometry



## ۱-۱-۴-۱- کروماتوگرافی مایع با عملکرد بالا

کروماتوگرافی مایع با عملکرد بالا (HPLC) از پرکاربردترین تکنیکهای جداسازی است. توانایی در آنالیز ترکیبات اسیدی، بازی و خنثی با دامنه وسیعی از قطبیت و ظرفیت بالای آن به همراه امکان اتوماتیک کردن آن، HPLC را به یک تکنیک مناسب برای تعیین خصوصیات تجزیه ای انواع مختلفی از نمونه ها تبدیل کرده است. کروماتوگرافی مایع با کارایی بالا به دو صورت فاز نرمال و فاز معکوس وجود دارد [۷].

## ۱-۱-۴-۲- کروماتوگرافی گازی با آشکارساز طیف سنج جرمی

جداسازی و شناسایی کمی و کیفی هیدروکربنهای آروماتیک چند حلقه ای با سیستم GC/MS به خوبی انجام می شود. شناسایی این ترکیبات از روی مقایسه ی پارامترهای بازداری آنها با ترکیبات استاندارد امکان پذیر است. این ترکیبات از نظر ساختمانی بسیار شبیه به هم بوده و اکثراً ایزومر هستند. بنابراین پس از یونیزاسیون در دستگاه طیف سنج جرمی ضمن تفکیک شدن و نوآرایی، الگوی طیفی یکسانی دارند. بنابراین حتی در صورت استفاده از دستگاههای GC/MS برای شناسایی آنها علاوه بر اطلاعات جرمی به پارامترهای بازداری نیاز داریم [۸].

## ۱-۱-۴-۱-۱- اساس کروماتوگرافی گازی

کروماتوگرافی گازی یک روش آنالیزی برای جداسازی و شناسایی اجزای مخلوطها است. شامل دو فاز ساکن مایع و متحرک گازی است. فاز متحرک، گاز بی اثر است و فقط نقش حمل اجزاء نمونه را بر عهده دارد. این دستگاه از یک ستون بلند با قطر داخلی تقزیباً کم حدود ۰/۵-۱/۵ میلی متر تشکیل شده است. فاز ساکن مایع توسط یک نگهدارنده در داخل ستون تعبیه می شود. نمونه‌های گازی به شکل بخار توسط روشهای مختلف تزریق، در سر ستون به همراه گاز حامل به درون ستون

تزریق می شوند. اساس جداسازی ترکیبات، تفاوت در خواص فیزیکی و شیمیایی و نوع برهمکنش آنها با فاز ساکن است. نمونه ممکن است در اثر تفاوت در خواص زیر برهمکنشهای متفاوتی با فاز ساکن ستون کروماتوگرافی داشته باشد:

۱-حلالیت

۲-جذب سطحی

۳-فراریت

۴-اندازه و شکل

زمان لازم برای رسیدن نمونه از لحظه ی تزریق به آشکارساز در انتهای ستون را زمان بازداری گویند. از انواع مختلف فاز ساکن می توان به آلومینا با خواص بازی، سیلیکاژل با خواص اسیدی و زغال چوب اشاره کرد. همچنین فازهای ساکن دارای خواص متنوع غیر قطبی تا قطبی هستند که در این تقسیم بندی نیز می توان به مثالهای زیر اشاره کرد.

۱- پلی سیکلوکسان (دارای سیلیسیم با اتصالات زنجیرهای آلکانی و غیر قطبی)

۲- پلی اتیلن گلیکول (دارای گروه عاملی هیدروکسیل و قطبی)

۱-۱-۴-۲-۲-آشکارساز

آشکارسازهای مختلفی مانند آشکار ساز شعله یونشی<sup>۱</sup> (FID)، الکترون گیر اندازی<sup>۲</sup> (ECD)، هدایت گرمایی<sup>۳</sup> (TCD) و اسپکتروسکوپی جرمی<sup>۴</sup> (MS) می توان در انتهای ستون به کار برد. گازهای حامل می توانند شامل هلیوم، هیدروژن، آرگون و نیتروژن باشند

<sup>۱</sup>-Flame Ionization Detector

<sup>۲</sup>-Electron Capture Ddetector

<sup>۳</sup>-Thermal conductive Detector

<sup>۴</sup>-Mass Spectrometry

که برای سیستم GC/MS هلیوم بهتر است زیرا فشار آن در سرعت جریانهای مختلف تقریباً ثابت است [۹].

### ۱-۱-۳-۴-اندیس بازداری<sup>۱</sup>

اندیس بازداری کواتس در سال ۱۹۵۸ برای اولین بار به عنوان یک پارامتر برای شناسایی ترکیبات شیمیایی با استفاده از کروماتوگرافی توسط کواتس معرفی گردید. تعیین این اندیس برای هر ترکیب از کروماتوگرام مربوط به مخلوط آن جسم و دو پارافین نرمال که زمان بازداری ترکیب مورد نظر در بین آن دو باشد، امکان پذیر است. در این مورد پارافین‌های نرمال به عنوان ترکیبات استاندارد در نظر گرفته می‌شوند. طبق قرارداد مقدار این اندیس برای یک پارافین نرمال صد برابر تعداد کربن‌های آن در نظر گرفته می‌شود. این مقدار برای پارافین‌های نرمال ثابت بوده و مستقل از نوع ستون، فاز ساکن، درجه حرارت و سایر شرایط می‌باشد. در حالیکه برای سایر ترکیبات با تغییر شرایط ستون، این اندیس تغییر می‌کند. مطالعات تجربی انجام شده نشان می‌دهند که نمودار تغییرات لگاریتم زمان بازداری تصحیح شده نسبت به تعداد اتم‌های کربن موجود در یک سری هم‌مرده خطی است. وجود رابطه‌ی خطی فوق‌الذکر امکان استفاده از معادله‌ی (۱-۱) را برای محاسبه‌ی اندیس بازداری برقرار می‌سازد. رابطه‌ی (۲-۱) محاسبه‌ی زمان بازداری تصحیح شده را نشان می‌دهد.

$$RI = 100N + 100 \frac{\log t_R(A) - \log t_R(N)}{\log t_R(N+1) - \log t_R(N)} \quad (1-1)$$

زمان بازداری تصحیح شده ( $t'_R$ ) از تفاضل زمان بازداری نمونه‌ی مورد نظر ( $t_R$ ) و زمان بازداری گونه‌ای که در ستون نگه‌داری نشده است ( $t_M$ ) بدست می‌آید.

$$t'_R = t_R - t_M \quad (2-1)$$

<sup>۱</sup> -Retention Index

در معادله (۱-۱)، RI اندیس بازداری،  $t_R(A)$  زمان بازداری تصحیح شده برای ترکیب A،  $t_{R(N)}$  و  $t_{R(N+1)}$  نیز زمان بازداری پارافین های نرمال با تعداد کربن N و N+1 بوده که زمان بازداری ترکیب A بین زمان بازداری پارافین های فوق می باشد.

در کروماتوگرافی گازی مقدار اندیس بازداری مستقل از نوع و کیفیت نگهدارنده و پارامترهای فیزیکی گاز حامل از قبیل سرعت جریان و فشار گاز وردی و خروجی از ستون است. از طرفی دیگر طبیعت شیمیایی فاز ساکن و متحرک و دمای ستون شدیداً بر اندیس بازداری یک ترکیب اثر می گذارند. بررسی های صورت گرفته نشان می دهد که اندیس بازداری کواتس تحت تأثیر خواص فیزیکی و شیمیایی مولکول قرار داشته و از آن می توان برای مقایسه قطبیت فازهای ساکن استفاده کرد.

۱-۱-۴-۴-مشکلات موجود در اندازه گیری هیدروکربنهای آروماتیگ

### چند حلقه ای باسیستم GC/MS

علی رغم توانمندی این سیستم برای جداسازی ترکیبات پیچیده از جمله هیدروکربنهای آروماتیگ چند حلقه ای مشکلاتی نیز در حین اندازه گیری این ترکیبات وجود دارد که می توان به موارد زیر اشاره کرد:

۱- هیدروکربنها با وزن مولکولی زیاد تمایل زیادی به اتصال به فاز ساکن مایع دارند و شسته نمی شوند.

۲- حضور ماتریکس مزاحمت جدی برای سیستم اسپکتروسکوپی جرمی است و تفکیک طیف ها مشکل شود.

۳- کنترل دما برای این سیستم مهم است، زیرا تغییرات دما باعث ایجاد نوسان در زمان بازداری می شود.

۴- استانداردهای مورد نیاز به راحتی قابل تهیه نیستند.

۵- تهیه ی نمونه جهت آنالیز نیاز به مراحل زیادی دارد که مستلزم صرف وقت و هزینه ی بسیار است.

۶- عدم قطعیت‌هایی در حین جمع آوری نمونه، استخراج و آماده سازی آن جهت آنالیز وجود دارد و همچنین حضور آلوده کننده های مختلف همراه نمونه باعث ایجاد مزاحمت و خطا در حین اندازه گیری می شود [۱۰].

## ۱-۲-هدف تحقیق

با توجه به اینکه هیدروکربن‌ها ی آروماتیک چند حلقه ای سمی و سرطان زا هستند به طور جدی سلامت انسان و محیط زیست را تحدید می کنند. به همین جهت از نظر دانش مندان بررسی اثرات این ترکیبات ضروری می باشد. بنابراین باید یک سری اطلاعات کیفی و کمی راجع به این ترکیبات کسب کرد. یکی از راه های آنالیز این ترکیبات که اکثراً ایزومر بوده و خواص فیزیکی و شیمیایی مشابهی دارند، اندازه گیری اندیس بازداری است. همان طور که در بخش ۱-۱-۴-۳-گفته شد این اندیس با روش های کروماتوگرافی قابل اندازه گیری است. با وجود مزایای تکنیک های کروماتوگرافی مشکلات زیادی نیز وجود دارد که در بخش ۱-۱-۴-۴-به برخی از آنها اشاره شد. وجود این محدودیت ها در روش های آزمایشگاهی باعث شد که در کنار این روش ها شیوه های تئوری نیز برای استفاده از نتایج تجربی توسعه یابند. مطالعات کمومتریکس راهی برای استخراج حداکثر اطلاعات از نتایج تجربی با استفاده از انجام یک سری محاسبات آماری و ریاضی می باشد. مطالعات کمی روابط ساختار-خاصیت (QSPR)<sup>۱</sup> یکی روش های کمومتریکس است که با استفاده از آن می توان روابط خطی یا غیر خطی بین ویژه گی های ساختاری و خواص ترکیبات برقرار کرده و خاصیت مورد نظر را برای ترکیبات پیش بینی کرد. در این تحقیق تصمیم گرفته شد که با بهره گیری از QSPR و استفاده از نتایج تجربی اندیس بازداری یک سری ترکیبات PAHs، این پارامتر را مدل سازی کرده و آن را برای سایر هیدروکربن های آروماتیک چند حلقه ای پیش بینی کنیم.

<sup>۱</sup>-Quantitative Structure-Property Relationship

## ۱-۳-مراحل تحقیق

در این تحقیق جهت مطالعه ی کمی ساختار-خاصیت اندیس بازداری هیدروکربن های آروماتیک چند حلقه ای از روش های آنالیز رگرسیون خطی چندگانه (MLR)<sup>۱</sup> و شبکه ی عصبی مصنوعی (ANN)<sup>۲</sup> استفاده شد. ابتدا ساختار هندسی ترکیبات با استفاده از نرم افزار HyperChem رسم شد. سپس با استفاده از نرم افزار Dragon مقادیر توصیف کننده ها محاسبه شدند. با بهره گیری از نرم افزار SPSS و محاسبه ی ضرایب همبستگی بین توصیف کننده ها، از بین زوج توصیف کننده هایی که با یکدیگر وابستگی داشتند توصیف کننده ای که کمترین وابستگی با اندیس بازداری را داشت حذف شد. همچنین توصیف کننده هایی که دارای مقادیر یکسان یا صفر برای ترکیبات بودند از بین سایر توصیف کننده ها حذف شدند. سپس با استفاده از آنالیز رگرسیون خطی چند گانه مدل خطی ساخته شد. در ادامه با استفاده از شبکه ی عصبی مصنوعی و انتخاب الگوریتم آموزشی Levenberg Marquardt یک شبکه ی پیش خور طراحی شد. توصیف کننده های برگزیده در مدل خطی به عنوان ورودی شبکه انتخاب شدند. سری داده ها به سه دسته ی آموزش، پیش بینی و تست تقسیم بندی شدند. در نهایت با استفاده از سری آموزش شبکه آموزش داده شد و اندیس بازداری شبیه سازی شد. برای ارزیابی نتایج حاصل از شبکه و مدل خطی اندیس بازداری برای سری پیش بینی محاسبه شد. همچنین با استفاده از روش ارزیابی تقاطعی اندیس بازداری برای سری داده ها توسط مدل خطی و غیر خطی پیش بینی شد.

---

<sup>۱</sup>-Multi linear Regression

<sup>۲</sup>-Artificial Neural Network

## ۱-۴- مروری بر کارهای انجام شده

در سال ۱۹۸۵ گروهی با استفاده از اندیس بازداري اندازه گیری شده با ستون ژل آبی<sup>۱</sup> و حلال THF و محاسبه عدد واینر<sup>۲</sup>، مقدار این پارامتر را برای این ترکیبات پیش بینی کردند. عدد واینر از محاسبه ماتریس فاصله بین اتمی بدست می آید [۱۱].

در سال ۲۰۰۰ فری راومارسیا<sup>۳</sup> بر روی ۴۸ هیدروکربن آروماتیک چند حلقه ای مطالعات ارتباط کمی ساختار-خاصیت<sup>۴</sup> انجام دادند. آنها از مقادیر تجربی اندیس بازداري که توسط ویس و ساندر با استفاده از کروماتوگرافی فاز معکوس در سال ۱۹۸۶ اندازه گیری شد، برای مدل سازی این پارامتر با روش رگرسیون خطی چند گانه استفاده کردند [۱۲].

در سال ۲۰۰۳ نیز گروهی با استفاده از روش رگرسیون خطی اندیس بازداري ۶۷ مولکول آروماتیک چندحلقه ای را پیش بینی کردند. مقادیر تجربی پارامتر بازداري با استفاده از کروماتوگرافی فاز معکوس با فاز ساکن اصلاح شده بوسیله ی اکتادسیل سیلان اندازه گیری شده بود. ضریب رگرسیون خطی<sup>۵</sup> برای این مدل خطی ۰/۸۸۲ است [۱۳].

در سال ۲۰۰۵ توسط گروهی با استفاده از روش رگرسیون خطی چندگانه اندیس بازداري هیدروکربنهای آروماتیک چند حلقه ای مدل سازی شد. مقادیر توصیف کننده ها با استفاده از روشهای نیمه تجربی AM1 و نرم افزار Dragon محاسبه شدند. چهار مدل طراحی شد.

مدل اول با توصیف کننده ی R13 که میزان شاخه ای شدن ترکیب را نشان می دهد با ضریب رگرسیون ۰/۹۵۷ بدست آمد. دومین مدل با توصیف کننده ی KHI2 یا اندیس Kier-Hall و ضریب رگرسیون خطی ۰/۹۳۳۲ است. سومین مدل با توصیف کننده ی NBR که اشاره به تعداد حلقه های

<sup>۱</sup> Water gel column

<sup>۲</sup> -Wiener number

<sup>۳</sup> - Ferreira and Marcia

<sup>۴</sup> -Quantitative Structure-Property relationship

<sup>۵</sup> - Linear regression coefficient



آروماتیک دارد با ضریب رگرسیون خطی  $0/6398$  می باشد. مقادیر تجربی اندیس بازداری توسط کراتس و واحن دول<sup>۱</sup> با کروماتوگرافی گازی بر روی ستون HP5880 بدست آمده است [۱۴]. در سال ۲۰۰۶ با استفاده از اندیس بازداری بدست آمده بر روی فاز ساکن DB-5 و SE-52 مقدار این پارامتر را مدل سازی کردند. مقدار ضریب همبستگی خطی برای مقادیر پیش بینی شده برای ستون DB-5،  $0/8939$  و میانگین درصد خطا  $1/43$  درصد است. همچنین برای ستون SE-52 ضریب همبستگی خطی  $0/94$  و میانگین خطا  $1/42$  درصد است [۱۵]. در سال ۲۰۰۸ با استفاده از روش برازش مرحله ای اندیس بازداری را برای یک سری از هیدروکربن ها مدل سازی کردند. ضریب همبستگی مقادیر پیش بینی شده اندیس بازداری  $0/992$  است. توصیف کننده های وارد شده در مدل شامل پتانسیل الکتریکی سطح مولکول، حجم و اندروالسی و انرژی بالاترین اوربیتال اشغال شده<sup>۲</sup> توسط الکترون های لایه ظرفیت است. مقادیر تجربی اندیس بازداری برای ۱۱۴ مولکول با ستون HF/6-31G بدست آمده است [۱۶].

---

<sup>۱</sup> -Vahlen dul & Krats

<sup>۲</sup> -High Orbital Molecular Occupied

فصل دوم

کمو متریکس

عبارت کمومتریکس توسط اسوانت ولد<sup>۱</sup> و بروس کووالسکی<sup>۲</sup> در اوایل ۱۹۷۰ تعریف شد [۱۷]. قبلاً عباراتی مانند بیومتریکس<sup>۳</sup> و اقتصاد سنجی<sup>۴</sup> در زمینه های علوم زیست و اقتصاد نیز معرفی شده اند. پس از آن انجمن بین المللی کمومتریکس تأسیس شد. از آن موقع به بعد کمومتریکس توسعه یافت و اکنون در شاخه های مختلف شیمی، به ویژه شیمی تجزیه مورد استفاده قرار می گیرد. هاوری و هیرش<sup>۵</sup> [۱۸] در اوایل سال ۱۹۸۰ پیشرفت های کمومتریکس را در مراحل مختلف رده بندی کردند. در این محدوده ی زمانی، شیمیدانان خود را به آنالیز داده ها، از جمله محاسبه ی پارامترهای آماری مانند میانگین، انحراف استاندارد، و حدود اطمینان، محدود کرده بودند. آن ها تحقیق خود را بطور ویژه، به همبستگی تعداد زیادی از داده های شیمیایی به ویژگیهای ملکولی مربوطه، اختصاص دادند. این اقدامات اولیه اساس یکی از شاخه های مهم کمومتریکس، ارتباط کمی ساختار - خاصیت<sup>۶</sup> (QSAR) را فراهم کرد.

دومین مرحله تکامل کمومتریکس در سال ۱۹۷۰ اتفاق افتاد. در این زمان کمومتریکس توجه شیمیدانان، به خصوص شیمیدانان تجزیه را به خود جلب کرد، زیرا روش های نوینی برای آنالیز داده ها در اختیارشان قرار داد. تکامل بعدی کمومتریکس که توسط هاوری و هیرش در مقاله آنها [۱۸] و سپس توسط

---

<sup>۱</sup> - Svant wold

<sup>۲</sup> - Bruce R. Kowalski

<sup>۳</sup> - Biometrics

<sup>۴</sup> - Econometrics

<sup>۵</sup> - Howery and Hirsch

<sup>۶</sup> -Quantitative Structure-Property Relationship

براون<sup>۱</sup> [۱۷] پیش بینی شده بود، از اوایل سال ۱۹۸۰ آغاز شد. کمومتریکس در

دانشگاه های اروپا و آمریکا با رشته های شیمی برای فارغ التحصیلان و دانشجویان تحصیلات تکمیلی ترکیب شد. علاوه بر این، کمومتریکس ابزاری برای شیمی دانان شد و یک شاخه وابسته به شیمی در بسیاری از دانشگاه های آمریکا و اروپا و بعضی دانشگاه ها در چین و سایر کشورها به شمار آمد.

## ۲-۲- ارتباط کمی ساختار - خاصیت

یکی از کاربردهای ویژه روشهای کمومتریکس توسعه ارتباط کمی ساختار - خاصیت و ارزیابی داده های تجزیه ای از طریق آن است. مولکولهای مشابه با تغییر کوچکی در ساختارشان می توانند خواص شیمیایی کاملاً متفاوتی داشته باشند. این نوع ارتباط بین ساختار مولکولی و تغییر در خاصیت شیمیایی، مرکز توجه در مطالعات QSPR می باشد [۱۹]. در واقع، QSPR سعی در پیدا کردن رابطه هماهنگ میان خاصیت شیمیایی و ویژگی های مولکولی، به منظور کاربرد این قواعد برای ارزیابی خاصیت ترکیبات جدید دارد [۲۰]. نظریه QSPR بر این اساس است که خاصیت یک ترکیب به ویژه گی های ساختاری آن بستگی دارد. ویژه گی های ساختاری با خصوصیات الکترونیکی، آبگریزی، فضای، هندسی و غیره ... توصیف می شوند [۲۱].

برای انجام یک مطالعه QSPR ابتدا باید مقادیر یک خاصیت شیمیایی که برای گروهی از ترکیبات اندازه گیری شده تهیه شود. سپس توصیف کننده های مناسب با استفاده از الگوریتم های ریاضی از ساختار مولکولی محاسبه شوند. مرحله بعد دستیابی به ارتباط ریاضی بین ساختار و خاصیت است.

---

<sup>۱</sup> - Brown

معمولترین تکنیکهای چند متغیری شامل رگرسیون خطی چند گانه (MLR)<sup>۱</sup>، آنالیز اجزای اصلی (PCA)<sup>۲</sup> و حداقل مربعات جزئی (PLS)<sup>۳</sup> می باشند. تکنیکی که معمولاً مورد استفاده قرار می گیرد، روش MLR است. با این حال این روش تنها زمانی می تواند بکار رود که تعداد مولکول های مورد بررسی بیشتر از تعداد توصیف کننده ها باشد و متغیرها با هم همبستگی زیادی نداشته باشند.

روش های QSPR به طور گسترده در پیش بینی خواص فیزیکی و شیمیایی مختلفی مثل نقطه جوش، نقطه ذوب، فشار بخار، حلالیت آبی و غیره... استفاده شده اند. در مطالعات QSPR امکان انتخاب تکنیک های مدل سازی مناسب برای ایجاد مدل وجود دارد. مطالعات QSPR شامل مراحل زیر می باشد.

۱- فراهم کردن سری داده ها

۲- بهینه کردن ساختار هندسی ترکیبات

۳- محاسبه توصیف کننده ها

۴- انتخاب بهترین توصیف کننده ها

۵- مدل سازی

۶- ارزیابی مدل

## ۲-۱-۲- جمع آوری سری داده ها

برای دستیابی به مدلی مناسب، اولین مرحله انتخاب یک دسته از مولکولها است که مقادیر تجربی یک خصوصیت شیمیایی یا فیزیکی آنها مانند زمان بازداری، فاکتور آبگریزی و غیره... در دسترس باشد. شرایط اندازه گیری خاصیت مورد نظر برای همه ترکیبت باید یکسان باشد. سری داده

---

<sup>۱</sup> -Multiple linear regression

<sup>۲</sup> -Principle component regression

<sup>۳</sup> -Partial Least Squares

ها به سه دسته آموزش، پیش بینی و تست تقسیم می شوند. عملیات مدلسازی روی سری آموزش انجام می شود و پس از محاسبه مدل مناسب توسط این سری، مدل مورد نظر به منظور ارزیابی اعتبار آن برای سری پیش بینی مورد استفاده قرار می گیرد. همچنین جهت تایید صحت اعتبار مدل خاصیت مورد نظر برای سری تست نیز با استفاده از مدل طراحی شده محاسبه می شود. هر چه سری داده ها بزرگتر بوده و تنوع گونه ها بیشتر باشد مدل بدست آمده از قدرت پیش بینی بیشتری برخوردار خواهد بود.

## ۲-۲-۲- بهینه سازی ساختار هندسی ترکیبات

در این مرحله با کمک شیمی محاسباتی و نرم افزارهای مناسب شکل ترکیبات بهینه شده و ویژه گی های ساختاری آنها به صورت مقادیر عددی محاسبه می شود. شیمی محاسباتی ساختارهای مولکولی را به صورت مقادیر عددی بیان می کند و رفتار آنها را معادلات کوانتومی و فیزیک کلاسیک شبیه سازی می کند. با استفاده از این روشها می توان بدون داشتن دانش از اصول اولیه و طریقه محاسبه، ساختارها و واکنش های شیمیایی را شبیه سازی کرد. روش های محاسباتی بر پایه ی مکانیک کوانتومی بوده و شامل گروه های زیر می باشند:

۱- روش های آغازین<sup>۱</sup>

۲- روش های نیمه تجربی<sup>۲</sup>

۳- روشهای تابع چگالی<sup>۳</sup>

۴- روش های مکانیک مولکولی<sup>۱</sup>

---

<sup>۱</sup>-Ab initio methods

<sup>۲</sup>-Seemi-empirical methods

<sup>۳</sup>-Density functional theory(DFT)

با استفاده از نرم افزار HyperChem ساختار ترکیبات رسم می شود. برای رسم شکل هندسی مولکول از روش نیمه تجربی AM1 استفاده می شود. در روش های نیمه تجربی یک سری از مقادیر تجربی وارد شده اند و فقط الکترون های لایه ی ظرفیت در نظر گرفته می شوند. با این نرم افزار می توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچشی، بار اتمها و انرژی تشکیل مولکول را بدست آورد [۲۲-۲۵].

## ۲-۲-۳- محاسبه توصیف کننده ها

۱۹

توصیف کننده های مولکولی مقادیر عددی هستند که ساختار یا شکل مولکول را توصیف می کنند و به پیش بینی فعالیت و خصوصیات مولکولها در آزمایش های پیچیده کمک می کنند [۲۶].

برخی از ویژگیهای لازم برای یک توصیف کننده مولکولی مناسب عبارتند از:

۱- ساده بودن

۲- مستقل بودن

۳- تفسیر ساختار مولکول (غنی بودن از نظر اطلاعات)

۴- عدم همبستگی با سایر توصیف کننده ها

۵- تغییر منظم با تغییر تدریجی در ساختارها

۶- وابستگی صحیح به اندازه مولکول

۷- تمایز بین ایزومرهای مختلف مولکول [۲۷]

انواع کاربردهای توصیف کننده های مولکولی به صورت زیر می باشند:

- آنالیز و مدلسازی فعالیت ترکیبات دارویی

- ارزیابی خصوصیات شیمیایی

- ارزیابی برخوردهای مولکولهای محیط زیستی

- ارزیابی و مدلسازی فعالیت مولکولها در سم شناسی

- پیش بینی خصوصیات، فعالیت، یا رفتارهای مولکولهایی که هنوز سنتز نشده اند.

توصیف کننده های مولکولی می توانند به روشهای مختلفی دسته بندی شوند. ساده ترین روش برای دسته بندی توصیف کننده ها براساس طبیعت آنها ( تجربی یا تئوری بودن آنها) است. دسته بندی دیگر بر اساس ابعاد مولکول است [۲۸]. این نوع دسته بندی به صورت زیر می باشد:

## ۲-۲-۳-۱- توصیف کننده های صفر بعدی

این دسته از توصیف کننده ها که از فرمول شیمیایی مولکول مشتق می شوند تعداد و نوع اتمها، جرم مولکولی، و حجم واندروالس را تعیین می کنند. این توصیف کننده ها به توصیف کننده های ساختاری نیز معروفند که علاوه بر موارد فوق شامل توصیف کننده های مربوط به نوع پیوندها و حضور حلقه ها در مولکول نیز می شوند. تعداد کل اتمها در مولکول، تعداد یک عنصر شیمیایی خاص ( C,N,O,H,F,... ) در مولکول، تعداد گروه های عاملی خاص در مولکول، تعداد پیوندهای ساده، دو گانه، سه گانه و آروماتیک در مولکول، تعداد کل حلقه ها براساس تعداد اتمهای آنها (حلقه های شش تایی، پنج تایی و...) در مولکول، وزن مولکولی و متوسط وزن اتمی مثالهایی از توصیف کننده های ساختاری می باشند.

## ۲-۲-۳-۲- توصیف کننده های یک بعدی

لیستی از نمایش زیر ساخت های یک مولکول می تواند به عنوان یک توصیف کننده تک بعدی در نظر گرفته شود و متشکل از لیستی از بخشهایی از مولکول (مثلاً گروه های عاملی، گروههای



استخلافی و...) باشد. توصیف کننده های گروههای عاملی، قطعات اتم مرکزی<sup>۱</sup>، توصیف کننده های تجربی و خصوصیات مولکول<sup>۲</sup> زیر گروه های این دسته می باشند. در زیر به توضیح هر یک از این زیر گروه ها می پردازیم:

## ۲-۲-۳-۱- گروه های عاملی

این زیر گروه شامل توصیف کننده های ساده ای می شود که به عنوان تعداد گروه های عاملی خاص در یک مولکول معرفی می شوند. این توصیف کننده ها با دانستن ترکیب مولکول و اتصالات اتمی محاسبه می شوند. تعداد اتمهای پذیرنده پیوند هیدروژنی (nHAcc)<sup>۳</sup>، تعداد کربنهای نوع اول، دوم و یا سوم در مولکول (n=CH<sub>2</sub>, n=CHR, n=CR<sub>2</sub>)، تعداد ایزوسیانات های آروماتیک (nNCOph) در مولکول (nNCOph) و نمونه هایی از این گروهند.

## ۲-۲-۳-۲- قطعات اتم مرکزی

این زیر گروه شامل توصیف کننده های مولکولی ساده ای می باشد که به عنوان تعدادی از یک اتم خاص در مولکول تعیین می شوند. این توصیف کننده ها با دانستن ترکیب مولکولی و اتصالات اتمی محاسبه می شوند. هیدروژن متصل به اکسیژن در فنول، انول و یا کربوکسیل، آمین نوع اول متصل به زنجیر آروماتیک، فلئور متصل به کربن نوع سوم با هیبرید SP<sup>3</sup> و ... نمونه هایی از این توصیف کننده ها هستند.

## ۲-۲-۳-۲- توصیف کننده های تجربی

<sup>1</sup> -Atom-centered fragment

<sup>2</sup> -Molecular properties

<sup>3</sup> -Number of acceptor atom for H-Bonds

اندیس غیر اشباعیت ( $U_i$ )<sup>۱</sup>، فاکتور آب دوستی ( $Hy$ )<sup>۲</sup>، نسبت آروماتیک ( $ARR$ )<sup>۳</sup> از جمله این توصیف کننده ها هستند.

## ۲-۲-۳-۴- خصوصیات مولکولی

این زیر گروه شامل توصیف کننده هایی می شود که به بیان یک خصوصیت فیزیکی شیمیایی از ترکیبات می پردازند. انکسار مولی گوس - کریپن ( $MR$ )<sup>۴</sup> و ضریب تقسیم آب - اکتانول موریگوچی ( $MLOGP$ )<sup>۵</sup> از جمله این توصیف کننده ها هستند.

## ۲-۳-۳-۲- توصیف کننده های دو بعدی

یک گراف مولکولی متشکل از اطلاعات توپولوژیکی یا دو بعدی است. گراف مولکولی چگونگی پیوند یافتن اتمها در مولکولها، نوع پیوند و همچنین برهمکنش اتمهای خاص را توصیف می کند. این دسته از توصیف کننده ها شامل زیرگروه های توپولوژیکی، توصیف کننده های ضرایب بار ( $BCUT$ )، و توصیف کننده های خود ارتباطی دو بعدی می باشند که در زیر به اختصار به هر کدام از این زیر گروه ها می پردازیم.

## ۲-۲-۳-۱- توصیف کننده های توپولوژیکی

---

<sup>۱</sup>-Unsaturation index

<sup>۲</sup>-Hydrophilic factor

<sup>۳</sup>-Aromatic ratio

<sup>۴</sup>-Ghose crippen molar refractivity

<sup>۵</sup>-Moguchi octanol-water coefficient

این توصیف کننده ها براساس نمایش گراف مولکول هستند. آنها در واقع مقادیری عددی از توپولوژی مولکولی هستند که با استفاده از عملگرهای جبری در ماتریسهای که گراف های مولکولی را نمایش می دهند و مقادیر آنها مستقل از تعداد رئوس یا شماره آنهاست، بدست می آیند. این توصیف کننده ها می توانند به یک یا چند تا از مشخصه های ساختاری مولکول مثل اندازه، شکل، تقارن، شاخه دار شدن، حلقوی بودن و... حساس باشند و همچنین می توانند اطلاعات شیمیایی را با توجه به نوع اتم و چندگانگی پیوند کدگذاری کنند [۲۹].

۲۳

## ۲-۲-۳-۲- توصیف کننده های BCUT

این توصیف کننده ها مقادیر ویژه ای از یک ماتریس ارتباطی تغییر شکل یافته به نام ماتریس ((بوردن))<sup>۱</sup> هستند. ماتریس بوردن یک گراف مولکولی وابسته به هیدروژن است که به صورت زیر تعریف می شود:

الف- مولفه های قطر اعداد اتمی عناصر هستند.

ب- مولفه های غیر قطری، اتمهای پیوندی هستند که برابر با  $10^{-1} \times \pi^*$  می باشند (  $\pi^*$  درجه پیوند است که برای پیوندهای یگانه، ۱، دوگانه، ۲، سه گانه، ۳، و آروماتیک، ۱/۵ می باشد).

ج- مولفه های غیر قطری مربوط به پیوندهای انتهایی هستند با ۰/۰۱ جمع می شوند.

د- بقیه مولفه ها با ۰/۰۰۱ تنظیم می شوند.

توصیف کننده های BCUT سه دسته از ماتریسها را در بر می گیرند که عناصر قطر آنها مربوط به مقادیر نسبی بار اتمی، مقادیر نسبی قطبش پذیری اتمی و توانایی های پیوند هیدروژنی می شوند. مقادیر ویژه بالا و پایین این ماتریس ها توصیف کننده های مجزایی را تشکیل می دهند [۳۰-۳۲].

---

<sup>۱</sup> -Borden matrix

## ۲-۲-۳-۴-توصیف کننده های سه بعدی

دسته دیگری از توصیف کننده های تئوری شامل توصیف کننده های سه بعدی است که با شروع از یک نمایش هندسی یا سه بعدی یک مولکول محاسبه می شوند. زیرگروه های این دسته شامل اندیس های آروماتیکی، خصوصیات مولکولی راندیک، توصیف کننده های هندسی، توصیف کننده های 3D- RDF, WHIM, MORSE و توصیف کننده ههای GETAWAY هستند که در زیر به اختصار این زیرگروه ها معرفی شده اند.

## ۲-۲-۳-۱-اندیس های آروماتیکی

این توصیف کننده ها اطلاعات را براساس درجه عدم استقرار الکترون در مولکول کدگذاری می کنند. از آنجا که مفهوم آروماتیسیته به وضوح تعریف نشده است، توصیف معمول آروماتیسیته به صورت عدم استقرار الکترونها  $\pi$  است که پایداری مولکولهای حلقوی و چند حلقه ای مزدوج را موجب می شود. اندیسهای آروماتیکی معمولاً از طول پیوند و مرتبه پیوند مولکولها محاسبه می شوند. بنابراین، مقادیر آنها به شدت به مرحله بهینه سازی ساختار مولکول و رمزگذاری پیوندهای آروماتیک بستگی دارند. مدل نوسانگر هماهنگ اندیس آروماتیک (HOMA)<sup>۱</sup> آروماتیسیته (AROM) و اندیس RC جاگ (RCI) نمونه هایی از این توصیف کننده ها هستند [۳۳-۳۷].

<sup>۱</sup>-Harmonic oscillator model of aromaticity index

## ۲-۲-۳-۴-۲-خصوصیات مولکولی راندیک

پروفیل های مولکولی، تناوبی از توصیف کننده های مولکولی هستند که توسط راندیک در سال ۱۹۹۵ پیشنهاد شدند و از فواصل هندسی درون اتمی یک مولکول مشتق می شوند. این توصیف کننده ها به دو دسته اصلی تقسیم می شوند. دسته اول dpk ها (۲۰ و... و ۲ و ۱) هستند که بیشتر

مربوط به ساختار کلی سه بعدی مولکول هستند. دسته دیگر spk ها هستند و مربوط به شکل مولکول می شوند. خصوصیات مولکولی راندیک می توانند در مدلسازی های QSPR بکار روند [۳۸-۴۰].

۲۵

## ۲-۲-۳-۴-۳-توصیف کننده های هندسی

این توصیف کننده ها به شیوه های متفاوتی تعریف می شوند ولی همیشه از ساختار سه بعدی مولکول مشتق می شوند. بطور کلی توصیف کننده های هندسی براساس تعدادی هندسه مولکولی بهینه شده که بوسیله روشهای شیمی محاسباتی بدست آمده اند، محاسبه می شوند. از آنجا که نمایش هندسی مولکول شامل موقعیت های نسبی اتمها در فضای سه بعدی است، معمولاً نسبت به توصیف کننده های توپولوژیکی اطلاعات بیشتر و قدرت تمایز بالاتری برای ساختارهای مشابه مولکولی و صورت بندی های مختلف مولکول تامین می کنند. با وجود محتوای اطلاعاتی بالای آنها، گاهی توصیف کننده های هندسی غیر قابل استفاده می شوند. برای محاسبه ی این توصیف کننده ها نیاز به بهینه سازی هندسی و پردازش کافی می باشد. علاوه بر این برای مولکولهای انعطاف پذیر، صورتبندی های زیادی در دسترس است که از یک طرف اطلاعات جدیدی نیز در اختیار خواهند گذاشت ولی از طرف دیگر مشکل پیچیدگی می تواند به طور مؤثری افزایش یابد. اکثر این توصیف کننده ها از ماتریس هندسی مشتق می شوند. ماتریس هندسی یک ماتریس مربع متقارن است که فاصله هندسی بین تمام جفت اتمها را جمع آوری می کند. مجموع سطر ماتریس هندسی درجه

فاصله هندسی نام دارد که حاصل جمع فاصله های هندسی بین یک اتم و سایر اتمها است. ضریب سه بعدی واینر (W3D)<sup>۱</sup>، ضریب سه بعدی بالابان (J3D)<sup>۲</sup> و درجه فاصله هندسی متوسط (AGDD)<sup>۳</sup> از جمله این توصیف کننده ها هستند [۴۱].

## ۲-۲-۳-۴- توصیف کننده های RDF

توصیف کننده های RDF براساس توزیع فاصله در نمایش سه بعدی مولکول هستند و یک تابع توزیع شعاعی را تشکیل می دهند [۴۲]. تابع توزیع شعاعی می تواند به عنوان توزیع احتمالی یافتن یک اتم در حجم کروی شعاع  $r$  تفسیر شود [۴۳].

## ۲-۲-۳-۵- توصیف کننده های 3D-MORSE<sup>۴</sup>

توصیف کننده های 3D-MORSE از معادله ای که در مطالعات پراش الکترونی بکار می رود و امکان نمایش سه بعدی مولکول را با مقادیری ثابت فراهم می کند، مشتق شده اند. خصوصیات اتمی گوناگونی مانند جرم اتمی، حجم و اندروالسی، الکترونگاتیوی و قطبش پذیری می توانند در محاسبه ی این توصیف کننده ها به کار روند. همچنین این دسته قادرند تفاوت های ساختاری ترکیبات هم خانواده را که از نظر ساختاری بسیار به هم نزدیک هستند نشان دهند. این توصیف کننده ها امکان مطالعه ارتباط بین ساختار سه بعدی ترکیلات آلی و خصوصیات فیزیکی، شیمیایی، و بیولوژیکی را فراهم می سازند [۴۴-۴۵].

---

<sup>۱</sup>-Winer index

<sup>۲</sup>-3D-balaban index

<sup>۳</sup>-Average geometric distance degree

<sup>۴</sup>-Molecule Sepresentation of Structure based on Electron diffraction

## ۲-۲-۳-۴-۶-توصیف کننده های WHIM<sup>۱</sup>

این توصیف کننده ها براساس اندیس های آماری هستند که با تجسم اتمها در محورهای اصلی محاسبه می شوند. توصیف کننده های WHIM با کسب اطلاعات سه بعدی مولکول مربوط به اندازه، تقارن، و توزیع اتمی مولکول، با حفظ قالب های ثابت مرجع ساخته می شوند. الگوریتم آن متشکل از انجام آنالیز اجزای اصلی روی مختصات کارترین مرکزی یک مولکول، با استفاده از یک ماتریس کوواریانسی وزن شده حاصل از طرح های وزن شده متفاوت برای اتمها، می باشد.

$$S_{ik} = \frac{\sum_{i=1}^A w_i (q_{ig} - \bar{q}_g)^A (q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad (1-2)$$

در رابطه ی (۱-۲)،  $S_{ik}$  کوواریانس وزن شده بین  $g$  و  $k$  امین کوئوردینه های اتمی،  $A$  تعداد اتم ها،  $W_i$  وزن  $i$  امین اتم،  $q_{ig}$  و  $q_{ik}$  به ترتیب  $g$  امین و  $k$  امین کوئوردینه های  $i$  امین اتم، و  $\bar{q}$  مقدار میانگین مربوطه را نشان می دهند. شش طرح وزن شده پیشنهاد شده اند:

۱- موقعیت وزن نشده (u) ۲- جرم اتمی (m) ۳- حجم واندروالسی (v) ۴- الکترونگاتیویته اتمی ساندرسون (e) ۵- قطبش پذیری اتمی (p) ۶- حالت الکتروتوپولوژیکی کبیر و هال<sup>۲</sup> (S)  
این توصیف کننده ها به دو دسته اصلی توصیف کننده های WHIM مسیری، و توصیف کننده های WHIM<sup>۱</sup> گروه تقسیم می شوند [۴۶-۴۹].

## ۲-۲-۳-۴-۷-توصیف کننده های GETAWAY<sup>۳</sup>

<sup>۱</sup>- Weighted Holistic Invariant Molecular descriptors

<sup>۲</sup>-Kier-hall

<sup>۳</sup>-Geometry,topology,and atom weights assembly (GETAWAY)

این توصیف کننده ها براساس یک ماتریس قدرت نفوذ، مشابه آنچه در آمار معرفی شده است و معمولاً برای رگرسیون تشخیصی بکار می رود، تعریف شده اند. این ماتریس قدرت نفوذی، ماتریس نفوذ مولکولی (MIM) نامیده می شود که از مختصات اتمهای مولکول در یک صورتبندی مشخص محاسبه می شود. توصیف کننده های GETAWAY به دو زیر مجموعه H-GETAWAY ها و R-GETAWAY ها تقسیم می شوند. دسته ی اول

با استفاده از اطلاعات حاصل از ماتریس نفوذ مولکولی محاسبه می شوند. دسته دوم اطلاعات را با فواصل هندسی بین اتمی در مولکول ترکیب می کنند [۵۲-۵۰].

۲۸

## ۲-۲-۴- انتخاب بهترین توصیف کننده ها

انتخاب بهترین توصیف کننده ها که بیشترین وابستگی را با خاصیت مورد بررسی دارند از مهمترین مراحل QSPR است. وجود توصیف کننده های زیاد و همچنین وجود همبستگی میان آنها محاسبات و مدل سازی را دشوار و پیچیده می کند. لذا باید توصیف گرهایی انتخاب شوند که با خاصیت مورد بررسی بیشترین وابستگی را داشته و تغییرات آن را به خوبی توصیف کنند. یکی از روش های انتخاب توصیف کننده روش برازش مرحله ای<sup>۲</sup> است که با محاسبه ضرایب همبستگی<sup>۳</sup> مشخص می شود. ضریب همبستگی میزان نزدیکی اطلاعات بدست آمده به خط رگرسیون را نشان می دهد. مقدار این ضریب بین صفر و یک است. صفر یعنی هیچ ارتباطی بین توصیف کننده و خاصیت وجود ندارد و یک یعنی یک ارتباط کامل وجود دارد .

---

<sup>۱</sup>- Leverage matrix

<sup>۲</sup>-Stepwise

<sup>۳</sup>-Intercorrelation coefficient



## ۲-۲-۵-مدل سازی

پس از ارزیابی توصیف کننده ها و انتخاب مناسب ترین آنها، باید مدلی طراحی کنیم که بیانگر ارتباط خاصیت مورد بررسی و توصیف کننده های برگزیده باشد و بتواند خاصیت را برای سایر ترکیبات پیش بینی کند. روش های بکار رفته در مدل سازی به دو دسته زیر تقسیم می شوند:

### الف-استفاده از روش های آماری چند متغیره

در این روش از کالیبراسیون چند متغیره نظیر رگرسیون خطی چند گانه و روش حداقل مربعات جزئی استفاده می شود. این روش ها<sup>۲۹</sup> خطی بوده و زمانی که داده ها رفتار خطی دارند، کارایی دارند.

### ب-استفاده از روش های غیر خطی

برای کم کردن انحراف بین مقادیر پیش بینی شده و مقادیر صحیح از یک برآزش غیر خطی در مدل سازی استفاده می شود. از میان روش های غیر خطی روش شبکه عصبی مصنوعی در مطالعات QSPR و QSAR بکار گرفته شده است.

## ۲-۲-۶-رگرسیون خطی چندگانه (MLR)

رگرسیون خطی چندگانه روشی است که برای ایجاد ارتباط خطی بین متغیر وابسته و یک یا چند متغیر مستقل بکار می رود. متغیر مستقل گاهی پیش بینی شونده و متغیرهای وابسته پیش بینی کننده نامیده می شوند. برای مثال روش MLR برای پیش بینی زمان بازداری ترکیبات مختلف که با تکنیک های مختلف کروماتوگرافی جداسازی شده اند مورد استفاده قرار می گیرد. در این

صورت اندیس بازداري تركيبات مورد بررسی متغير وابسته، و توصيف کننده های ساختاری تركيبات، متغيرهای مستقل هستند. داده ها به دو دسته ی آموزش و پيش بينی تقسيم می شوند. با استفاده از سری آموزش مدل طراحی می شود. برای تعداد  $n$  مشاهده تجربی و تعداد  $p$  متغير مستقل، مدل رگرسيون خطی چندگانه توسط معادله (۲-۲) نمایش داده می شود:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2-2)$$

که در معادله (۲-۲)،  $\beta_0, \beta_1, \dots, \beta_p$  ضرایب رگرسيون<sup>۱</sup> و  $x_1, x_2, \dots, x_p$  متغيرهای مستقل (توصيف کننده های عددی) می باشند و  $y_i$  متغير وابسته هورد نظر (برای مثال زمان بازداري) می باشد. طی فرایند بدست آوردن یا محاسبه مدل مناسب، فاکتورهای آماری محاسبه می شوند که صحت مدل رگرسيون برای سری آموزش را نشان می دهند. این فاکتورها شامل ضريب همبستگی<sup>۲</sup>  $(R)$ ،  $F$ ،  $t$  و ضريب همبستگی میان توصيف کننده ها می باشد.

آماره  $F$ : آزمون  $F$  یا آزمون فیشر در واقع آزمون معنی دار بودن آماری در تحلیل رگرسيون ساده و چند متغيره، مشابه آزمون های معنی دار بودن آماری در تحلیل واریانس است [۵۳] نسبت  $F$  از رابطه ۲-۳ بدست می آید:

$$F = \frac{\sum \frac{(\hat{y}_i - \bar{y})^2}{p}}{\sum \frac{(y_i - \bar{y})^2}{n - p - 1}} \quad (3-2)$$

<sup>۱</sup> - Regression coefficient

<sup>۲</sup> - Correlation coefficient

در رابطه ی (۲-۳)، مقدار واقعی متغیر وابسته،  $y_i$  مقدار محاسبه شده ی آن توسط مدل خطی و  $\bar{y}$  میانگین مقادیر تجربی متغیر وابسته می باشد.

آماره R: ضریب همبستگی یا ضریب تعیین بین مقدار تجربی و پیش بینی شده برای متغیر مستقل را نشان می دهد. مقدار این آماره بین صفر و یک متغیر است. مقدار کوچک آن نشان دهنده این مسئله است که ارتباط خطی اندکی بین متغیر وابسته و متغیرهای مستقل وجود دارد یا چنین رابطه ای وجود ندارد.

آماره  $R^2$ : مجذور ضریب همبستگی چند متغیره که مبین واریانس مشترک بین متغیرهای وابسته و مستقل است. به آن ضریب تعیین نیز می گویند، چون نسبتی از واریانس متغیر وابسته را که توسط متغیرهای مستقل به حساب آمده معین می کند [۵۳].

آماره t: آزمون معنی دار بودن حضور هر متغیر وابسته در معادله رگرسیون را بیان می کند. مقدار t از تقسیم مقدار ضریب هر متغیر مستقل بر خطای استاندارد آن بدست می آید. این مقدار برای کلیه متغیرهای موجود در معادله رگرسیون و خارج آن محاسبه می شود [۵۳].

ضریب همبستگی میان توصیف کننده ها: یا همبستگی ((پیرسون))<sup>۱</sup> اندازه ای از ارتباط خطی بین دو متغیر است. مقادیر ضریب همبستگی بین -۱ تا +۱ متغیر هستند که علامت آن نشان دهنده جهت ارتباط خطی (مستقیم (+) یا عکس

---

<sup>۱</sup> -Pearson

بودن (-) آن) و قدر مطلق آن بزرگی این همبستگی را نشان می دهد [۵۳]. در ادامه به مراحل انجام رگرسیون خطی پرداخته شده است.

## ۲-۶-۱- آنالیز مدل‌های آماری و انتخاب مدل مناسب

مدل در حقیقت یک رابطه ریاضی است که معرف روابط بین متغیر وابسته و متغیر یا متغیرهای مستقل است و با استفاده از روشهای آماری گوناگون، می توان مدل مورد نظر را بدست آورد. برای این منظور در روش MLR چندین روش رگرسیون مختلف وجود دارد که عبارتند از:

### الف) روش ورود اجباری<sup>۱</sup>

فرایندی برای انتخاب متغیرها است که تمامی آن توصیف کننده ها در یک مرحله وارد می شوند. از آنجا که در این روش اثرات مثبت و منفی متغیرهای مستقل برای ورود در مدل نادیده گرفته می شوند روش مطلوبی نیست.

### ب) روش انتخاب جلوبرنده<sup>۲</sup>

تکنیکی مرحله ای برای انتخاب متغیرها است که طی آن متغیرها یکی پس از دیگری وارد مدل می شوند. اولین متغیر که برای ورود به مدل انتخاب می شود دارای بالاترین همبستگی مثبت یا منفی با متغیر وابسته است. این روند تا زمانی که دیگر متغیری وجود نداشته باشد که دارای شرط ورود باشد، ادامه دارد.

---

<sup>۱</sup> -Enter

<sup>۲</sup> -forward

### ج) روش عقب برنده<sup>۱</sup>

در این تکنیک تمام توصیف کننده ها وارد معادله می شوند و سپس متغیر با کوچکترین همبستگی جزئی با متغیر وابسته حذف می شود. پس از اینکه اولین متغیر حذف شد، از بین متغیرهای باقی مانده در معادله، متغیری که کوچکترین همبستگی را با متغیر وابسته دارد حذف می شود. این فرایند زمانی که متغیری با کمترین همبستگی وجود نداشته باشد متوقف خواهد شد.

### د) روش رگرسیون مرحله ای<sup>۲</sup>

این روش مشابه روش انتخاب جلو برنده است با این تفاوت که در این روش در هر مرحله برای متغیرهای وارد شده در مدل، آماره  $F$  محاسبه می شود و متغیر با کوچکترین آماره  $F$  بدون معنی حذف خواهد شد و مدل مناسب به دست می آید.

### ه) روش حذفی<sup>۳</sup>

در این تکنیک تمام توصیف کننده های یک بلوک در یک مرحله حذف می شوند [۵۴].

## ۲-۷- شبکه های عصبی مصنوعی

---

<sup>۱</sup> -backward

<sup>۲</sup> -Stepwise

<sup>۳</sup> -Remove

در دهه های اخیر شیمیدانان هم زمان با استفاده از کامپیوتر و انجام روشهای آماری پیچیده، برای کشف همبستگی های چند متغیری بین خروجی و متغیرهای ورودی با جزئیات بیشتر در تلاش بوده اند. با افزایش صحت و دقت روش های اندازه گیری تجزیه ای واضح شده است که همه اثرهایی که مورد نظر هستند بوسیله رگرسیون های ساده و چند متغیره خطی قابل توصیف نیستند. دسته ای از روشها که مورد استفاده و استقبال زیادی توسط شیمیدانان واقع شده اند، شبکه های عصبی مصنوعی هستند.

مفهوم استفاده از شبکه های عصبی در حل مسایل منطقی در سال ۱۹۴۰ بوجود آمد. در سال ۱۹۴۳، ((مک کولوچ))<sup>۱</sup> و ((پیتز))<sup>۲</sup> مقاله ای در مورد روشی جدید

که امروزه تحت عنوان شبکه های عصبی از آن یاد می شود، به چاپ رساندند [۵۵]. پس از آن در سال ۱۹۴۹، هب<sup>۳</sup>، روانشناسی در دانشگاه مک گیل، یک قاعده یادگیری قابل توجیه را (که امروزه به قاعده هبمان معروف است) منتشر کرد و بعضی از جنبه های تئوری الگوریتم های آموزشی برای شبکه عصبی را شکل داد [۵۶]. از آن زمان تا اواسط سال ۱۹۸۰ روش محاسبات عصبی به دلیل نتایج ضعیف آن در حل مسایل تجربی کمتر مورد توجه قرار گرفت. استفاده از شبکه های عصبی از سال ۱۹۸۶ و پس از انتشار کتابی توسط رامل هارت<sup>۴</sup> و مک کللاند<sup>۵</sup> به طور چشمگیری افزایش یافت [۵۶]. پس از آن زمان کاربردهای شبکه عصبی مصنوعی در شیمی و علوم وابسته به آن مانند بیوشیمی،

---

<sup>۱</sup> -McCulloch

۴ -Pitts

<sup>۱</sup> -McCulloch

<sup>۲</sup> -Pitts

<sup>۳</sup> -Hebb

<sup>۴</sup> -Rumelhart

<sup>۵</sup> -McClelland

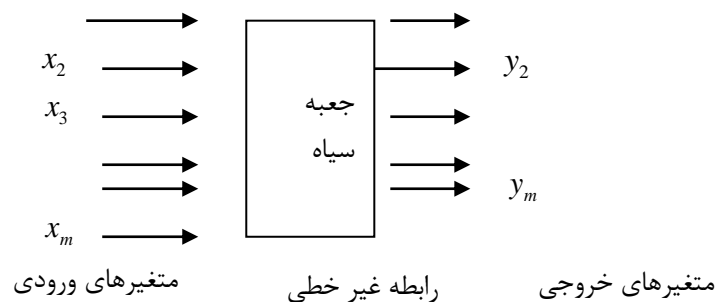
مهندسی شیمی و داروسازی رواج پیدا کرد [۵۷]. شبکه های عصبی چه در بعد آنالیز و توسعه و چه در بعد پیاده سازی سخت افزاری از نظر کمی و کیفی و توانایی، در حال رشد و پیشرفت می باشند و تکنیک های مختلف محاسبات عصبی از لحاظ تعداد در حال افزایش است [۵۸].

## ۲-۲-۷-۲- ماهیت شبکه عصبی مصنوعی

شبکه عصبی مصنوعی یک الگوی پردازش اطلاعات است که از سیستم های عصبی بیولوژیکی مانند مغز الهام می گیرد. این الگو تشکیل شده از تعداد زیادی از عناصر پردازش کننده که به هم متصل هستند و بطور هماهنگ برای

حل مسائل ویژه، عمل می کنند. شبکه های عصبی مصنوعی مانند انسان ها با مثال آموزش می بینند. آموزش در سیستمهای بیولوژیکی درگیر تنظیمات در اتصالات سیناپسی است. در شبکه عصبی مصنوعی نیز به همین صورت است. شبکه های عصبی مصنوعی به سادگی قابل توصیف نیستند. شاید نزدیکترین توصیف مقایسه آن با یک جعبه سیاه باشد که چندین ورودی و چندین خروجی دارد که اکثرا به طور موازی با واحدهای پردازشگر ارتباط دارند. مهمترین مسئله ای که درباره شبکه های عصبی باید به خاطر داشت این است که در صورتی که رابطه غیر خطی بین ورودی ها و خروجی ها را بررسی کنند بهترین روش هستند (شکل ۱-۲). به دلیل این واقعیت که در ابتدا از نوع رابطه بین پاسخ و ورودی ها اطلاعی نداریم بهتر است برای تفسیر همسویی اطلاعات در شبکه عصبی، همیشه از روش های آماری استاندارد استفاده کنیم. این روش برای

محاسبه رابطه خطی بین ورودی‌ها و خروجی‌ها نیز می‌تواند بکار رود ولی نتیجه آن، به خوبی نتایج حاصل از روشهای آماری ساده نمی‌باشد [۵۹].



شکل ۱-۲ شبکه عصبی مصنوعی به عنوان یک جعبه سیاه با در نظر گرفتن رابطه غیر خطی بین متغیرهای چندتایی ورودی و متغیرهای خروجی

## ۲-۷-۳- دلیل استفاده از شبکه های عصبی مصنوعی

شبکه های عصبی با توانایی چشمگیرشان در استخراج مفاهیم از اطلاعات پیچیده یا مبهم، می‌توانند به منظور استخراج مدلها و شناسایی نظامهایی که برای تشخیص توسط انسانها یا تکنیک های کامپیوتری بسیار پیچیده هستند، بکار روند. از جمله محاسن شبکه عصبی مصنوعی می‌توان به موارد زیر اشاره کرد:

۱- یادگیری تطبیقی: شبکه توانایی یادگیری چگونگی انجام ماموریت های سخت براساس اطلاعات داده شده برای آموزش یا آزمایش اولیه را دارد.

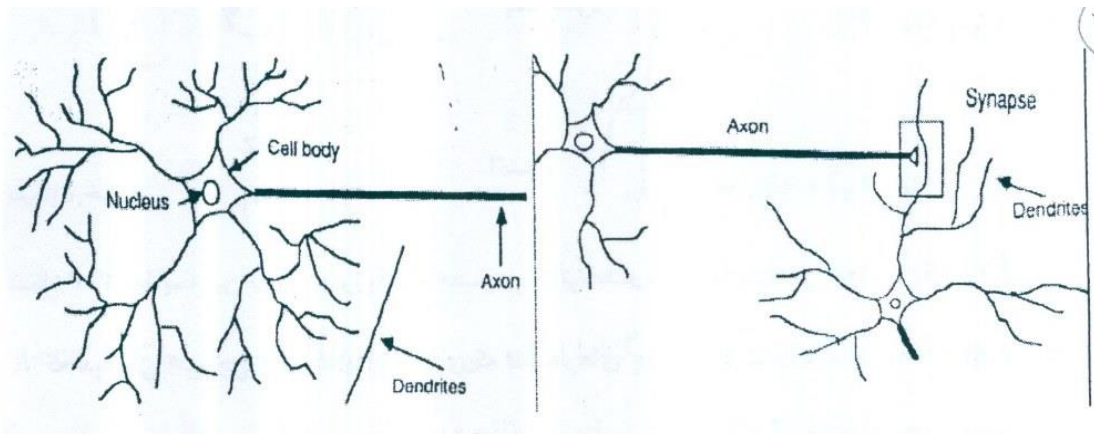
۲- خودسازمان دهی: یک شبکه عصبی مصنوعی می‌تواند سازماندهی یا ارائه اطلاعات خود را که در حین یادگیری خود کسب کرده، خلق کند [۶۰].



۳ — عملکرد موازی: محاسبات شبکه عصبی مصنوعی ممکن است بطور همسو انجام گیرد. وسایل سخت افزاری ویژه ای طراحی و تولید شدند که مزیت این قابلیت را تفسیر می کنند [۶۰].

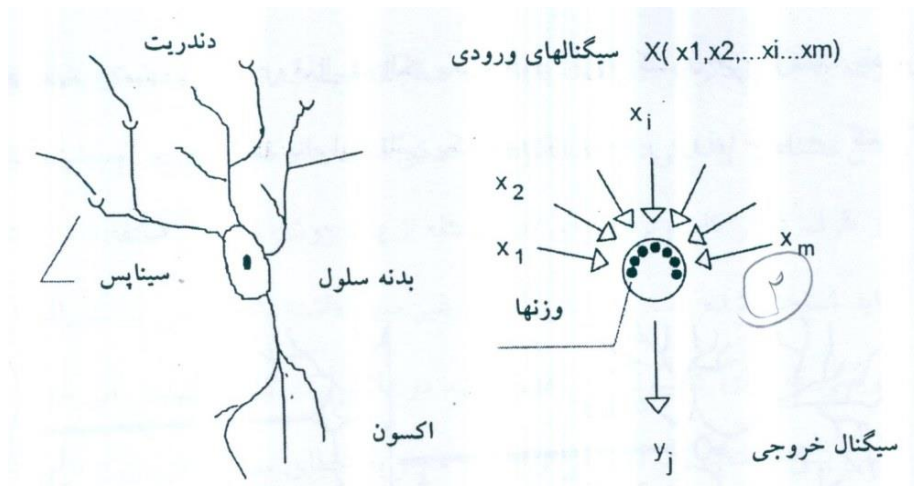
## ۲-۲-۷-۴- نرون بیولوژیکی

خیلی مسائل هنوز در مورد اینکه چگونه مغز خودش را آموزش می دهد تا اطلاعات را پردازش کند، نا شناخته است. در مغز انسان یک نرون سیگنالها را از سایر نرونها توسط بخشی از ساختار خود به نام دندریت جمع آوری می کند. نرون سیگنال های الکتریکی را از طریق جزء بلند و باریکی به نام آکسون انتقال می دهد که در انتها به صدها شاخه تقسیم می شود. در انتهای هر شاخه ساختاری به نام سیناپس وجود دارد که فعالیت را از اکسون به سیگنالهای های الکتریکی تبدیل می کند که فعالیت را در سایر نرونها تقویت یا متوقف می کند. وقتی نرون ورودی تقویت کننده دریافت می کند؛ با آستانه ورودی مقایسه شده و سپس پیکهای فعالیت الکتریکی خود را به اکسون می فرستد. یادگیری از طریق تغییر تاثیر سیناپسها رخ می دهد بنابراین تاثیر یک نرون روی سایر نرونها تغییر می کند. شکل (۲-۲) یک نرون عصبی را نشان می دهد [۵۹].



شکل ۲-۲ اجزای تشکیل دهنده نرون (سمت چپ) ساختار سیناپس (سمت راست) [۵۸]

نرون مصنوعی همانند یک نرون بیولوژیکی، تعداد زیادی سیگنال  $x_i$  را از نرون های مجاور دریافت کرده و آنها را با روشی از پیش تعیین شده پردازش می کند. بسته به خروجی این پردازش، نرون  $g$  در مورد تولید خروجی  $y_g$  تصمیم می گیرد. سیگنال خروجی (در صورتی که ایجاد شود) بسته به اینکه نرونهای دوتایی یا مقدار حقیقی بکار ببریم، به ترتیب می تواند صفر یا یک باشد، یا می تواند هر مقدار حقیقی بین صفر و یک داشته باشد (شکل ۲-۳).



شکل ۲-۳ مقایسه بین نرون بیولوژیکی و مصنوعی. دایره بدنه سلول را نشان می دهد که یک فرایند ریاضی ساده را که سیگنالی خروجی از سیگنال های ورودی که با بردارهای چندتایی  $X$  نشان داده شده اند، تولید می کند [۵۸].

ابتدا ترکیبی خطی از متغیرهای ورودی  $x_1, x_2, \dots, x_i, \dots, x_m$  در ضریب های  $w_{gi}$  به نام وزن ها ضرب می شوند. مجموع ورودی به همراه وزن ها از تابع انتقال عبور کرده و خروجی را بوجود می آورند. خروجی  $y_i$  روی  $g$  امین نرون به صورت زیر محاسبه می شود:

ابتدا ورودی شبکه براساس معادله ۲-۴ محاسبه می شود.

$$Net_g = \sum_{i=1}^m w_{gi} x_i \quad (۴-۲)$$

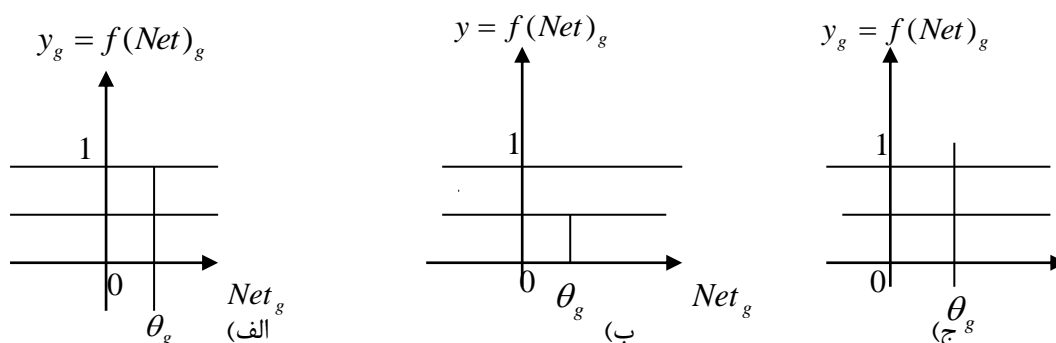
و سپس  $Net_g$  به عنوان یک شناسه در تابع انتقال (معادله ۲-۵) وارد می شود.

$$y_g = out_g = 1 / \{1 + \exp[-a_g (Net_g + \theta_g)]\} \quad (۵-۲)$$

وزنها ( $w_{gi}$ ) در نرون مصنوعی هم ارزشهای قدرت سیناپس در نرون های بیولوژیکی، بین اکسون ها که سیگنال ها را ایجاد می کنند و دندریت ها که سیگنال ها را دریافت می کنند، هستند. هر قدرت سیناپس بین اکسون و دندریت (به عبارتی هر وزن) تعیین می کند چه نسبتی از سیگنال های ورودی به درون بدنه سلول عبور کند [۵۹].

## ۲-۲-۷-۵- تابع انتقال

اثبات شده که اطلاعات در مغز توسط تعدیل کردن ثابت سیناپس ها برای سیگنال های ورودی های متفاوت، خروجی های بهتری ایجاد می کنند، که همان سیگنال هایی هستند که باید جواب یا واکنش مناسب بدن را ایجاد کنند. نتایج به عنوان ورودی های جدید دائماً برگشتی هستند. مشابه مغز، نرون مصنوعی، تعدیل قدرت سیناپس ها را با تعدیل مکرر وزن ها  $w_{ig}$  در نرون ها براساس تفاوت بین خروجی نرون ( $y_i$ ) و مقدار واقعی آن ( $t_g$ ) تقلید می کند. اگر تابع خروجی  $f(Net_g)$  یک تابع آستانه ای دوتایی باشد (شکل ۲-۴ الف) خروجی نرون تنها دو مقدار صفر یا یک خواهد داشت. با این حال در اکثر موارد کاربردی شبکه عصبی مصنوعی تشکیل شده از نرونهایی است که تابع انتقال سیگموئیدی دارند (شکل ۲-۴ ج، معادله ۲-۵). در بعضی موارد، تابع انتقال در نرون مصنوعی می تواند حالت رادیال<sup>۱</sup> داشته باشد.



شکل ۲-۴ سه تابع انتقال مختلف: توابع آستانه ای (الف) سیگموئیدی (ب) و رادیال (ج) پارامتر در هر سه تابع

مقدار  $Net_g$  در محدوده ای که نرون بیشترین گزینش پذیری را دارد، تعیین می کند.

پارامتر دیگر  $a_g$  که در معادلات ۲-۴ و ۲-۵ دیده می شود، روی شیب تابع انتقال تاثیر می گذارد.

<sup>۱</sup> -Radial

دانستن نوع تابع مورد استفاده در تمام نرون های شبکه، بدون در نظر گرفتن مکان آنها و چگونگی ارتباط آنها با سایر نرون ها، اهمیت دارد. آنچه که در طول یادگیری یا آموزش تغییر می کند تابع نیست، بلکه وزن ها و پارامترهای تابع هستند که موقعیت مقدار آستانه و شیب تابع انتقال را کنترل می کنند (معادلات ۲-۵ و ۲-۶).

$$y_i = out_g = e^{-[a_g (Net_g + \theta_g)]^2} \quad (۲-۶)$$

می توان نشان داد که دو پارامتر مقدار آستانه ( $\theta_g$ ) و شیب تابع انتقال  $a_g$  در تمام شبکه های عصبی که چنین تابع انتقالی هستند، طی فرایند آموزش به همان صورتی که وزنها آموزش می بینند، تصحیح می شوند. دلیل این ادعا در زیر به صورت ریاضی نشان داده شده است.

ابتدا باید مقدار قراردادی تابع  $f(u_g)$  (معادله ۲-۶) را به شکلی بسط یافته با وارد کردن معادله ۲-۴ بنویسیم:

$$u_g = a_g (Net_g + \theta_g) = \sum_{i=1}^m a_g w_{gi} x_i + a_g \theta_g \quad (۲-۷)$$

$$y_g = out_g = a_g w_{g1} x_1 + a_g w_{g2} x_2 + \dots + a_g w_{gi} x_i + \dots + a_g w_{gm} x_m + a_g \theta_g \quad (۲-۸)$$

در شروع فرایند آموزش هیچ کدام از ثابت های  $a_g, \theta_g$  یا  $w_{gi}$  معلوم نیستند. بنابراین ایرادی ندارد که پاسخ هایی مانند  $a_g w_{gi}, a_g \theta_g$  به سادگی مانند پاسخ های یک متغیره و یا دو متغیره نوشته شوند. بنابراین با نوشتن معادله دو مجهولی به صورت معادله ای با یک پارامتر مجهول معادله ۲-۸ به صورت ۲-۹ بدست خواهد آمد:

$$u_i = w_{g1} x_1 + w_{g2} x_2 + \dots + w_{gi} x_i + \dots + w_{gm} x_m + w_{gm} + 1 \quad (۲-۹)$$

حال با اضافه کردن یک متغیر  $x_{m+1}$  به تمام بردارهای ورودی  $x$  بردار تقویت شده ای به صورت  $x' = (x_1, x_2, \dots, x_m, x_{m+1})$  بدست خواهیم آورد. سپس متغیر اضافه شده برای تمام ورودی ها

بدون استثناء، برابر یک تنظیم می شود؛ بنابراین بردار تقویت شده  $x'$  را می توان به صورت  
 $(x_1, x_2, \dots, x_{i \dots x_m, 1})$  نوشت.

دلیل این فرض ساده است. ما می خواهیم آخرین وزن  $w_{gm} = 1$  ایجاد شده از طرح  $a_g \theta_g$  را به یک  $Net_g$  جدید و یگانه، شامل تمام پارامترها (وزن ها، آستانه، و شیب) برای تعدیل (یادگیری) نرون به همان شکل، ترکیب کنیم. تنها چیزی که برای بدست آوردن و بکار بردن اختلاف خروجی نرون و مقدار واقعی، لازم است اضافه کردن یک جزء مساوی یک به هر بردار ورودی است.

این متغیر ورودی اضافه شده که غالباً مقدار یک دارد، بایاس نامیده می شود. بایاس نرون را انعطاف پذیرتر و انطباق پذیرتر می کند. اکنون معادله برای بدست آوردن سیگنال خروجی از یک نرون خاص را از ترکیب معادله ۹-۲ بسته به اولویت ما برای شکل دهی تابع انتقال با معادله ۵-۲ یا ۶-۲ به صورت زیر بدست می آید:

$$y_i = out_g = 1/[1 + \exp(-Net_g)] \quad (۱۰-۲)$$

$$y_i = out_g = \exp(-Net_g^2) \quad (۱۱-۲)$$

$$Net_g = \sum_{i=1}^{m+1} w_{gi} x_i \quad (۱۲-۲)$$

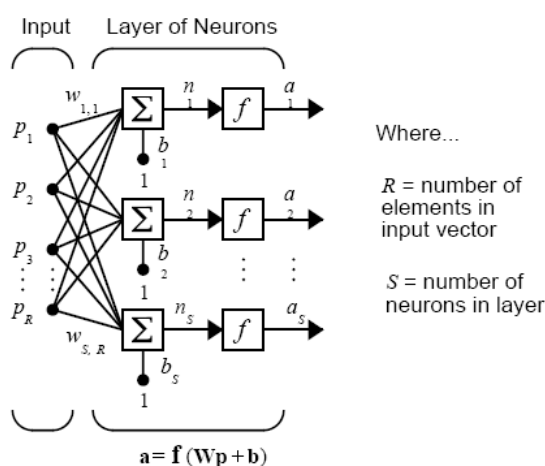
توجه داشته باشید که تنها تفاوت میان معادلات (۱۲-۲) و (۴-۲) در علامت جمع در معادلات (۱۲-۲) بین ۱ تا  $m+1$  به جای ۱ تا  $m$  در (۴-۲) است. اکنون که مقداری با نرون های مصنوعی و عملکرد آن با سیگنالهای خروجی، آشنا شدیم می توانیم نرونهای جمعی را مورد بررسی قرار دهیم. تعداد زیادی از سری های نرونها که اکثر آنها به همدیگر مرتبط هستند، شبکه عصبی نامیده می شود. اگر نرونهای مجزا با کامپیوترها شبیه سازی می شوند بنابراین دسته ای از نرون ها هم می توانند به همان خوبی شبیه سازی شوند. دسته های نرونها که بوسیله کامپیوترها شبیه سازی می شوند، شبکه های عصبی مصنوعی نامیده می شوند.

## ۲-۲-۷-۶- ساختار شبکه ی عصبی مصنوعی

شبکه های عصبی مصنوعی می توانند از تعداد زیادی نرون تشکیل شوند. نرون ها می توانند در ساختار لایه ای سازماندهی شوند. در ادامه انواع شبکه ی تک لایه و چند لایه شرح داده خواهد شد.

### ۲-۲-۷-۶-۱- شبکه ی تک لایه

در شکل ۲-۵ یک شبکه ی تک لایه با R عنصر ورودی و S نرون نشان داده شده است.



شکل ۲-۵- یک شبکه ی تک لایه ای

در این شبکه هر عنصر از بردار ورودی p به هر نرون ورودی از طریق ماتریس وزنی W متصل می شود. I امین نرون یک جمع کننده ( $\sum$ ) دارد که ورودی های وزن دار شده و بایاس را برای تشکیل خروجی عددی خودش، n (i) جمع آوری می کند. n های مختلف در کنار یکدیگر یک بردار ورودی شبکه S عنصری بردار n را می دهند. سرانجام لایه ی خروجی نرون یک بردار ستونی a را تشکیل می دهد. عناصر بردار ورودی در شبکه از میان ماتریس وزنی W وارد می شوند.

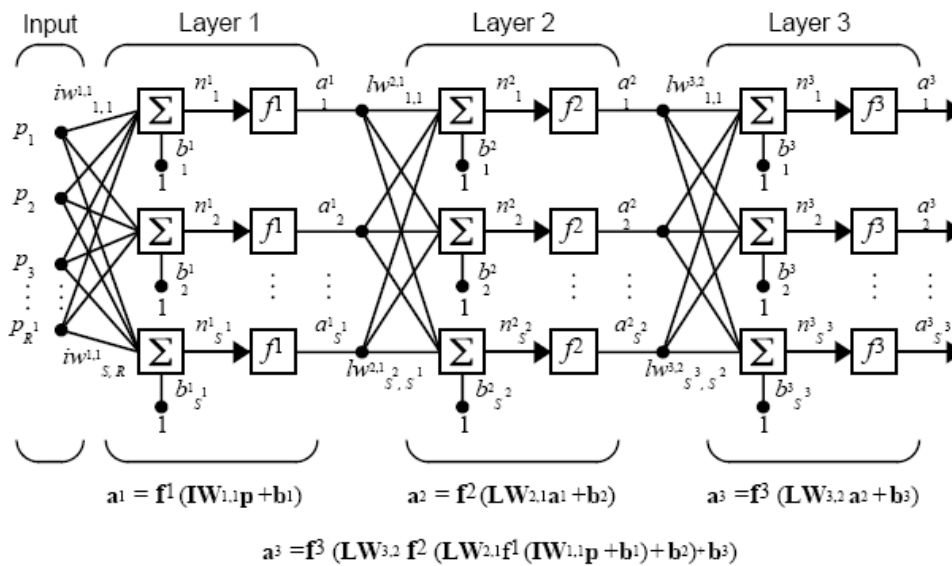
$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,R} \\ w_{2,1} & w_{2,2} & \dots & w_{2,R} \\ \dots & \dots & \dots & \dots \\ w_{S,1} & w_{S,2} & \dots & w_{S,R} \end{bmatrix} \quad (13-2)$$

در معادله ی (۱۳-۲)، اندیس  $i$  در نماد  $w_{i,j}$  نشان دهنده ی نرون مقصد و اندیس  $j$  نشان دهنده ی منبع ورودی است (شکل ۲-۵).

## ۲-۲-۷-۶-۲- شبکه ی چند لایه

یک شبکه می تواند چندین لایه داشته باشد. هر لایه شامل ماتریس وزنی  $w$ ، یک بردار بایاس  $b$  و یک بردار خروجی  $a$  می باشد. شبکه ی نشان داده شده در شکل (۶-۲) دارای  $R$  عنصر ورودی،  $S_1$  نرون در لایه ی اول،  $S_2$  نرون در لایه ی دوم و ... می باشد. لایه های مختلف تعداد متفاوتی نرون می توانند داشته باشند. یک ورودی ثابت هم می تواند به عنوان بایاس به هر نرون اضافه شود. لایه ای که خروجی را محاسبه می کند لایه ی خروجی نامیده می شود. شبکه ی سه لایه ای نشان داده شده در شکل ۶-۲ دارای یک لایه ی خروجی و دو لایه ی مخفی می باشد. شبکه های چند لایه کاملاً قدرتمند هستند [۶۱].





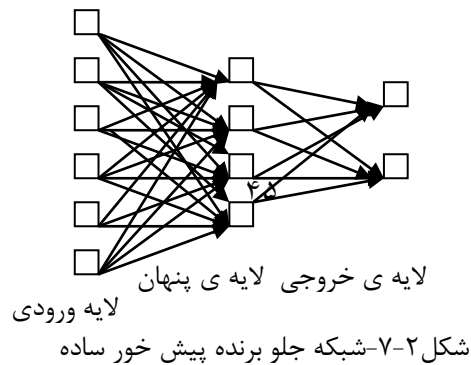
شکل ۲-۶- یک شبکه ی سه لایه ای

## ۲-۲-۷-۷- انواع شبکه های عصبی

### ۲-۲-۷-۷-۱- شبکه های جلوبرنده (پیش خور)<sup>۱</sup>

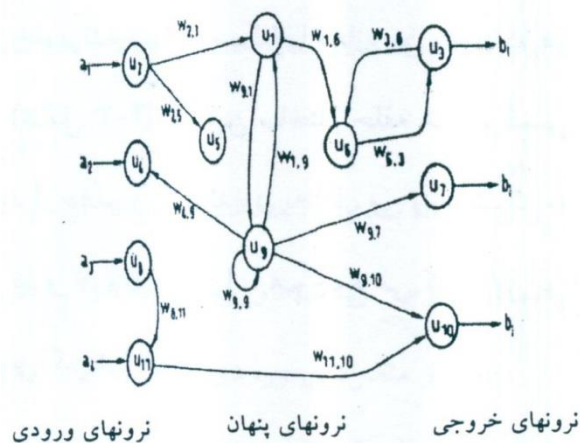
شبکه های عصبی جلوبرنده تنها به سیگنال ها اجازه می دهند که در یک مسیر، از ورودی به سمت خروجی حرکت کنند (شکل ۲-۷). در این ساختار حلقه های برگشتی که طی آن خروجی هر لایه، آن لایه را تحت تاثیر قرار دهد، وجود ندارد. به نظر می رسد شبکه های جلوبرنده شبکه های مستقیمی هستند که ورودی ها و خروجی ها را با هم جمع می کنند.

<sup>۱</sup> -Feed forward



### ۲-۲-۷-۲- شبکه های برگشتی (پس خور)<sup>۱</sup>

شبکه های برگشتی سیگنالهایی دارند که از طریق حلقه ها در دو مسیر، حرکت می کنند. شبکه های عقب برنده بسیار قوی هستند و می توانند به شدت پیچیده باشند (شکل ۸-۲). این شبکه ها پویا هستند و موقعیت آنها به طور پیوسته تا رسیدن به نقطه تعادل در حال تغییر است. آنها در موقعیت تعادل باقی می مانند تا اینکه ورودی جدیدی وارد شود و نقطه تعادل دیگری مورد نیاز باشد. ساختار شبکه های برگشتی به متقابل یا برگشتی بودن آنها بر می گردد. به هر حال مورد دوم معمولاً برای مشخص کردن اتصالات عقب برنده در ساختارهای تک لایه ای بکار می رود [۶۰].



شکل ۸-۲- نمونه ای از یک شبکه پیچیده برگشتی [۶۰]

<sup>۱</sup> -Feed backward

## ۲-۲-۷-۸-طراحی شبکه

۴۶

۲-۲-۷-۸-۱-ایجاد شبکه

در این بخش نحوه ی طراحی یک شبکه ی پیش خور با الگوریتم پس انتشار خطا<sup>۱</sup> شرح داده خواهد شد. اولین مرحله ی ایجاد شبکه فراهم کردن اجزاء شبکه<sup>۲</sup> است. با استفاده از توابع خلق شبکه این کار انجام می شود. تابع newff یکی از توابع خلق شبکه است که یک شبکه ی پیش خور ایجاد می کند. این تابع نیازمند یک ماتریس از بردارهای ورودی R عنصری نمونه، یک ماتریس از بردار هدف S عنصری و آرایه ای که در بر گیرنده ی اندازه ی لایه ها (تعداد نرون در لایه) است، می باشد.

## ۲-۲-۷-۸-۲-تعیین مقدار اولیه ی وزن و بایاس

قبل از آموزش شبکه ابتدا باید به وزن ها و بایاس مقدار اولیه داده شود. تابع خلق شبکه به طور خودکار این کار را انجام می دهد. اگر بخواهیم به آنها دوباره مقدار اولیه دهیم، با دستور init<sup>۳</sup> این کار را انجام می دهیم.

## ۲-۲-۷-۸-۳-شبیه سازی

تابع sim شبکه را شبیه سازی می کند. این تابع ورودی و اجزای شبکه را می گیرد و خروجی را پس می دهد.

---

<sup>۱</sup> -Backpropagation

<sup>۲</sup> -Network object

<sup>۳</sup> Initialize

## ۲-۲-۷-۹-آموزش شبکه

آموزش شبکه فرایندی است که طی آن وزن ها و بایاس شبکه جهت محاسبه ی خروجی مطلوب تعیین می شوند. از جمله روش های موجود برای آموزش شبکه می توان به افزایشی<sup>۱</sup> و مجموعه ای<sup>۲</sup> اشاره کرد. در آموزش افزایشی وزن ها و بایاس شبکه در هر زمانی که یک ورودی به شبکه داده می شود تنظیم می شوند. در شیوه ی مجموعه ای وزن ها و بایاس ها بعد از اینکه همه ی ورودی ها و هدف ها ارائه شدند، بهینه می شوند.

## ۲-۲-۷-۱۰-قوائد یاد گیری

یک قائده ی یادگیری فرایندی برای اصلاح وزن ها و بایاس های شبکه است. این قائده به عنوان یک الگوریتم آموزشی در نظر گرفته می شود. یادگیری به دو روش با راهنما و بدون راهنما انجام می شود. الگوریتم پس انتشار یا خطا یک روش متداول یادگیری است که در ادامه شرح داده خواهد شد.

## ۲-۲-۷-۱۰-۱-الگوریتم پس انتشار

---

<sup>۱</sup> -Incremental

<sup>۲</sup> -Batch

پس انتشار خطا یک روش متداول یادگیری با راهنما است و برای شبکه های پیش خور مفید است. لازم به ذکر است که برای انجام هر فرایند آموزش باید سری داده ها را به سه دسته تقسیم کرد. اولین دسته را سری آموزش<sup>۱</sup> می گویند. دومین دسته پیش بینی<sup>۲</sup> و سومین دسته، سری تست<sup>۳</sup> است. هر کدام از این سه سری داده ها باید شامل تقریباً یک سوم کل داده ها باشند. سری آموزش باید بیشترین تعداد و سری تست کمترین تعداد را داشته باشند. پس انتشار خطا [۶۲و۶۳] حداقل به دو لایه از نرونها تحت عنوان لایه پنهان و لایه خروجی نیاز دارد. لایه ورودی غیر فعال است. بنابراین در شکل در نظر گرفته نمی شود. آموزش با استفاده از تکنیک پس انتشار خطا نیاز به راهنما دارد. یعنی باید یک سری از ورودی<sup>۴۸</sup> I و زوج هدف  $\{x_s, T_s\}$  داشته باشیم. بردار هدف شامل مقادیر تجربی خروجی شبکه است. الگوریتم پس انتشار خطا این نام را به خاطر فرایندی که طی آن اتفاق می افتد دریافت کرده است: وزن نرونها ابتدا در لایه خروجی تصحیح می شوند، سپس در لایه پنهان (در صورتی که یک لایه باشد) و در پایان در لایه پنهان اول که همان لایه ای است که سیگنال ها را مستقیماً از ورودی دریافت می کند.

دلیل تصحیح وزنها به این شیوه این است که خطای دقیق شبکه عصبی تنها در لایه خروجی معلوم است. پس از آن که  $n$  نرون لایه خروجی، مقدار  $y_i$  را ارائه دادند، می توانند به عنوان بردار چند بعدی در نظر گرفته شود.  $Y(y_1, y_2, \dots, y_{i...y_n})$  این پاسخ ها با مقدار واقعی  $t_i$  بردار هدف  $T_s$  که بردار ورودی XS را همراه می کند، مقایسه می شوند. در این روش، خطا روی هر گره خروجی  $i$  دقیقاً شناخته شده است.

$$\delta_{gi} = y_i - t_i \quad (۱۴-۲)$$

آموزش شبکه عصبی با به حداقل رساندن یک تابع خطا (E) با تغییروزنها ( $W_{mh}, W_{hi}$ ) بدست می آید.

---

<sup>۱</sup> -Training set

<sup>۲</sup> -Prediction set

<sup>۳</sup> -Test set

$$E = \sum_p E_p = \frac{1}{2} \sum_p \sum (a_{pm} - o_{pm})^2 \quad (15-2)$$

که  $E_p$  خطای P امین مسیر آموزش که با سری توصیف کننده ها و فعالیت مربوط به ترکیبات  $a_{pm}$  مربوط به مقدار تجربی m امین متغیر وابسته است. یکی از راه های استاندارد کاهش خطا قانون دلتا است [64-65]. الگوریتم این قانون براساس یک فرایند تکرار شونده برای تغییر وزنه های شبکه از مقدار قبلی آنهاست. نحوه تغییر وزنه توسط معادلات 16-2 الف و 16-2 ب نشان داده شده اند.

$$W_{mh}^{n+1} = W_{mh}^n - \eta \frac{E}{W_{hi}} \quad (16-2 \text{ الف})$$

۴۹

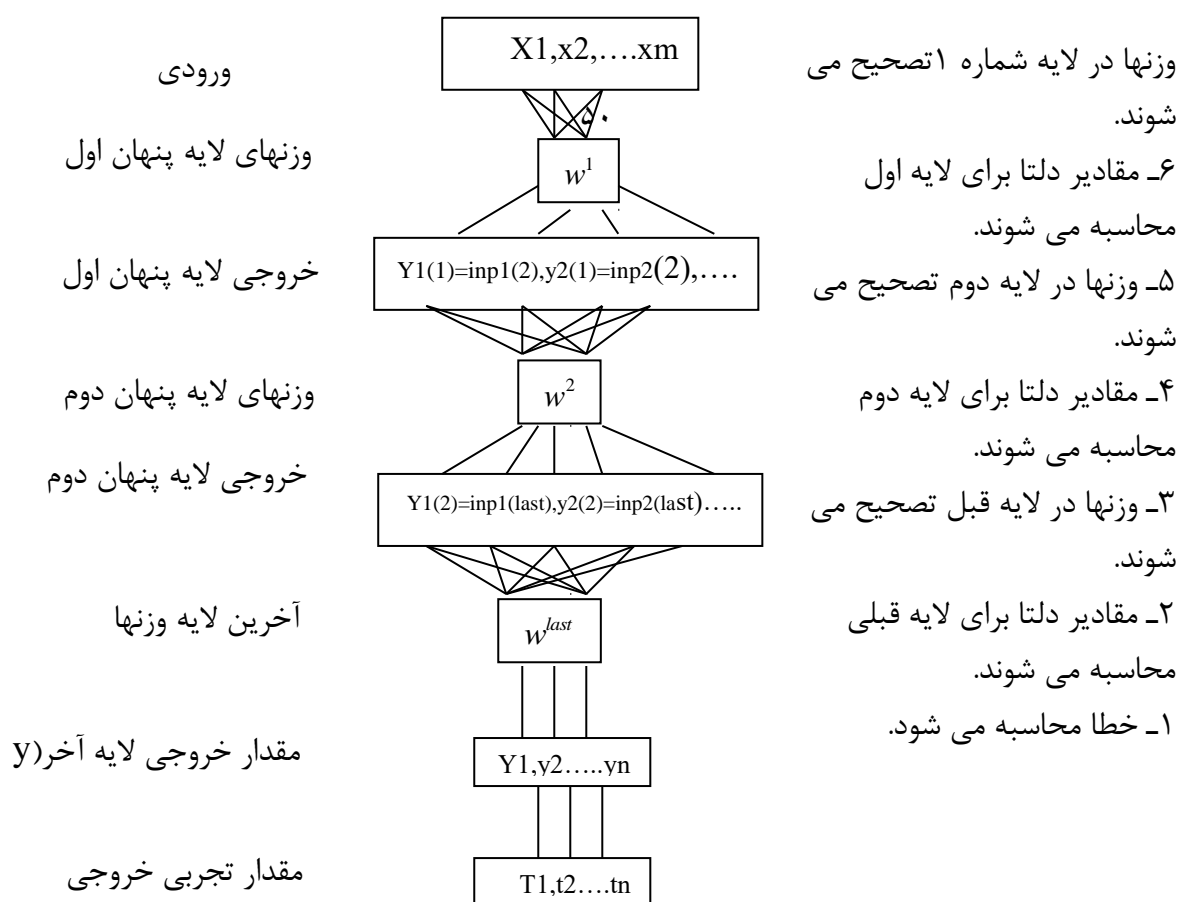
$$W_{hi}^{n+1} = W_{mh}^n - \eta \frac{E}{W_{hi}} \quad (16-2 \text{ ب})$$

در معادلات (16-2 الف) و (16-2 ب) بالا نویس های n نشان دهنده چرخه های پیاپی در فرایند کاهش خطا و  $\eta$  سرعت آموزش است. معادلات مشابهی برای بایاس بین لایه پنهان و لایه ورودی ( $\theta_h$ ) و بایاس بین لایه خروجی و لایه پنهان ( $\theta_m$ ) نیز بکار می رود. در شکل 9-2 مراتب تصحیح وزن ها طی آموزش با پس انتشار خطا نشان داده شده است.

اولین وزنه هایی که تصحیح می شوند، وزنه های لایه خروجی هستند و پس از تصحیح این وزنه ها با این فرض که خطا وقتی که سیگنال ها از آخرین لایه پنهان به لایه خروجی عبور می کنند به طور یکنواخت توزیع می شود، وزنه ها در لایه پنهان آخر تصحیح می شوند. آخرین وزنه هایی که تصحیح می شوند وزنه های لایه بالاتر هستند. به طریق مشابهی وزنه ها در تمام لایه های اضافه شده (به استثنای لایه خروجی) محاسبه می شوند. الگوریتم آموزش در توابع پس انتشار خطا در صورتی که ترم ممنتوم به عبارت تصحیح وزنه اضافه شود، به صورت زیر است:

$$\Delta w_{gi}^l = \eta \delta_i^l y_i^{l-1} + \mu \Delta w_{gi}^{l(previous)} \quad (17-2)$$

در رابطه ی (۲-۱۷) پارامترهای  $\mu$  و  $\eta$  سرعت آموزش و ممنتوم نامیده می شوند و معمولاً بین ۰/۱ تا ۰/۹ هستند. بالانویس (previous) نشان دهنده این است که برای تصحیح مناسب وزنها در کنار محاسبه خطاها ذخیره کردن وزنها از تصحیح قبلی و استفاده از آن برای تصحیح بعدی ضروری است. از طرف دیگر ترم  $\delta_g^l$  خطای موجود در نرون (گره)  $i$  ام را در لایه  $l$  ام نشان می دهد.



شکل ۲-۹ ترتیب تصحیح وزنها طی آموزش پس انتشار خطا از پایین به بالا است.  $i$  امین خروجی  $K$  امین لایه به صورت  $(k(y_i))$  نمایش داده شده است.

آموزش پس انتشار خطا به صورت چرخه ای انجام می گیرد که به آنها دوره گفته می شود. یک دوره، تناوبی است که طی آن زوج های ورودی - هدف  $\{x_s, T_s\}$  در شبکه وارد می شوند. وزنها بعد از اینکه

هر زوج  $\{x_s, T_s\}$  یک بردار خروجی  $Y_s$  تولید کرد و خطاها محاسبه شدند با هم جمع می شوند. پس از هر دوره، خطای RMS گزارش می شود:

$$RMS = \left[ \frac{\sum_{s=1}^r \sum_{i=1}^n (t_{si} - y_{si})^2}{rn} \right]^{1/2} \quad (18-2)$$

مقدار RMS تناسب خروجی ها را در مقایسه با مقادیر بردار هدف تخمین می زند. هدف آموزش با سرپرست دستیابی به کمترین مقدار ممکن برای RMS در کمترین زمان ممکن است.

۵۱

## ۲-۲-۷-۱۰-۲- الگوریتم لونیبرگ-مارکواردت<sup>۱</sup>

یکی از الگوریتم های پس انتشار خطا لونیبرگ-مارکواردت است که به شیوه ی مجموعه ای عمل می کند. این روش به منظور رسیدن به سرعت آموزش بالا بدون نیاز به محاسبه ی ماتریس هسیان<sup>۲</sup> طراحی شده است. تابع اجرای این الگوریتم مجموع مربعات خطای خروجی (MSE)<sup>۳</sup> شبکه است. ماتریس هسیان به صورت معادله ی ۲-۱۹ محاسبه می شود.

$$H = J^T J \quad (19-2)$$

گرادیان نیز با معادله ی ۲-۲۰ محاسبه می شود.

$$g = J^T e \quad (20-2)$$

که در معادله ی (۲۰-۲)  $J$  ماتریس ژاکوبین<sup>۴</sup> است و مشتقات اولیه ی خطای شبکه با در نظر گرفتن وزن ها و بایاس را در بر می گیرد.  $e$  برداری از خطای شبکه است. ماتریس

<sup>۱</sup> -Levenberg-Marquardt

<sup>۲</sup> -Hessian

<sup>۳</sup> Mean square error

<sup>۴</sup> -Jacobian matrix



ژاکوپین از طریق تکنیک پس انتشار که نسبت به ماتریس هسیان ساده تر است، محاسبه می شود. با مینیمم کردن تابع خطا تغییرات وزن چنین محاسبه می شود:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{e} \quad (21-2)$$

در رابطه ی (21-2)  $\mu$  یک کمیت عددی است. اگر صفر باشد، این همان روش نیوتن است و به راحتی قابل حل است. زمانی که  $\mu$  بزرگ است، از روش گرادیان نزولی باید حل شود که محاسبه آن طولانی و وقتگیر است. روش نیوتن در نزدیکی خطای کم و کوچک روشی دقیق و سریع است. بنابراین  $\mu$  بعد از هر مرحله آموزش به طور موقت کاهش می یابد و تنها زمانی که تابع اجرا افزایش یافت، افزایش داده می شود [66].

مراحل طی شده حین آموزش شبکه با الگوریتم لوببرگ-مارکواردت چنین است:

۱- ارائه ورودی به شبکه، تعیین مقدار اولیه ی وزن و  $\mu$

۲- محاسبه خروجی شبکه و خطای مربوط به خروجی

۳- محاسبه ماتریس ژاکوپین،  $\mathbf{J}(x)$ ، که  $x$  وزن و بایاس مربوط به ورودی است.

۴- محاسبه  $\nabla$  یا تغییرات وزن طبق الگوریتم مربوطه

۵- محاسبه مجدد خطای شبکه

اگر خطای محاسبه شده در این حالت کمتر از حالت اول باشد در این حالت مقدار  $\mu$  توسط فاکتور کاهش دهنده کاهش یافته و مقدار جدید  $\mu$  و همچنین مقدار جدید وزن  $x = x_{prev} + \nabla x$  به مرحله اول بر می گردد.

اگر خطا در این مرحله افزایش داشته باشد در این حالت فقط مقدار  $\mu$  توسط فاکتور افزایش دهنده افزایش می یابد و به مرحله ۴ بر می گردد.

پارامترهای موجود در تابع train lm به قرار زیر هستند:

تعداد دوره‌های چرخش (epochs)، مقدار نهایی تابع اجرا (show)، زمان آموزش (time)، کمترین مقدار برای گرادیان خطا (mingrad)، مقدار اولیه  $\mu$  در رابطه مربوط به محاسبه تغییرات وزن ( $\mu$ )، کاهش دهنده مقدار  $\mu$  (mudec)، افزایش دهنده مقدار  $\mu$  (muinc)، فاکتور کاهش دهنده مقدار حافظه ای که این تابع برای انجام محاسبات مربوط به خطا به آن نیاز دارد (memreduce). پارامتر  $\mu$  یک مقدار اولیه برای  $\mu$  تعیین می کند. سپس زمانی که بعد از هر چرخه آموزشی خطای شبکه یا همان مقدار تابع اجرا (MSE) افزایش یابد، پارامتر  $\mu$ -inc مقدار  $\mu$  را افزایش می دهد. وقتی گرادیان خطای شبکه کاهش یابد پارامتر  $\mu$ -dec مقدار  $\mu$  را کاهش می دهد.

آموزش زمانی به پایان می رسد که:

۱- ماکسیمم تعداد دوره های چرخش انجام شده باشد.

۲- تابع اجرا به کمترین مقدار ممکن رسیده باشد [۶۷-۶۸].

## فصل سوم

مطالعات کمی ساختار-خاصیت بر روی اندیس بازداری  
هیدروکربن های آروماتیک چند حلقه ای

### ۳-۱- مقدمه

در این تحقیق به منظور مطالعه ی ارتباط کمی ساختار خاصیت برای مدل سازی اندیس بازداري هیدروکربن های آروماتیک چند حلقه ای، از روش رگرسیون خطی چند گانه (MLR) و شبکه ی عصبی مصنوعی (ANN) استفاده شد. به همین منظور با استفاده از اندیس بازداري یک سری از این ترکیبات که با سیستم GC/MS اندازه گیری شده است و به کارگیری روش رگرسیون خطی ابتدا مدل خطی طراحی شد. سپس با استفاده از شبکه ی عصبی مصنوعی این پارامتر شبیه سازی شد. در نهایت نتایج حاصل از رگرسیون خطی و شبکه ی عصبی مصنوعی با هم مقایسه شدند. بخش تجربی این کار شامل مراحل زیر می باشد.

۱- انتخاب سری داده ها

۲- محاسبه ی توصیف کننده ها

۳- مدل سازی

۴- ارزیابی مدل


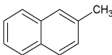
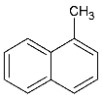
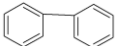
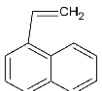
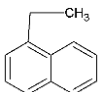
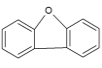
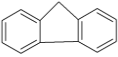
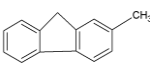
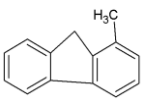
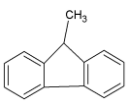
۳-۲- انتخاب سری داده ها

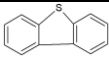
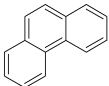
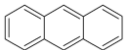
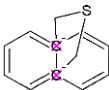
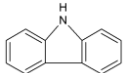

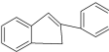
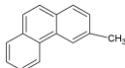
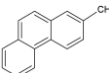
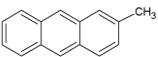

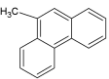

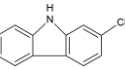
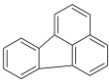
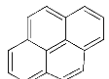
سری داده ها شامل مقادیر اندیس بازداری هیدروکربن های آروماتیک چند حلقه ای است که در سال ۲۰۰۳ رایان و لمیکس با استفاده از سیستم GC/MS اندازه گیری کردند. آنها از یک ستون HP-5MS با فاز ساکن قطبی برای شویش نمونه ها بهره گرفتند. این ترکیبات در اثر یک آتش سوزی در یک انبار نگه داری لاستیک ماشین در کانادا تولید شدند. لاستیک نوعی پلیمر است که از استیرن، پلی بوتادی ان، کربن سیاه، نایلون و سیم فلزی تشکیل شده است در حین احتراق این محصول ساخت بشر ترکیبات متنوعی از هیدروکربن های آروماتیک های چند حلقه ای تولید می شوند. فرآیند تشکیل یک هیدروکربن آروماتیک به این شکل است که در اثر پیرولیز یک آلکان یک اولفین تشکیل می شود و طی فرآیند دیلزآلدر اولفین تبدیل به حلقه آروماتیک می شود. از واکنش بین ترکیبات آروماتیک و اولفین ها نیز هیدروکربن های آروماتیک سنگین تر و فشرده تر تشکیل می شود. این ترکیبات بسیار سمی و سرطان زا هستند و برای سلامت انسان مضر می باشند. به همین جهت اندازه گیری این ترکیبات حائز اهمیت است [۶۹].



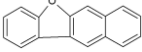
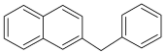
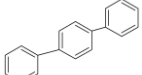
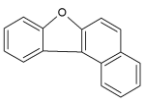
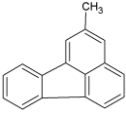
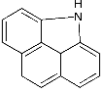
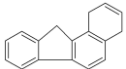
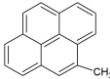
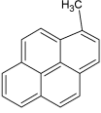
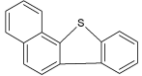
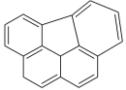
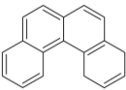
### ۳-۲-۱- نتایج

نتایج این بررسی نشان داد که دسته ی وسیعی از هیدروکربن های آروماتیک چند حلقه ای حین آتش سوزی تولید شدند. این ترکیبات اکثرا ایزومر بوده و خواص فیزیکی شیمیایی مشابهی دارند. در جدول ۳-۱ نام، شکل و اندیس بازداری این ترکیبات که توسط رایان و لمیکس اندازه گیری شده، در این جدول آمده است.

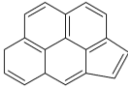
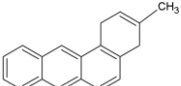
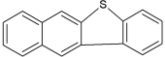
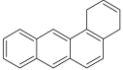
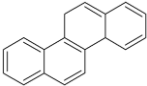
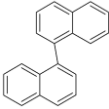
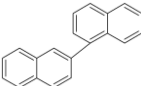
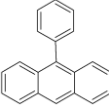
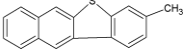
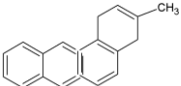
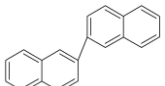
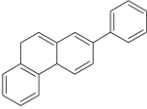
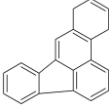
جدول ۱-۳- نام، شکل و مقادیر اندیس بازداري هيدروکربن های آروماتیک چند حلقه ای  
۵۶

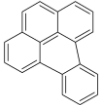
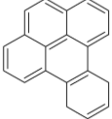
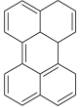
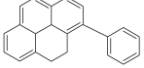
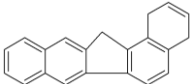
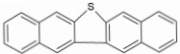
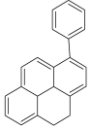
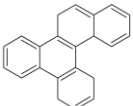
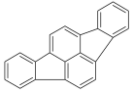
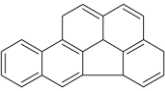
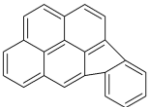
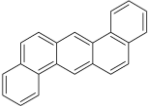
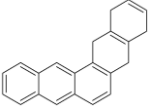
شماره	شکل	نام ترکیب	اندیس بازاری
1		naphthalene	200/00
2		2-Methyl naphthalene	221/10
3		1-Methyl naphthalene	224/12
4		Biphenyl	236/56
5		Acenaphthylene	248/27
6		Acenaphthene	253/56
7		Di Benz furan	259/63
8		Flourene	270/39
9		2-Methyl flourene	288/28
10		1-Methyl flourene	289/24
11		9-Methyl flourene	292/28

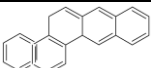
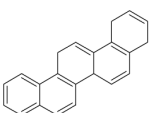


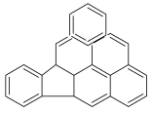
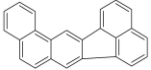
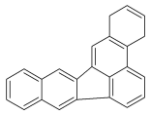
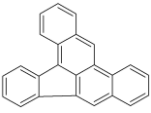
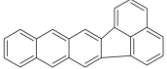
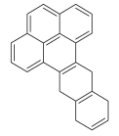

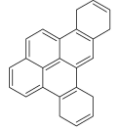

12		DiBenzo thiophene	295/95
13		Phenanthrene	300/00
14		Anthracene	301/53
15		Naphtha[2,3-b]thiophene	304/28
16		Carbazole	309/80
17		$\Delta V$ 1H-Indene,1-phenyl	311/12
18		1H-Indene,2-phenyl	314/15
19		3-Methyl Phenanthrene	318/83
20		3-Methyl Phenanthrene	319/67
21		2-Methyl Anthracene	321/12
22		Cyclopenta[def]Phenanthrene	325/73
23		9-Methylphenanthrene	322/74
24		1-Methylphenanthrene	323/54
25		2-Methylcarazol	326/94
26		fluoranthene	344/96
27		Acephenanthrylene	348/14

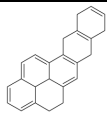
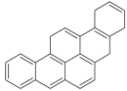
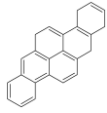
28		Phenanthro[4-5 bcd]thiophene	349/42
29		pyrene	352/41
30		Benzo[b]naphtha[2,3-d]furan	353/47
31		Phenyl methyl naphthalene	356/04
32		P-trephenyl	357/65
33		$\Delta \wedge$ Benzo[b]naphtha[1,2-d]furan	357/86
34		2-Methylfluoranthene	362/09
35		4H-Benzo[def]carbazole	362/82
36		Benzo[a]flourene	366/28
37		4-Methylpyrene	368/75
38		1-Methey pyrene	373/55
39		Benzo[b]naphtha[1,2-d]thiophene	391/73
40		Benzo[ghi]fluoranthene	390/18
41		Benzo[c]Phenanthrene	390/57



42		Cyclopenta[cd]pyrene	391/10
43		Methyl benzo[a]Anthracene	392/70
44		Benzo[b]naphtha[2,3.d]thiophene	394/76
45		Benzo[a]Anthracene	398/36
46		Chrysene	400/00
47		Δ4,1-Binaphthalene	402/77
48		1,2-Binaphthalene	404/58
49		9-phenyl Anthracene	406/47
50		Methyl benzo[b]naphtha[2,3-d]thiophene	409/54
51		Methylbenz[a]anthracene	410/27
52		2,2Binaphthalene	421/55
53		2-Phenylphenanthrene	424/19
54		Benzo[b]fluorescence	443/93

55		Benz[c]acephenanthrylene	448/07
56		Benzo[c]pyrene	452/93
57		Perylene	457/98
58		Phenylpyrene	459/01
59		DiBenzo[a,h]fluorene	456/60
60		Dinaphthothiophene	481/96
61		1-Phenyl pyrene	486/34
62		Benzo[b]triphenylene	483/27
63		Indeno[1,2,3-cd]fluorancene	488/38
64		Indeno[7,1,2,3-cdef]chrysene	491/53
65		Indeno[1,2,3-cd]pyrene	494/91
66		DiBenzo[ah]Anthracene	495/89
67		Pentaphene	496/67

68		Benzo[b]chrysene	498/86
69		Picene	500/00
70		Benzo[ghi]perylene	501/64
71		Benzo[def,mno]chrysene	505/97
72		DiBenzo[be]fluoranthene	533/22
73		Naphtha[1,2-k]fluoranthene	536/92
74		DiBenzo[b,k]fluranthene	539/59
75		DiBenzo[a,e]fluoranthene	540/71
76		Naphtho[2,3,k]fluoranthene	543/06
77		Naphtha[2,3,e]pyrene	547/71
78		Coronene	550/43
79		DiBenzo[a,e]pyrene	551/65
80		DiBenzo[a,l]pyrene	552/82

۸۱		Naphtho[2,3-a]pyrene	551/10
۸۲		DiBenzo[a,i]pyrene	556/32
۸۳		DiBenzo[a,h]pyrene	559/90

### ۳-۳- محاسبه توصیف کننده ها

توصیف کننده ها مقادیر عددی هستند که بیانگر ویژگی خاصی از مولکول هستند و با فعالیت شیمیایی آن ارتباط دارند. با توجه به روش مورد بررسی (بررسی ارتباط کمی ساختار - خاصیت) ابتدا ساختار مولکولها بوسیله نرم افزار HyperChem به صورت تقریبی رسم شد. پس از اتصال اتم های هیدروژن و انتخاب گزینه ی compute، با کمک روش نیمه تجربی AM1 ساختار ترکیبات بهینه شد.

با این نرم افزار اطلاعات زیادی نظیر زوایای پیوندی، طول پیوند ها، زوایای پیچشی، بار اتم و انرژی تشکیل مولکول را می توان بدست آورد. سپس ساختارهای رسم شده توسط این نرم افزار به عنوان ورودی به نرم افزار Dragon وارد شدند. محاسبه توصیف کننده ها توسط این نرم افزار صورت گرفت که در حدود ۱۴۸۱ توصیف کننده از هیجده گروه مختلف برای هر ترکیب بدست آمد. در جدول ۲-۳ فهرست این توصیف کننده ها آورده شده است.

جدول ۳-۲- نام توصیف کننده های محاسبه شده توسط نرم افزار Dragon<sup>۶۳</sup>

نام توصیف کننده	نام توصیف کننده
۱-توصیف کننده ی زیر ساختاری <sup>۱</sup>	۲-توصیف کننده ی هندسی <sup>۲</sup>
۳-توصیف کننده ی توپولوژی <sup>۳</sup>	۴-توصیف کننده ی RDF <sup>۴</sup>
۵-شمارنده های مولکول <sup>۵</sup>	۶-توصیف کننده ی سه بعدی مورس <sup>۶</sup>
۷-توصیف کننده ی BCUT <sup>۷</sup>	۸-توصیف کننده ی WHIM <sup>۸</sup>
۹-شاخص بار <sup>۹</sup>	۱۰-توصیف کننده های GETAWAY <sup>۱۰</sup>
۱۱-خود ارتباطی دو بعدی <sup>۱۱</sup>	۱۲-گروه های عاملی <sup>۱۲</sup>

<sup>۱</sup>- Constitutional Descriptors

<sup>۲</sup>- Geometrical Descriptors

<sup>۳</sup>- Topological Descriptors

<sup>۴</sup>- RDF

<sup>۵</sup>-Molecular walk count

<sup>۶</sup>- Molecule representation of structure backed on electron diffraction

<sup>۷</sup>- Charge indices

<sup>۸</sup>- Weighted holistic invariant molecular descriptors

<sup>۹</sup>- Charge indices

<sup>۱۰</sup>- Geometry, topology, and atom weights assembly

<sup>۱۱</sup>-2D autocorrelation

<sup>۱۲</sup>- Functional group

۱۳-توصیف کننده های بار <sup>۱</sup>	۱۴-اجزاء میان اتمی <sup>۲</sup>
۱۵-شاخص آروماتیسیته <sup>۳</sup>	۱۶-توصیف کننده های تجربی <sup>۴</sup>
۱۷-پروفیل مولکولی راندیک <sup>۵</sup>	۱۸-خصوصیات <sup>۶</sup>

### ۳-۴-انتخاب توصیف کننده های مناسب

طبیعی است که تعداد زیاد توصیف کننده ها باعث پیچیدگی محاسبات می شود و همچنین احتمال وجود توصیف کننده هایی که با هم برهمکنش دارند افزایش می یابد. لذا تعدادی از این توصیف کنندهها که دارای ۹۰٪ مقادیر یکسان و یا تنها دارای ۱۰٪ مقادیر غیر صفر بودند حذف شدند. با استفاده از نرم افزار SPSS ضریب همبستگی میان توصیف کننده ها و اندیس بازداری محاسبه شد. سپس از بین زوج توصیف کنندههایی که دارای همبستگی بیش از ۰/۹ با هم بودند توصیف کنندههایی که دارای همبستگی کمتری با متغیر وابسته بودند از سری داده ها حذف شدند.

۶۴

### ۳-۵- آنالیز رگرسیون خطی چندگانه (MLR)

برای ساختن مدلی که بیانگر ارتباط ساختاری ترکیبات مورد بررسی با فاکتور بازداری آنها باشد، از روش رگرسیون خطی چندگانه (MLR) استفاده شد. توصیف کنندههای محاسبه شده توسط نرم افزار Dragon، به عنوان متغیر های مستقل و اندیس بازداری به عنوان متغیر وابسته وارد نرم افزار

<sup>۱</sup>- Charge descriptors

<sup>۲</sup>- Atom-centered fragments

<sup>۳</sup>- Aromatic indices

<sup>۴</sup>-Empirical descriptor

<sup>۵</sup>- Rancid Molecular profiles

<sup>۶</sup>- Properties

SPSS شدند در نهایت با استفاده از روش آنالیز، گزینه ی رگرسیون خطی و انتخاب روش مرحله ای<sup>۱</sup> چندین مدل مختلف بدست آمد با توجه به خصوصیات آماری آنها از جمله ضریب همبستگی<sup>۲</sup> خطای استاندارد (SE)<sup>۳</sup>، مدلی که دارای بیشترین مقدار R و کمترین مقدار خطای استاندارد بود به عنوان بهترین مدل انتخاب شد. جدول ۳-۳ آماره های چند مدل طراحی شده با استفاده از روش رگرسیون خطی را نشان می دهد.

جدول ۳-۳- آماره های مدل های بدست آمده از روش MLR

شماره	توصیف کننده	R	SE	F
۱	SNar <sup>۱</sup>	0/990	13/4236	4139/466
۲	SNar, Moe06m <sup>۲</sup>	0/992	12/034	2585/719
۳	SNar, Mor06m, GATS1e <sup>۳</sup>	0/994	10/3587	2336/157
۴	SNar, Mor06m, GATS1e, RPCG	0/996	9/1223	2265/208

با توجه به اطلاعات بدست آمده از این جدول مدل چهارم به عنوان مدل نهایی انتخاب می شود. زیرا آماره F و R بزرگی دارد و همچنین مقدار خطای استاندارد آن کم می باشد. هر چه توصیف کننده های کمتری وارد شبکه عصبی شوند شبیه سازی پارامتر مورد نظر راحت تر انجام می شود. در جدول ۳-۴ نام توصیف کننده ها و اثر متوسط آنها را برای مدل منتخب نشان می دهد.

جدول ۳-۴- ضریب رگرسیون و اثر متوسط توصیف کننده ها

1-Stepwise  
2-Interrelation coefficient  
3-Standard Error

توصیف کننده	علامت	ضریب	اثر متوسط
توصیف کننده توپولوژی	SNar	۲۵/۹۰۶	57/805
سه بعدی مورس	Mor0۶ m	10/925	6/367
خود ارتباطی دوبعدی	GATS1e	49/140	6/884
بار نسبی مثبت	RPCG	120/007	4/885
ثابت		-100/456	-6/327
N=83	F=2265/208	R=0/996	SE=9/123

در جدول ۳-۴، N تعداد داده‌های موجود در سری آموزش، F مقدار مربوط به آماره ی F یا آزمون فیشر و R ضریب همبستگی خطی بین اندیس بازداری پیش بینی شده توسط رگرسیون خطی چنگدگانه و مقدار تجربی آن و SE خطای استاندارد رگرسیون است. توصیف کننده‌های وارد شده در مدل هر کدام یک خصوصیت ساختاری از مولکولها را بیان می‌کنند که نشان دهنده تأثیر بسزای ساختار مولکولهای مورد بررسی بر زمان بازداری آنها در ستون کروماتوگرافی است. از طرفی ضریب همبستگی میان توصیف کننده های وارد شده در مدل منتخب نشان دهنده عدم همبستگی قابل توجه میان این توصیف کننده ها است. جدول (۳-۵) ماتریس همبستگی توصیف کننده های وارد شده در مدل حاصل از رگرسیون خطی را نمایش می دهد. همچنین جدول (۳-۶) مقادیر عددی توصیف کننده ها را برای ترکیبات مورد بررسی نشان می دهد.

جدول ۳-۵- ماتریس همبستگی توصیف کننده های وارد شده در مدل منتخب

	Mor06m	RPCG	GATS1e	SNar
--	--------	------	--------	------



Mor06m	1	0/72	-0/596	0/679
RPCG	-0/072	1	0/055	-0/379
GATS1e	-0/569	0/055	1	0/618
SNar	0/679	0/379	-0/618	1

۶۷

جدول ۳-۶- مقادیر توصیف کننده های وارد شده در مدل خطی

شماره ترکیب	Mor06m	RPCG	GATS1e	SNar
1	0/815	0/125	1/611	7/742
2	0/817	0/096	1/658	11/538
3	0/666	0/095	1/658	11/038
4	2/114	0/292	1/313	10/633
5	1/530	0/100	1/516	11/326
6	1/188	0/101	1/516	11/326
7	2/502	0/124	1/208	12/019
8	0/565	0/179	2/230	10/633
9	0/811	0/091	1/614	12/019
10	1/070	0/093	1/614	12/019
11	1/711	0/092	1/614	11/731
12	0/925	0/113	1/736	8/148
13	1/689	0/092	1/614	11/731

14	1/337	0/093	1/611	11/731
15	1/147	0/105	1/429	12/830
16	1/745	0/092	1/614	11/731
17	1/846	0/092	1/614	11/731
18	0/660	0/157	2/227	11/038
19	0/936	0/102	1/401	13/523
20	1/823	0/100	1/401	13/523
21	2/310	0/279	1/249	12/830
22	1/823	0/100	1/401	13/523
23	1/052	0/112	1/736	8/148
24	1/203	0/083	1/095	14/216
25	1/224	0/075	1/658	13/405
26	2/172	0/072	1/621	14/099
27	0/618	0/105	1/561	14/216
28	1/076	0/093	1/493	13/929
29	0/951	0/173	2/152	12/830
30	1/182	0/087	1/493	14/216
31	2/139	0/092	1/493	13/929
32	2/149	0/092	1/493	13/929
33	2/476	0/263	1/307	14/216
34	1/209	0/101	1/674	9/129
35	1/369	0/102	1/313	15/720
36	2/232	0/084	1/465	14/909
३७	1/672	0/107	1/313	15/720
38	2/262	0/078	1/544	15/315
39	2/670	0/237	1/307	14/216

40	2/055	0/084	1/465	14/909
41	2/338	0/084	1/465	14/909
42	1/467	0/076	1/516	16/296
43	1/663	0/076	1/516	16/296
44	1/344	0/075	1/516	16/296
45	0/968	0/106	1/674	8/841
46	2/909	0/220	1/369	14/622
47	2/262	0/078	1/544	15/315
48	2/370	0/072	1/516	16/296
49	2/549	0/072	1/516	16/296
50	1/414	0/085	1/378	17/107
51	2/828	0/084	1/378	17/107
52	2/828	0/084	1/378	17/107
53	2/756	0/084	1/378	17/107
54	1/928	0/072	1/432	18/493
55	1/525	0/075	1/453	17/800
56	1/266	0/103	1/840	8/141
57	3/268	0/200	1/303	17/800
58	3/042	0/072	1/432	18/493
59	3/193	0/076	1/432	18/493
60	0/955	0/085	1/308	19/304
61	1/774	0/086	1/308	19/304
62	1/772	0/085	1/308	19/304
63	2/886	0/072	1/432	18/493
64	2/646	0/072	1/432	18/493
65	2/869	0/072	1/432	18/493

66	3/112	0/072	1/432	18/493
67	0/467	0/127	1/584	10/633
68	3/105	0/084	1/308	19/304
69	2/831	0/084	1/308	19/304
70	1/862	0/073	1/362	20/690
71	2/091	0/073	1/362	20/690
72	1/679	0/073	1/362	20/690
73	1/578	0/073	1/362	20/690
74	3/444	0/072	1/362	20/690
75	3/444	0/072	1/362	20/690
76	3/403	0/083	1/25	21/501
77	3/637	0/072	1/362	20/690
78	0/868	0/096	1/658	11/038
79	3/742	0/072	1/362	20/690
80	3/003	0/072	1/362	20/690
81	3/244	0/072	1/362	20/690
82	3/264	0/072	1/362	20/690
83	0/868	0/096	1/658	11/038

۳-۶-مدل سازی به روش شبکه عصبی

۳-۶-۱-مقدمه

در این تحقیق مدل شبکه عصبی مصنوعی با استفاده از برنامه نویسی در نرم افزار مطلب بدست آمد. الگوریتم شبکه Levenberg Marquardt و تابع انتقال سیگموئیدی است. معیار این الگوریتم برای پایان آموزش و بهینه کردن مقادیر وزن ورودی ها جهت پیش بینی پارامتر مورد بررسی، میزان خطای شبکه می باشد که طبق تابع اجرای تعیین شده محاسبه می گردد. در این روش تابع اجرا، میانگین مجذور خطای استاندارد (RMSE) است که به فرمول آن در بخش ۲-۲-۷-۱۰-۱ اشاره شده است. شبکه از لایه ی ورودی با ۴ گره، لایه ی پنهان و لایه ی خروجی با یک گره تشکیل شده است. تعداد گره لایه ی پنهان حین آموزش شبکه تعیین می شود. مقادیر توصیف کننده های وارد شده در مدل خطی به عنوان ورودی به شبکه داده می شود. مقدار تجربی اندیس بازداری نیز به عنوان هدف به دنبال ورودی ها وارد شبکه می شوند. سری داده ها به سه دسته ی آموزش، ارزیابی و پیش بینی تقسیم می شوند. با استفاده از توابع مناسب مقدار اولیه ای بین ۱- و ۱- برای وزنه های سری آموزش در نظر گرفته میشود. سپس مقادیر انتخابی وزنها در مقادیر توصیف کننده ها ضرب شده و ماتریس حاصل از این مقادیر وارد شبکه می شوند. سرانجام مقدار اندیس بازداری با استفاده از مقدار تجربی سری آموزش شبیه سازی می شود. مقدار اندیس بازداری با استفاده از مدل شبیه سازی شده، برای سری پیش بینی محاسبه می شود. مطابق تابع اجراء مربوطه مقدار خطای شبکه از تفاضل مقدار خروجی محاسبه شده و مقدار تجربی اندیس بازداری محاسبه می شود. مقادیر وزنها طی آموزش شبکه آنقدر تغییر می کنند تا مقدار خطای شبکه به کمترین مقدار ممکن برسد. پارامتر هایی که حین آموزش شبکه باید بهینه شوند شامل تعداد گره لایه پنهان و تعداد چرخه های آموزش می باشد. سرعت آموزش وزن و ممنتوم نیازی به بهینه کردن ندارد زیرا مقدار اولیه آنها توسط گرادیان نزولی `learnngdm` تعیین می شود و به دلیل بالا بودن سرعت آموزش نیازی به بهینه کردن آنها در حین آموزش شبکه نیست.

### ۳-۶-۲-انتخاب ورودی شبکه

در مدل سازی شبکه عصبی مصنوعی، از توصیف کننده های مدل منتخب MLR به عنوان ورودی استفاده شد وبااستفاده از الگوریتم آموزشی Levenberg Marquardt پیش بینی اندیس بازداری این ترکیبات انجام گرفت. ملاک انتخاب مقدار بهینه برای پارامترهای مربوطه زمانی است که میانگین خطای استاندارد (SE) طبق معادله زیر برای سری پیش بینی مینیمم شود.

$$SE = \sqrt{\frac{\sum(t_i - y_i)^2}{n}} \quad (۱-۳)$$

در رابطه ی (۱-۳)،  $t_i$  خروجی شبکه،  $y_i$  مقدار هدف (مقادیر تجربی اندیس بازداری) و  $n$  تعداد مولکول سری پیش بینی است.

۷۲

### ۳-۶-۳-انتخاب اولیه مقدار وزن ها و بایاس

با انتخاب تابع خلق شبکه ی newff، یک شبکه به روش پیش خور برگشتی<sup>۱</sup> طراحی شد. مقدار اولیه ای به صورت تصادفی به وزن ها داده شده و وزن ها طی آموزش بهینه می شوند. سرعت آموزش با این الگوریتم بسیار بالا است و نیازی به بهینه کردن سرعت آموزش وزن حین آموزش نیست زیرا با تغییر مقدار سرعت آموزش وزن (lr) تغییری در خطای شبکه ایجاد نمی شود [۶۱].

### ۳-۶-۴-بهینه کردن تابع انتقال و گره لایه ی پنهان

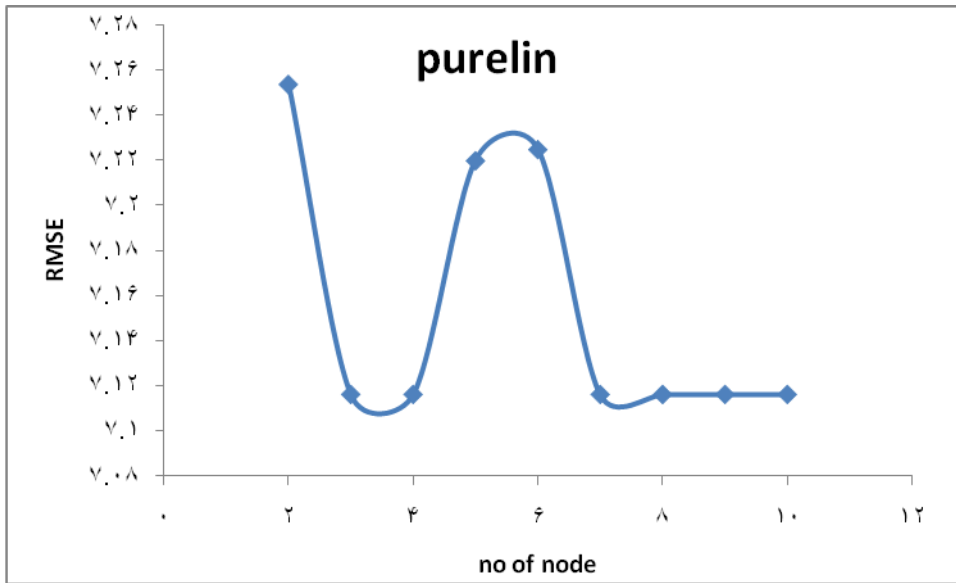
---

<sup>۱</sup>Feedforward

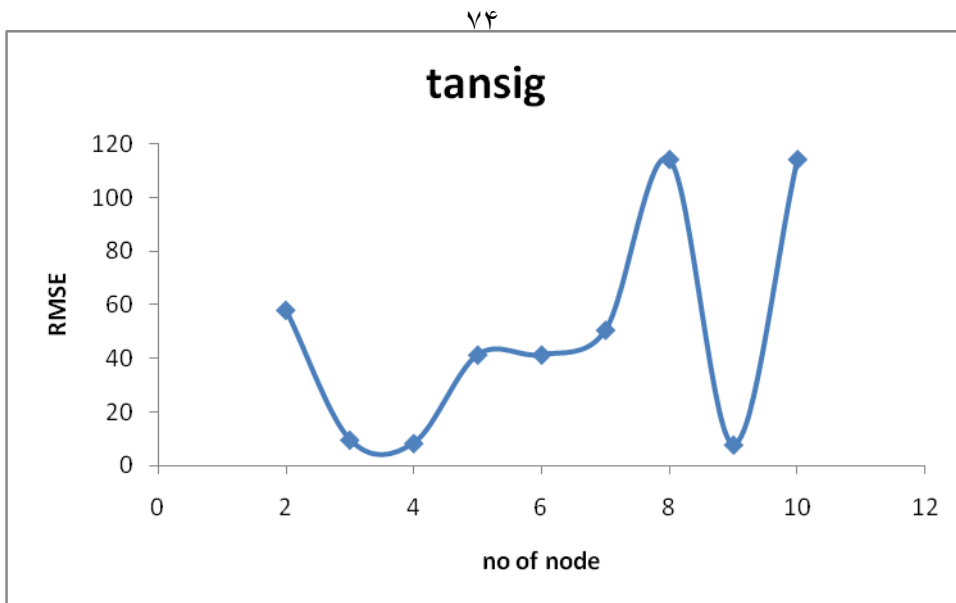
جهت انتخاب تابع انتقال مناسب برای لایه ی پنهان شبکه از سه تابع انتقال تانژانتی (tansig)، لگاریتمی (logsig) و خطی (purelin) استفاده کردیم. در این برنامه با انتخاب هر تابع مقدار گره را تغییر می دهیم. جدول 3-7 مقادیر گره و میانگین مجذورخطای استاندارد (RMSE) را برای سه تابع انتقال نشان می دهد. با توجه به مقادیر خطا برای توابع انتقال در جدول 3-7 مشاهده می کنیم که کمترین مقادیر خطا برای تابع خطی مشاهده می شود. شکل های 3-1، 3-2 و 3-3 نمودار تغییرات خطا برحسب تغییرات گره را برای توابع انتقال ذکر شده نشان می دهد. بررسی این نمودار ها نیز نشان می دهند که تابع خطی با سه گره دارای کمترین مقدار خطا است.

جدول 3-7-مقادیر گره توابع انتقال

گره	RMSE tansin	RMSE logsig ۷۳	RMSE purelin
۲	58/906	114/780	7/254
۳	9/636	7/470	7/116
۴	8/384	97/757	7/116
۵	41/370	88/124	7/220
۶	41/370	58/985	7/225
۷	50/632	58/353	7/116
۸	114/280	7/856	7/116
۹	7/810	10/460	7/116
۱۰	114/280	7/487	7/116

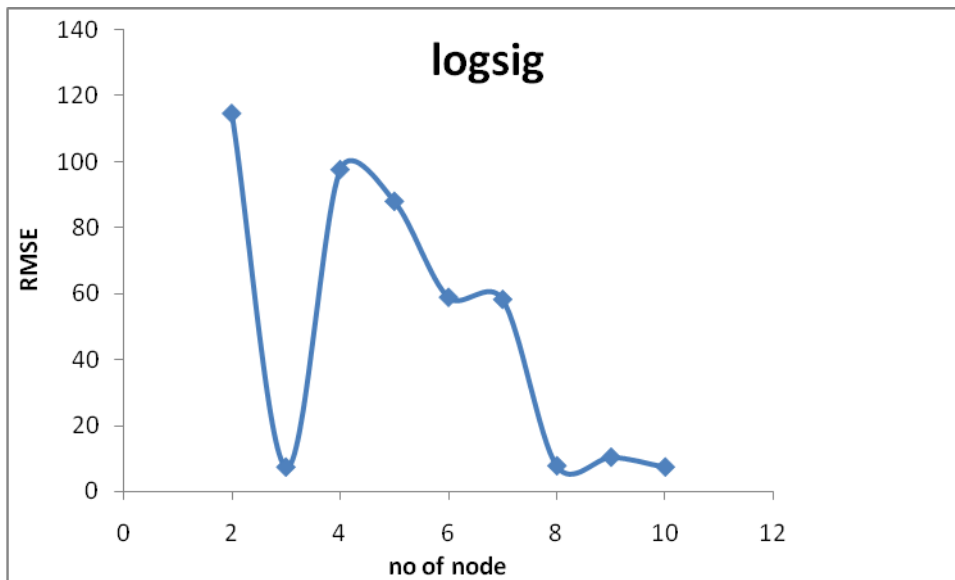


شکل ۳-۱- تغییرات گره و مقدار خطای مربوطه ی تابع خطی



شکل ۳-۲- تغییرات گره و خطای تابع تانژانتی

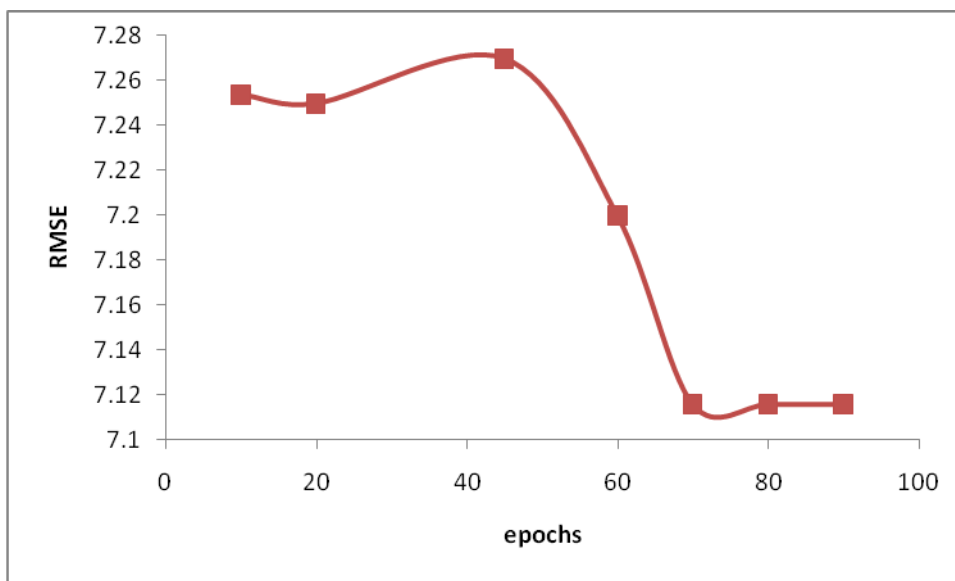




شکل ۳-۳- نمودار تغییرات گره و خطای تابع لگاریتمی

### ۳-۶-۵- بهینه سازی تعداد چرخه های آموزشی (epochs)

برای جلوگیری از آموزش بیش از حد شبکه (Overtraining)، منحنی تعداد چرخه های آموزش بر حسب خطا برای سری پیش بینی رسم می شود. مقدار بهینه جایی است که میزان خطای سری پیش بینی مینیمم مقدار ممکن شود. در این تحقیق با انتخاب تابع خطی و قرار دادن تعداد سه گره در لایه ی پنهان شبکه، مقدار چرخه ی آموزش را تغییر می دهیم تا به مقدار بهینه ی آن برسیم و خطای استاندارد به کمترین حد خود برسد. شکل ۳-۴ نمودار تغییرات خطا بر حسب تغییرات چرخه های آموزش را برای تابع خطی نشان می دهد. مشاهده می شود که تغییرات خطا در حین تغییر در چرخه های آموزش برای تابع خطی ناچیز است. در نهایت تعداد ۹۰ چرخه ی آموزش به عنوان مقدار بهینه انتخاب شد.



شکل ۳-۴- نمودار تغییرات چرخه های آموزشی و خطای مربوطه

جدول ۳-۸- شرایط بهینه برای آموزش شبکه برای هیدروکربن های آروماتیک چند حلقه ای را نشان

می دهد.

جدول ۳-۸- شرایط بهینه ی شبکه

۳	تعداد گره لایه ی پنهان
purelin	تابع انتقال
۹۰	تعداد دورهای آموزش (epochs)
۱۰ <sup>-۸</sup> -۳۰	مینیمم شیب عملکرد (min grad)
۱	سرعت آموزش وزن
۰/۱	ممنتوم
۰/۰۱	فاکتور کاهش دهنده ی پارامتر تطبیق مارکورت (dec mu)

۰/۱	فاکتور افزایش دهنده ی پارامتر تطبیق مارکورت ( $inc\ \mu$ )
۵۰	مقدار ماکزیمم برای پارامتر تطبیق مارکورت ( $\mu\ max$ )
۰/۱	پارامتر تطبیق مارکورت ( $\mu$ )

## فصل<sup>۷۷</sup> چهارم

# بحث و نتیجه گیری

#### ۴-۱- مقایسه روش های رگرسیون خطی چند گانه و شبکه عصبی مصنوعی با

#### استفاده از سری ارزیابی و تست

برای مقایسه ی مدل خطی و غیر خطی طراحی شده توسط شبکه ی عصبی مصنوعی مقادیر اندیس بازداری توسط این دو مدل برای سری پیش بینی محاسبه گردید. در روش خطی داده ها به سه دسته ی آموزش، پیش بینی و ارزیابی تقسیم بندی شدند. سپس با استفاده از مقادیر توصیف کننده های وارد شده در مدل برگزیده برای سری آموزش، انتخاب روش رگرسیون خطی و روش ورود اجباری<sup>۱</sup> ضرایب توصیف کننده ها در مدل خطی استخراج شد. در نهایت مقدار اندیس بازداری برای سری پیش بینی محاسبه شد.

در روش غیر خطی با اجرا کردن برنامه، شبکه با استفاده از سری آموزش پارامتر مورد نظر را شبیه سازی کرده و مقدار اندیس بازداری را برای سری پیش بینی

---

<sup>۱</sup>-Eneter

محاسبه کرد. با رسم نمودار مقدار تجربی اندیس بازداری بر حسب مقدارپیش بینی شده آن می توان توانایی مدل خطی و غیر خطی را در پیش بینی اندیس بازداری مقایسه کرد. جدول (۱-۴) و (۲-۴) مقدار اندیس بازداری محاسبه شده با مدل خطی و غیر خطی برای سری پیش بینی را نشان می دهد.

جدول ۱-۴- مقدار اندیس بازداری پیش بینی شده برای سری پیش بینی باروش MLR<sup>۷۸</sup>

شماره ترکیب	مقادیر تجربی RI	مقادیر محاسبه شده با MLR	درصد خطا
۴	295/95	298/11	0/732
۸	309/80	312/96	1/021
۱۲	221/10	219/73	-0/616
۲۸	362/09	356/66	-1/498
۳۲	373/55	368/05	-1/470
۳۶	390/57	392/01	0/369
۵۲	452/93	451/18	-0385
۵۶	253/65	245/04	-3/360

۶۰	488/38	484/75	-0/742
۷۶	550/43	556/01	2/649
۸۰	551/10	343/85	-1/316
۸۲	559/90	546/65	-2/366

جدول ۴-۲- مقادیر محاسبه شده اندیس بازداري برای سری پیش بینی با روش ANN

شماره ترکیب	مقادیر تجربی RI	مقادیر محاسبه شده با ANN	درصد خطا
۴	295/95	298/12	0.733
۸	309/80	312/97	1/022
۱۲	221/10	219/74	-0/615
۲۸	362/09	356/67	-1/497
۳۲	373/55	368/06	-1/469
۳۶	390/57	392/02	0/370
۵۲	452/93	451/19	-0383
۵۶	253/65	245/04	-3/359
۶۰	488/38	484/76	-0/741

۷۶	550/43	565/02	2/650
۸۰	551/10	543/85	-1/315
82	559/90	546/66	-2/365

جهت ارزیابی بیشتر صحت اعتبار مدل های خطی و غیر خطی مقدار اندیس بازداری برای سری تست نیز پیش بینی می شود. جدول (۴-۳) و (۴-۴) مقادیر محاسبه شده ی اندیس بازداری سری تست توسط مدل خطی و غیر خطی را نشان می دهد.

۸۰

جدول ۴-۳-مقادیر پیش بینی شده برای سری تست با روش MLR

شماره ترکیب	مقادیر تجربی RI	مقادیر محاسبه شده باروش MLR	درصد خطا
۱۶	322/74	312/63	-3/129
۲۰	348/14	350/42	0/657
۲۴	353/47	369/29	1/475
۳۸	392/70	406/06	3/403
۴۰	398/36	390/11	-2/070

۴۴	406/47	419/90	3/305
۴۸	421/55	430/56	2/138
۶۴	496/67	486/42	-2/064
۶۸	501/64	507/75	1/218
۷۲	539/59	530/53	-1/677

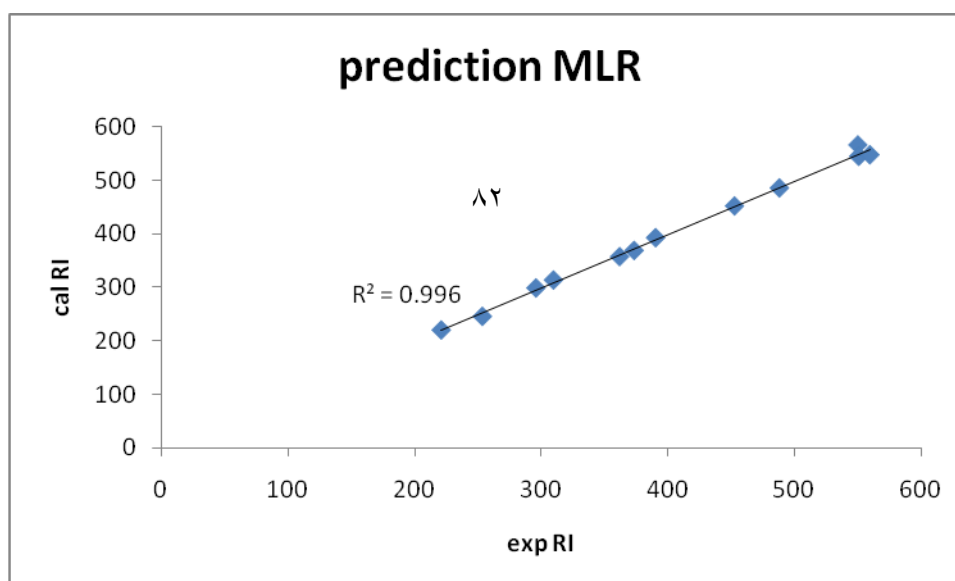
جدول ۴-۴- مقادیر محاسبه شده اندیس بازداري برای سري تست با روش ANN

شماره تركيب	مقادير تجربی RI	مقادير محاسبه شده با ANN	درصد خطا
۱۶	322/74	312/64	-3/128
۲۰	348/14	350/43	0/657
۲۴	353/47	369/29	4/476
۳۸	392/70	401/54	3/403
۴۰	398/36	390/11	-2/070

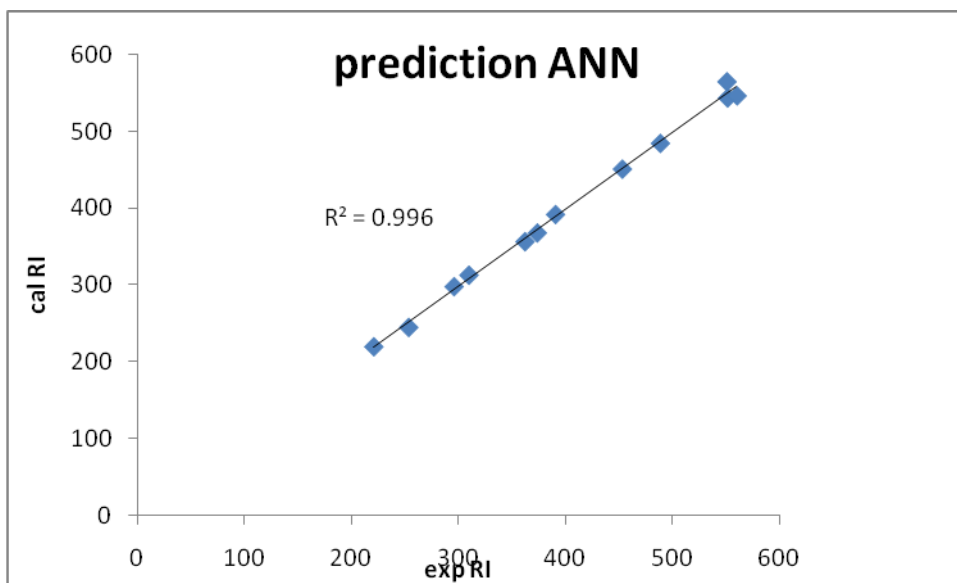


۴۴	406/47	416/93	3/306
۴۸	421/55	430/57	2/138
۶۴	496/67	486/42	-2/063
۶۸	501/64	507/76	1/219
۷۲	539/59	530/54	-1/676

شکل ۱-۴ و ۲-۴ نمودار مقادیر تجربی اندیس بازداری هیدروکربنها ی آروماتیک چند حلقه ای بر حسب مقادیر محاسبه شده توسط مدل خطی و غیر خطی برای سری پیش بینی می باشد.

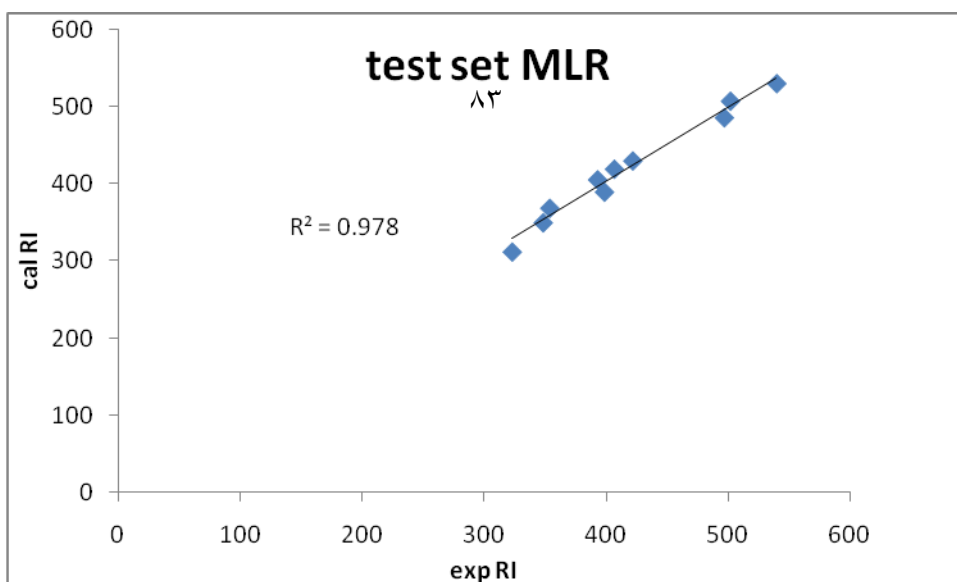


شکل ۱-۴-نمودار مقادیر تجربی اندیس بازداری بر حسب مقادیر پیش بینی شده با روش MLR

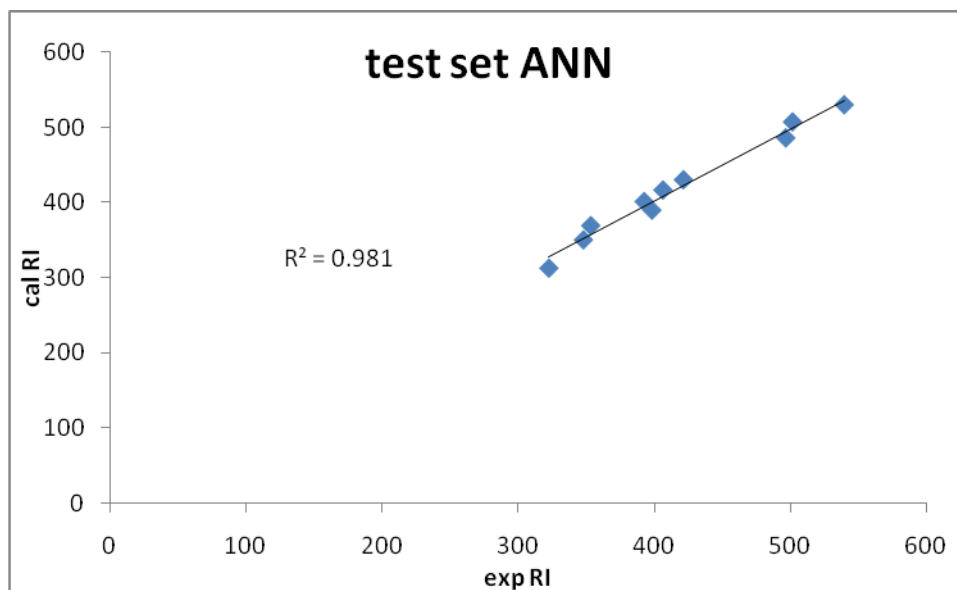


شکل ۴-۲- مقادیر تجربی اندیس بازداري بر حسب مقادير پيش بيني شده توسط روش ANN

همچنين شكل هاي ۴-۳ و ۴-۴ نمودار مقادير تجربی بر حسب مقادير پيش بيني شده ی اندیس بازداري توسط دو مدل مذکور برای سری تست می باشد.



شکل ۴-۳- نمودار مقادير تجربی اندیس بازداري بر حسب مقادير پيش بيني شده با روش خطی



شکل ۴-۴- نمودار مقادیر تجربی برحسب مقادیر پیش بینی شده توسط روش غیر خطی

مقادیر ضریب همبستگی نمودارهای مقادیر تجربی بر حسب مقادیر محاسبه شده با روش غیر خطی بیشتر است. این نشان می دهد که روش غیر خطی در پیش بینی اندیس بازداری توانمند تر است...

#### ۴-۲- بررسی مدل خطی و غیر خطی با استفاده از روش رد مرحله ای تک تک<sup>۱</sup>

به منظور بررسی بیشتر مدل خطی و <sup>۸۴</sup> مدل طراحی شده توسط شبکه ی عصبی مصنوعی با استفاده از روش رد مرحله ای تک تک مقدار اندیس بازداری را محاسبه میکنیم. روش رد مرحله ای تک تک یکی از روش های ارزیابی تقاطعی<sup>۲</sup> می باشد. در روش رگرسیون خطی یک ترکیب را از سایر داده ها جدا می کنیم و با بقیه و استفاده از روش Enter ضرایب توصیف کننده های وارد شده در مدل را استخراج می کنیم. سپس مقدار اندیس بازداری را برای ترکیب کنار گذاشته شده محاسبه می کنیم. در روش غیر خطی یک ترکیب را به عنوان پیش بینی و بقیه

<sup>۱</sup>-leave one out

<sup>۲</sup>-Cross validation

ترکیبات را به عنوان سری آموزش در نظر می گیریم و شبکه را آموزش می دهیم. بارسم منحنی مقدار تجربی اندیس بازداری بر حسب مقدار پیش بینی شده آن و محاسبه ضریب همبستگی خطی میتوان این دو مدل را مقایسه کرد. جدول ۴-۵ و ۴-۶ مقادیر پیش بینی شده اندیس بازداری با روش مرحله ای تک تک توسط مدل خطی و غیر خطی را نشان می دهد.

جدول ۴-۵- مقادیر تجربی زمان بازداری و پیش بینی شده توسط MLR , برای سری داده ها

شماره ترکیب	مقدار تجربی RI	مقدار محاسبه شده توسط MLR	درصد خطا
۱	200/00	203/41	1/705
۲	289/24	287/341	-0/656
۳	292/28	285/70	-2/251
۴	295/95	298/17	0/7517
۵	300/00	296/02	-1/325
۶	301/53	292/21	-3/088

۷	304/28	313/56	3/0520
۸	309/80	313/21	1/101
۹	311/12	309/95	-0/373
۱۰	314/15	313/03	-0/354
۱۱	318/83	312/27	-2/056
۱۲	221/10	219/81	-0/579
۱۳	319/67	312/00	-2/398
۱۴	321/12	308/18	-4/029
۱۵	325/73	328/61	0/885
۱۶	322/74	312/51	-3/166
۱۷	323/54	313/59	-3/074
۱۸	326/94	318/82	-2/482
۱۹	344/96	۸۶ 340/96	-1/158
۲۰	348/14	350/70	0/735
۲۱	349/42	352/71	0/942
۲۲	352/41	350/57	-0/520
۲۳	224/12	220/65	-1/546
۲۴	353/47	369/70	4/593
۲۵	356/04	350/50	-1/553
۲۶	357/65	377/55	5/565
۲۷	357/86	364/24	1/785

۲۸	362/09	356/50	-1/541
۲۹	362/82	371/07	2/274
۳۰	366/28	364/48	-0/489
۳۱	368/75	368/15	-0/162
۳۲	373/55	368/14	-1/446
۳۳	391/73	390/43	-0/330
۳۴	236/56	244/02	3/156
۳۵	390/18	398/94	2/245
۳۶	390/57	392/26	0/434
۳۷	391/10	402/84	3/003
۳۸	392/70	406/54	3/524
۳۹	394/76	<sup>۸۷</sup> 388/88	-1/488
۴۰	398/36	390/15	-2/058
۴۱	400/00	393/85	-1/536
۴۲	402/77	421/83	4/734
۴۳	404/58	423/94	4/786
۴۴	406/47	420/30	3/404
۴۵	248/27	180/31	-27/372
۴۶	409/54	403/07	-1/578
۴۷	410/27	406/13	-1/006

۴۸	421/55	430/95	2/231
۴۹	424/19	432/95	2/065
۵۰	443/93	435/72	-1/848
۵۱	448/07	451/50	0/766
۵۲	452/93	451/36	-0/346
۵۳	457/98	450/40	-1/653
۵۴	459/01	490/74	6/913
۵۵	456/60	457/77	0/257
۵۶	253/56	244/53	-3/559
۵۷	481/96	484/69	0/568
۵۸	486/34	490/94	0/946
۵۹	483/27	۸۸ 493/47	2/111
۶۰	488/38	483/96	-0/903
۶۱	491/53	493/72	0/446
۶۲	494/91	493/374	-0/310
۶۳	495/89	488/92	-1/404
۶۴	496/67	486/26	-2/095
۶۵	498/86	488/63	-2/050
۶۶	500/00	491/23	-1/752
۶۷	259/63	273/97	5/523

۶۸	501/64	508/07	1/282
۶۹	505/97	504/88	-0/215
۷۰	533/22	531/43	-0/333
۷۱	536/92	533/89	-0/563
۷۲	539/59	529/65	-1/840
۷۳	540/71	528/60	-2/238
۷۴	543/06	527/06	-2/946
۷۵	547/71	548/78	0/196
۷۶	550/43	565/95	2/820
۷۷	551/65	550/78	-0/157
۷۸	270/39	288/59	6/732
۷۹	552/82	<sup>۸۹</sup> 551.91	-0/164
۸۰	551/10	543/59	-1/362
۸۱	556/32	546/03	-1/848
۸۲	559/9	546/05	-2/472
۸۳	288/28	287/96	-0/110

جدول ۴-۶- مقادیر پیش بینی شده اندیس بازداري توسط روش ANN برای سری داده ها



شماره ترکیب	مقدار تجربی RI	مقدار محاسبه شده توسط ANN	درصد خطا
۱	200/00	203/40	1/700
۲	289/24	287/34	-0/655
۳	292/28	285/36	-2/365
۴	295/95	298/17	0/752
۵	300/00	296/03	-1/323
۶	301/53	292/23	-3/084
۷	304/28	313/56	3/049
۸	309/80	313/21	1/100
۹	311/12	309/06	-0/662
۱۰	314/15	313/08	-0/340
۱۱	318/83	312/28	-2/054
۱۲	221/10	219/82	-0/578
۱۳	319/67	312/01	-2/396
۱۴	321/12	308/19	-4/026
۱۵	325/73	327/33	0/491
۱۶	322/74	312/52	-3/166
۱۷	323/54	313/60	-3/072
۱۸	326/94	318/83	-2/480

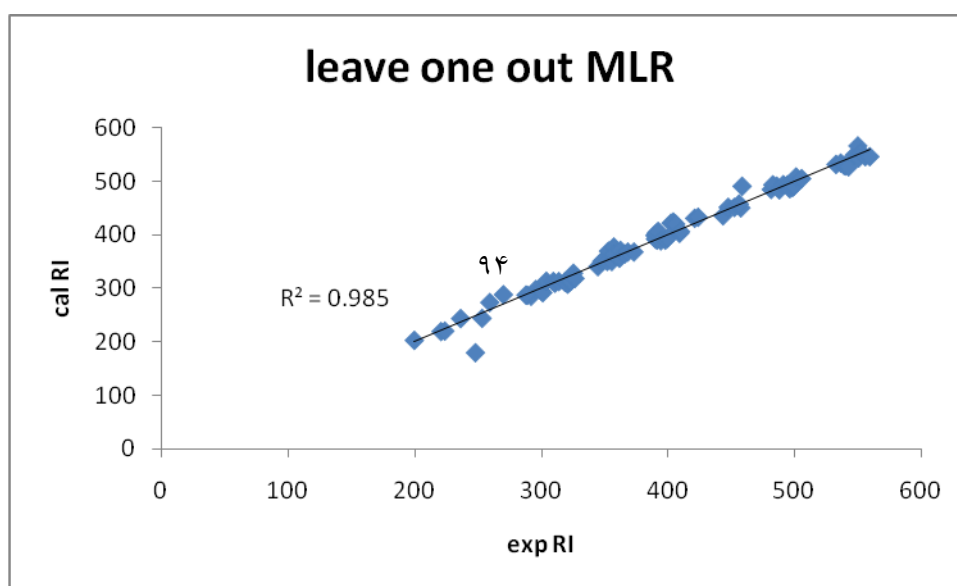
۱۹	344/96	326/52	-5/345
۲۰	348/14	350/70	0/735
۲۱	349/42	352/71	0/941
۲۲	352/41	350/58	-0/519
۲۳	224/12	220/66	-1/543
۲۴	353/47	369/68	4/585
۲۵	356/04	350/50	-1/556
۲۶	357/65	377/53	5/558
۲۷	357/86	356/52	-0/374
۲۸	362/09	368/14	1/670
۲۹	362/82	371/05	2/268
۳۰	366/28	364/50	-0/485
۳۱	36875	<sup>۹۱</sup> 368/15	-0/162
۳۲	373/55	368/15	-1/445
۳۳	391/73	390/43	-0/331
۳۴	236/56	244/02	3/153
۳۵	390/18	398/94	2/245
۳۶	390/57	392/26	0/432
۳۷	391/10	402/80	2/991
۳۸	392/70	406/50	3/514

۳۹	394/76	388/89	-1/486
۴۰	398/36	390/16	-2/058
۴۱	400/00	393/23	-1/692
۴۲	402/77	421/80	4/724
۴۳	404/58	423/88	4/770
۴۴	406/47	420/31	3/404
۴۵	248/27	233/36	-6/005
۴۶	409/54	403/08	-1/577
۴۷	410/27	406/14	-1/006
۴۸	421/55	430/90	2/218
۴۹	424/19	432/90	2/053
۵۰	443/93	435/73	-1/847
۵۱	448/07	۹۲ 431/51	-3/695
۵۲	452/93	451/36	-0/34663
۵۳	457/98	450/42	-1/650
۵۴	459/01	490/70	6/903
۵۵	456/60	457/70	0/240
۵۶	253/56	224/54	-11/445
۵۷	481/96	448/69	-6.903
۵۸	486/34	491/10	0/978

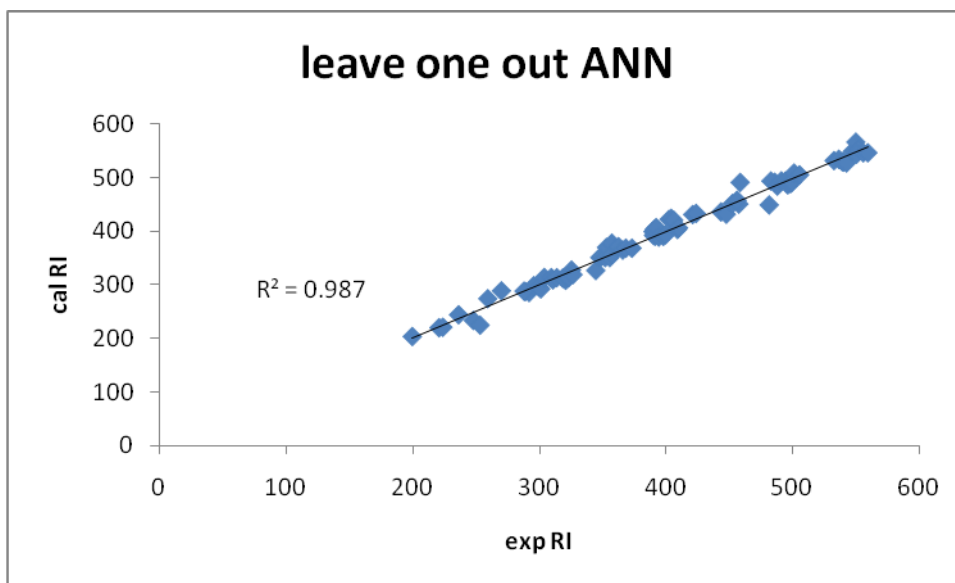
۵۹	483/27	493/40	2/096
۶۰	488/38	483/97	-0/902
۶۱	491/53	493/72	0/445
۶۲	494/91	493/38	-0/309
۶۳	495/89	488/93	-1/403
۶۴	496/67	486/26	-2/095
۶۵	498/86	488/64	-2/048
۶۶	500/00	491/25	-1/750
۶۷	259/63	273/95	5/515
۶۸	501/64	508/14	1/295
۶۹	505/97	504/88	-0/215
۷۰	533/22	531/45	-0/331
۷۱	536/92	۹۳ 533/91	-0/560
۷۲	539/59	529/66	-1/840
۷۳	540/71	528/61	-2/237
۷۴	543/06	527/07	-2/944
۷۵	547/71	548/70	0/180
۷۶	550/43	565/90	2/810
۷۷	551/65	550/77	-0/159
۷۸	270/39	288/59	6/7310

۷۹	552/82	551/92	-0/162
۸۰	551/1۰	543/60	-1/360
۸۱	556/32	546/04	-1/847
۸۲	559/90	546/08	-2/468
۸۳	288/28	287/96	-0/111

شکل های ۴-۵ و ۴-۶ نمودار مقادیر تجربی برحسب مقادیر محاسبه شده به روش رد مرحله ای تک تک توسط مدل خطی و شبکه ی عصبی مصنوعی می باشد..



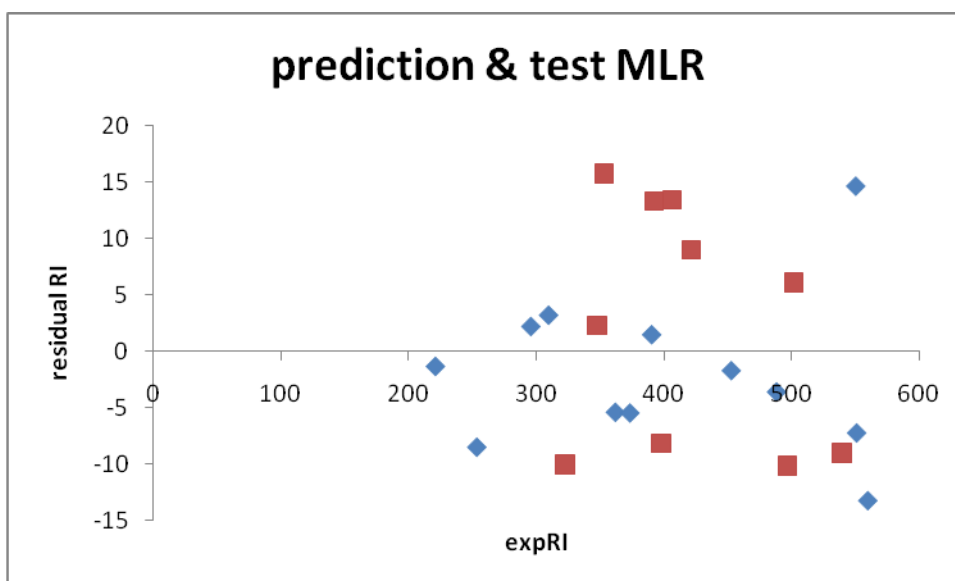
نمودار ۴-۵-نمودار مقادیر پیش بینی شده با روش رگرسیون خطی برای سری داده ها



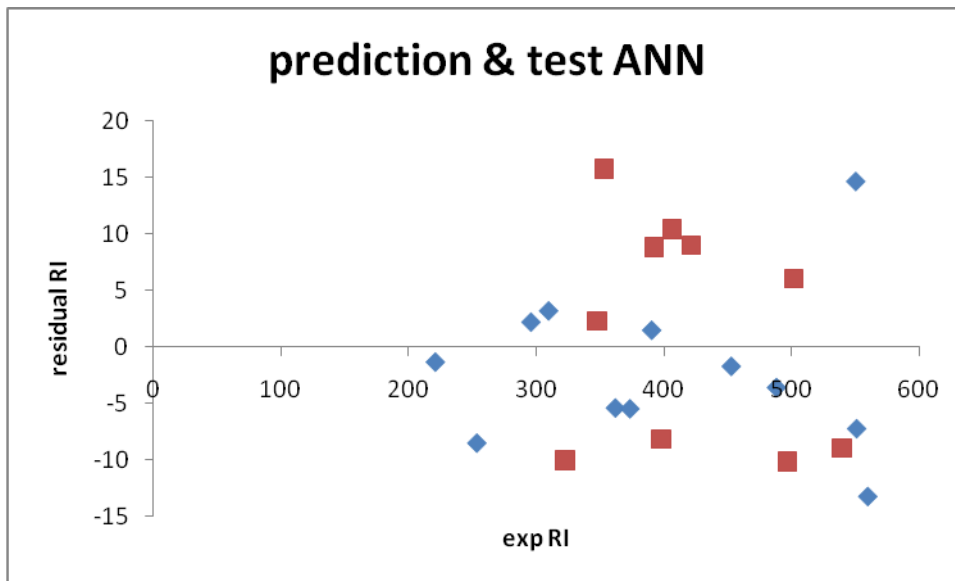
نمودار ۴-۶- مقادیر پیش بینی شده با مدل غیر خطی برای سری داده‌ها

مقدار ضریب همبستگی این نمودار ها نشان می دهد که مقادیر محاسبه شده ی اندیس بازداری با شبکه ی عصبی مصنوعی به مقدار واقعی آن نزدیک تر است. شکل های ۴-۷ و ۴-۸ نمودار باقیمانده ها که از تفاضل مقدار محاسبه شده از مقدار واقعی اندیس بازداری بدست می آید، بر حسب مقدار واقعی این پارامتر برای سری پیش بینی و تست می باشد. نقاط بزرگ مربوط به سری پیش بینی و نقاط کوچک مربوط به سری تست است.

۹۵

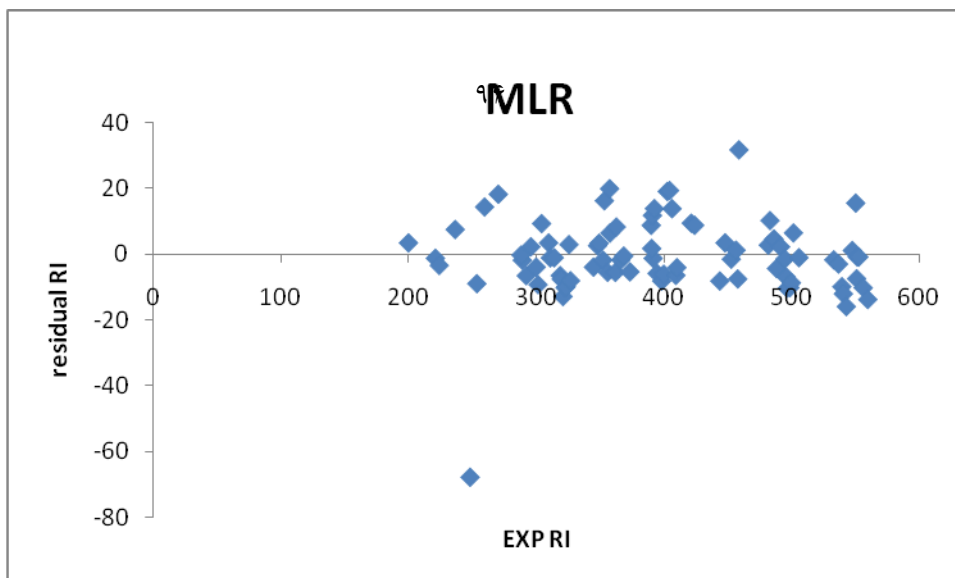


نمودار ۴-۷- نمودار مقادیر باقی مانده بر حسب مقادیر تجربی اندیس بازدارنده برای روش MLR

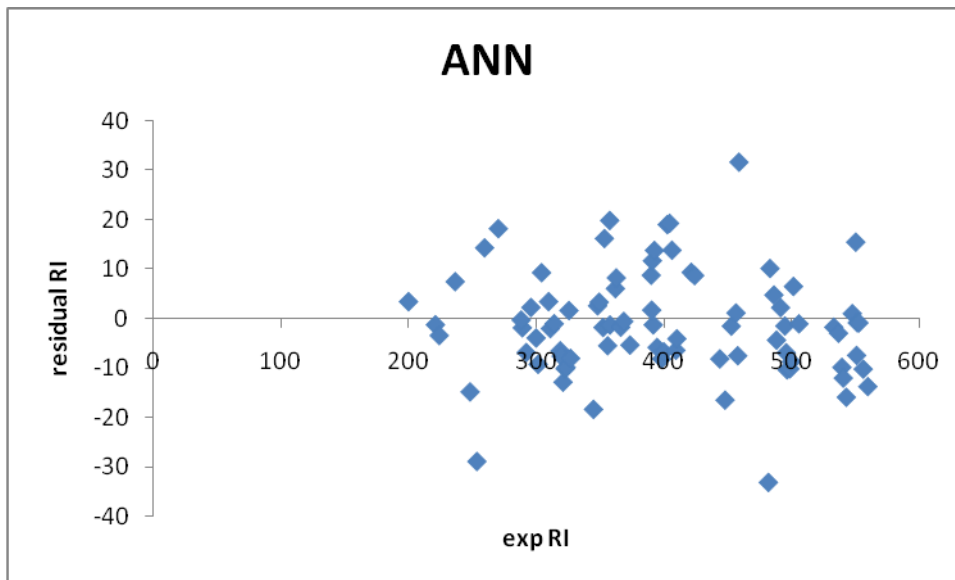


نمودار ۴-۸- مقادیر باقی مانده بر حسب مقادیر تجربی اندیس بازدارنده برای روش ANN

شکل های ۴-۹ و ۴-۱۰ نیز نمودار باقی مانده های محاسبه شده باروش رد مرحله ای تک تک توسط رگرسیون خطی و شبکه ی عصبی مصنوعی می باشد.



شکل ۴-۹- نمودار باقیمانه های محاسبه شده توسط MLR با روش رد مرحله ای تک تک



شکل ۴-۱۰- نمودار باقی مانده های محاسبه شده توسط ANN با روش رد مرحله ای تک تک

تقارن پراکنده گی نقاط در دو طرف محور افقی در نمودار باقی مانده ها نشان دهنده ی نبود خطای معین (سیستماتیک) است. تمرکز نقاط به سمت محور افقی نیز نشان دهنده ی نزدیک بودن مقادیر پیش بینی شده به مقادیر واقعی اندیس بازداری است. در نمودار های مربوط به روش غیر خطی تقارن نقاط در دو طرف محور افقی بیشتر است. جهت بررسی و مقایسه ی این دو روش در پیش بینی اندیس بازداری به محاسبه ی چند پارامتر پرداختیم. آماره های دوروش رگرسیون خطی و شبکه عصبی مصنوعی در جدول ۴-۷ آمده است که می توان کارایی این دو روش را مقایسه کرد.

جدول ۴-۷- مقایسه پارامترهای آماری بدست آمده از روش ANN و MLR

	پیش بینی	تست	پیش بینی	تست	سری داده ها	سری داده ها
	MLR	MLR	ANN	ANN	MLR	ANN



REP	1/874	2/610	1/778	2/180	2/945	2/678
RMSE	7/796	10/442	7/110	10/430	11/783	10/716
MAE	1/396	2/388	1/390	2/209	2/291	2/183
D	1/377	2/414	1/370	2/410	2/185	2/168
F	1/750	1/250	1/750	1/250	19/500	19/500
R <sup>2</sup>	0/996	0/978	0/997	0/981	0/985	0/987

روابط (۱-۴)، (۲-۴)، (۳-۴)، (۴-۴)، (۵-۴) و (۶-۴) طریقه ی محاسبه ی پارامترهای جدول ۷-۴ را نشان می دهند

۹۸

$$MSE = \frac{\sum (y_i - \hat{y}_i)}{n} \quad (۱-۴)$$

$$RED = \frac{100}{\bar{Y}} \times \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \quad (۲-۴)$$

$$MAE = \frac{\sum |y_i - \hat{y}_i| / y_i}{n} \times 100 \quad (۳-۴)$$

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

$$F = \frac{\sum (\hat{y}_i - \bar{y})^2 / p}{\sum (\hat{y}_i - \bar{y})^2 / (n - p - 1)}$$

(۴-۴)

(۵-۴)

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (۶-۴)$$

در فرمول های (۱-۴)، (۲-۴)، (۳-۴)، (۴-۴)، (۵-۴) و (۶-۴)  $y_i$  اشاره دارد به مقدار تجربی اندیس بازداری، مقدار  $\hat{y}_i$  اشاره دارد به اندیس بازداری پیش بینی شده و  $\bar{y}$  مربوط به میانگین تجربی مقدار اندیس بازداری است.  $n$  مربوط به تعداد ترکیبات مورد بررسی و  $p$ ، تعداد توصیف کننده های وارد شده در مدل MLR و شبکه عصبی مصنوعی می باشد. نتایج نشان می دهد که ضریب همبستگی تا حدودی برای مدل غیر خطی بهتر از مدل خطی است و این بیانگر این است که مقادیر پیش بینی شده با مدل غیر خطی در اطراف خط رسم شده متمرکزتر هستند. مقدار مجذور خطای استاندارد نیز برای مدل شبکه ی عصبی مصنوعی از ۹۹ مدل خطی کمتر است. نتایج در نهایت نشان می دهد که مدل طراحی شده با استفاده از شبکه ی عصبی مصنوعی در پیش بینی خاصیت مورد بررسی توانایی بیشتری دارد.

۳-۴- بررسی ارتباط توصیف کننده های ساختاری وارد شده در مدل با اندیس

### بازداری هیدروکربنهای آروماتیک چند حلقه ای

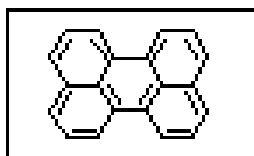
در این قسمت با توجه به توصیف کننده های وارد شده در مدل های ANN و MLR یک بررسی

اجمالی روی اثرات مختلف موجود در اندیس بازداری ترکیبات مورد مطالعه صورت خواهد گرفت. بهترین مدل انتخاب شده شامل چهار توصیف کننده است که هر کدام بیانگر خصوصیات دو بعدی، هندسی و سه بعدی مولکولهای مورد بررسی است. اولین توصیف کننده وارد شده در مدل با بزرگترین اثر متوسط SNar از گروه توصیف کننده های توپولوژی است. این گروه از دسته ی توصیف کننده های دو بعدی هستند که از گراف مولکولی حاصل می شوند. گراف مولکولی چگونگی پیوند یافتن اتمها در مولکول، نوع پیوند و همچنین همچنین برهمکنش های اتمهای خاص را توصیف می کند.

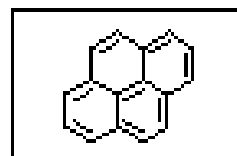
این توصیف کننده بیانی از اندازه، شکل، تقارن، شاخه ای بودن و میزان حلقوی بودن ترکیب است. در جدول ۴-۸ مقادیر این توصیف کننده برای چند هیدروکربن آروماتیک با مقادیر اندیس بازداری آنها مقایسه شده است. شکل های ۴-۱۱ (الف، ب، ج و د) ساختار این هیدروکربن ها را نشان می دهد.

هرچه مولکول تعداد حلقه های بیشتری دارد و شاخه ای تر است مقدار توصیف کننده ی SNar آن بیشتر است این توصیف کننده با اندیس بازداری ارتباط مستقیم دارد. با افزایش تعداد شاخه ها و تعداد حلقه ها میزان برهمکنش مولکول با ستون کروماتوگرافی افزایش یافته و زمان بازداری ترکیب افزایش می یابد.

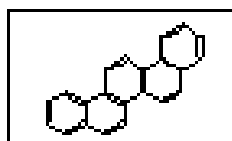
توصیف کننده ی بعدی Mor06m از گروه توصیف کننده های سه بعدی مورس است. این توصیف کننده با مقادیر جرم اتمی وزن دار شده است. با افزایش جرم اتمی مولکول مقدار این توصیف کننده افزایش می یابد. با افزایش این توصیف کننده مقدار اندیس بازداری افزایش می یابد. در جدول ۴-۸ مقادیر این توصیف کننده برای چند ترکیب PAH آمده است.



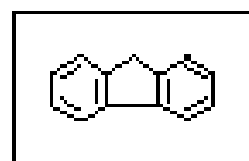
(ب) پریلن



(الف) پیرن



(د) فلوئورن



(ج) پیسن

شکل ۴-۱۱- ساختار چند ترکیب آروماتیک چند حلقه ای

جدول ۴-۸- مقدار SNar, Mor06m و اندیس بازداری برای تعدادی از ترکیبات

	فلوئورن	پیرن	پریلن	پیسن
SNar	11/038	13/523	17/107	18/493
Mor06m	0/868	1/823	2/756	3/112
RI	270/39	352/41	457/98	۵۰۰

سومین توصیف کننده GATS1e از گروه توصیف کننده های خود ارتباطی دوبعدی ۲D و از گروه Geary Autocorrelation است. این توصیف کننده نقش خاصیت فیزیولوژیکی الکترونگاتیویته را در مولکول نشان می دهد. رابطه ی ۴-۶ نحوه ی محاسبه ی این توصیف کننده را نشان می دهد.

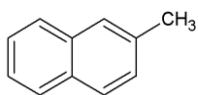
$$C(d) = \frac{\frac{1}{2\Delta} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - w_j)^2}{\frac{1}{A-1} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad (۴-۶)$$

در رابطه ی (۴-۶)، مقدار الکترونگاتیویته اتم،  $\bar{w}$  مقدار میانگین آن روی مولکول، A تعداد اتم و d فاصله ی توپولوژیک می باشد. با توجه به اثر متوسط این

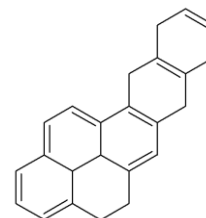
توصیف کننده تاثیر چندانی در اندیس بازداري ندارد. با تغيير در ساختار تركيبات تغييرات اين توصيف کننده ناچيز است. به هر حال اين توصيف کننده با انديس بازداري ارتباط مستقيم دارد.

چهارمين توصيف کننده RPCG از گروه توصيف کننده هاي بار بوده و بياني از ميزان بار مثبت در مولكول است. توصيف کننده هاي بار از طريق روش هاي كوانتوم مولكولي قابل محاسبه هستند. ميزان اثر اين توصيف کننده نيز در انديس بازداري هيدروكربن هاي آروماتيك چند حلقه اي ناچيز و ارتباط آن مستقيم است [۷۰].

شكل هاي ۴-۱۲ (الف، ب، ج و د) تصوير چند هيدروكربن آروماتيك چند حلقه اي را نشان مي دهد.

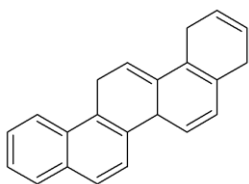


۱۰۲

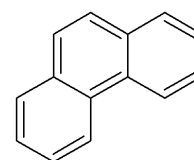


شكل ۴-۱۰ (ب) متيل نفتالن

شكل ۴-۱۰ (الف) نفتو[2,3-a]پيرن



شكل ۴-۱۰ (د) پيسن



شكل ۴-۱۰ (ج) فنانترن

شكل ۴-۱۲- ساختار تعدادي از تركيبات PAH

مقادیر توصیف کننده ی RPCG ، GATS1e و اندیس بازداری برای تعدادی از هیدروکربنهای آروماتیک چند حلقه ای در جدول ۴-۹ آمده است.

جدول ۴-۹-مقادیر توصیف کننده های RPCG ، GATS1e و اندیس بازداری برای چند PAH

	فنانترن	پیسن	متیل نفتالن	نفتو پیرن
GATS1e	1/516	1/432	1/736	1/362
RPCG	0/100	0/072	0/113	0/072
RI	300/00	500/00	221/10	551/10

از جمله مسائلی که در کمومتریکیس و مطالعات QSPR و QSAR وجود دارد این است که مدل به دست آمده در پیش بینی خاصیت مورد بررسی برای ترکیباتی که کاملاً جدا از سری ترکیبات مورد استفاده هستند، تا چه اندازه موفق است از این جهت می توان یک محدوده ی کاربرد پذیری<sup>۱</sup> برای مدل طراحی شده تعریف کرد که خارج از این محدوده مدل مربوطه در پیش بینی خاصیت مورد نظر

<sup>۱</sup>-Application Domain

ناتوان است. در این راستا می توان از روش های دسته بندی<sup>۱</sup> در کنار روش های توسعه ی مدل بهره گرفت.

از آنجایی که در این تحقیق از توصیف کننده های منتخب در مدل MLR به عنوان ورودی شبکه ی عصبی استفاده شده است لذا می توان از روشهایی چون الگوریتم ژنتیک (GA) و الگوریتم طرح ریزی متوالی<sup>۲</sup> (SPA) به منظور انتخاب توصیف کننده استفاده کرد. به جای شبکه ی عصبی مصنوعی می توان از روش های SVM<sup>۳</sup> و LS-SVM<sup>۴</sup> استفاده کرد.

## فهرست منابع لاتین

- [1]-Li-Bin Liu, Yuki HASHI, Min Liu, Yanlin WET and Jin-ming LIN. Analitical science June 2007" Determination of Polycyclic aromatic hydrocarbons in urban Air of Beijing by GC/MS", vol 23.
- [2]-European communities 2001" Ambient Air Pollution by PAH", *position paper –July 27 -2001.* ۱۰۴
- [3]-Int .j. Environ Sci. Tech. (2005)"vol2, No3.pp 275-286.
- [4]-U.S. Department of Health and Human Servers. (2000)"PAH prepared by the Wisconsin department of health and family serices division of public health". POH 4606.

---

<sup>۱</sup>-Clustering

<sup>۲</sup>- Successive Projection Algorithm

<sup>۳</sup>- Supported Vector Machine

<sup>۴</sup>-Least Squares Supported

- [5]- Gerhard Scherer, Stefanis frank, Kirsten Riedel Irintrud Meger-Kossion and Thomas Ranner. (2000)"Biomontoring of exposure to PAHs of nano occupationally exposed persons".vol 373-380.
- [6]-Canadian soil quality Guidelines. (2008)"Carcinogenic and other PAHs".PN1401-15BN 978-1-896997-79-7 PDF
- [7]-Josef Cvacka, Jiri Barek, Arnold G.fogg.JosinoC.Moreira and Jiri Zima. (1998)"High performance liquid chromatography of nitrated PAHs". *Analyst*, vol 133(9R-18R).
- [8]-Thomas J.McDonald.Bowang, Susanne J.McDonaldan James M.Brooks."Quantitative Dtermination of aromatic hydrocarbon using ion monitoring GC/MS".TID-Brooks International/B&B Laboratories Colleg station Texas 77845.
- [9]-Ronald A.Hites.Indiana University." Gas chromatography Mass spectrometry". *Chapter 31*.
- [10]-OlaEngkvist-Piotr Borowski, AgnetBemgard Gunnar karlstrom Rolahd Lindh and Anders Colmsjo. (1996)"On the relation between retention index and interaction between the solute and the column in Gas-Liquid chromatography".*J.chem .Inf Compu Sci* 1996, 36, 1153-1161.
- [11]-N. Adler, D.Babic, and N.Trinajstic. (1985)"On the calculation of the HPLC parameters for PAHs". *Fresenius Z Anal chem.* 332: pp: 426-429.
- [12]-Heszalc MARY NOWSKI,Marta Peta and Janusz JANECZEC.(2000)"Composition and source of polycyclic aromatic compound in deposited dust from selected sites around the upper Silesia,Poland".*Geological Quarterly* 2004,48(21:169-180).
- [13]-Fabiana Alvesdelima Ribeiro, Marcia Miguel Gastero Ferreira.(2003)"QSPR model of boiling point,octanol-water partition coefficient and retention index of PAHs".*Jurnal of molecular structure* 663,page 109-126.
- [14]-Rang. Jing Hu.HuanLin, Rui-sheng Chun-Xia Xu Xiao-Jun Yao, Man-Cang Liu. (2005)"QSPR prediction of GC retention index forfruitrogen-contuning PAH from heuristically compouted molecular descriptor".*talanta –vol 68-155ue1.15,page* 109-126.
- [15]-Skrbic, B Onjia, A. (2006)"Prediction of the Lee retention index of PAHs by ANN". *Journal of chromatography. A*.vol, 1108(no.2).
- [16]-Xu HY Zou Jw, jiang YJ HuGx YuQs. (2008)"Quantitative structure-chromatographic RI for PAHs".*Jurnal chromatogram A, journal of women's Health*.
- [17]-Wold, S.Chemometrics: What do we mean with it, and what do we want from it? S.D, Has the chemo metrics revolution ended? Some views on he past, present and future of Chemometrics, paper on chemo metrics: Philosophy, History and Direction of InCINC '94,  
[http:// www.emsl .pnl.gov:2080/docs/incinc/homepage.html](http://www.emsl.pnl.gov:2080/docs/incinc/homepage.html)



- [18]-Howery, D.G.Hirsch, R.F. (1983)"chemo metrics in the chemistry curriculum". *J.chem.Ed.***60**, pp: 656-659.
- [19]-Zupan, J., Gasteiger, J. (1991)"Neural network –a new method for solving chemical problem or just a passing phase". *Anal .Chim.Acta*, 248, pp: 1-30.
- [20]-Anderson, G.G.Kaufmann, P. (2000)"Development of a generalized neural network". *Chemometr.Intel.Lab.Syst.***50**, pp: 101-105.
- [21]-Hilbert, D.B. (1993)"Genetic algorithms in chimestry". *Chemometr.Intel.Lab.Syst*, 19, pp: 277-293. JL24-KM5F-76CA-A3XP
- [22]-M. Arab Chamjangali, M. Beglari, G. Bagherian, (2007), "Prediction of cytotoxicity data(CC50) of anti-HIV 5-phenyl-1-phenyl amino-1H-imidazole derivatives by artificial neural network with Levenberg-Marquardt algorithm", *Jurnal of Molecular Graphics and Modeling*,**26**, pp 360-367.
- [23]-M. Arab Chamjangali, M. Beglari, G. Bagherian, (2009), "Prediction of cytotoxicity data (CC50) of anti-HIV 1-[5-chlorophenyl Sulphonyl]-1H-Pyrrole Derivatives Using Calculated Molecular Descriptors and Levenberg-Marquardt Artificial Neural Network", *Chem. Biol Derug Des* 2009
- [24]- C. Hansch, T. Fujita, (1964), "ρ-σ-π Analysis: a method for the correlation of biological activity and chemical structure", *J. Am. Chem. Soc*; 86, 1616.
- [25]-T. Clark,(1985) "A Handbook Computational Chemistry", John Wiley & Sons,New York.
- [26]-Karelson, M. (2000)"Molecular descriptors in QSAR/QSPR".Wiley-VCH, New York.
- [27]-R.Todeschini, V.Consonni. (2000)"Hand book of Molecular Descriptor ".Wiley\_VCH, Weinheim
- [28]-Hammet, LP. (1970)"Physical organic chemistry. Reaction rate, Equilibrium and Mechanism ".*2th ed.McGraw-Haill, New, New York*.
- [29]-R.Todeschini, V.Consonni. (2000)"Handbook of Molecular Descriptor ".Wiley\_VCH Weinheim.,
- [30]-Burden, F.R. (1989)"Molecular identification number for substructure searches". *J.Chem.Inf.Comput.Sci.***29**, pp: 225-227.
- [31]-Burden, F.R. (1997), *Quant.Struct.Act.Relat.***16**, pp: 309-314.
- [32]-Pearlman, R.S., Smith, K.M. (1998)"Metric Validation and the Receptor-Relevant Subspace Conceot". *J.Chem.Inf.Comput.Sci*, 39, pp: 28-35.

[33]-Krygowski, T.M., Ciesielski, A., Bird, C.W., Kotschy, A. (1995) "Aromatic character of the Benzene ring present in various topological environments in Benzenoid hydrocarbons. Nonequivalence of indices of aromaticity". *J.Chem.Inf.Comput.Sci.* 35, pp: 203-210.

[34]-Krygowski, T.M., Cyranski, M., Ciesielski, A., Swirska, B., Leszczynski. (1996) "Separation of the energetic and geometric contributions to aromaticity. 2. Analysis of the aromatic character of benzene ring in their various topological environments in the benzenoid hydrocarbons. Crystal and molecular structure of coronene". *J.Chem.Inf.Comput.Sci.* 36, pp: 1135-1141.

[35]-Jug, K., Croat. (1984), *Chem.Acta.* 57, pp: 941-953.

[36]-Bird, C.W. (1986) "The application of a new aromaticity index to six-membered ring heterocycles, Tetrahedron". 42, pp: 89-92.

[37]-Silverman, B.D. (2000) "Three-Dimension moment of molecular property fields". *J.Chem.Inf.Comput.Sci.*, 40, pp: 1470-1476.

[38]-Randic, M. (1995). *New J.Chem.* 19, pp: 781-791.

[39]-Randic, M. (1995) "Molecular Shape Profiles ". *J.Chem.Inf.Comput.Sci.*, 35, pp: 373-382.

[40]-Randic, M., Razinger. (1995) "On characterization of molecular shapes". *J.Chem.Inf.Comput.Sci.*, 35, pp: 594-606.

[41]-Randic, M., Kleiner, A.F., DeAlba, L.M. (1994) "Distance/Distance Matrixes". *J.Chem.Inf.Comput.Sci.*, 34, pp: 277-286.

[42]-Gasteiger, J., Sadowski, J., Schuur, J., Selzer, P., Steinhauer, L., Steinhauer, V. (1996) "Chemical information in 3D space". *J.Chem.Inf.Comput.Sci.*, 36, pp: 103-107.

[43]-Hemmer, M., Steinhauer, V., Gasteiger, J. (1999) "The application of the 3D structure of organic molecules from their infrared spectra". *Vibrat.Spectrosc.*, 19, pp: 151-164.

[44]-J., Gasteiger, J., Sadowski, P., Selzer, P., Steinhauer, V. (1996) "Chemical information in 3D space". *J.Chem.Inf.Comput.Sci.*, 36, pp: 1030-1037.

[45]-J.H.Schuur, P.Selzer, J.Gasteiger. (1996) "The coding of the three-Dimensional structure of molecules by molecular transforms and its application to structure-spectra Correlation and studied of biological activity". *J.Chem.Inf.Comput.Sci.*, 36, pp: 334-344.

[46]-R.Todeschini, P.Grammatica. (1997), *Quant.Struct.Act.Relat.*, 16, pp: 113-119.

[47]-R.Todeschini, P.Grammatica, E.Marego, and R.Provenzani. (1996) "Modeling and prediction by using WHIM descriptors in QSAR studies: submitochondrial particles (SMP) as toxicity biosensors of chlorophenols". *Chemosphere*, 33(1), pp: 71-79.

- [48]-R.Todeschini, M.Lasagni, E.Marengo .(1994), *J.Chemom* .,8,pp:263-273.
- [49]-R.Todeschini, M.Vghi, R.Provenzani, A.Finzio, P.Grammatica. (1996)"Modeling and prediction by using WHIM descriptors in QSAR studies: toxicity of heterogeneous chemical on *Daphnia magna*". *Chemosphere*, 32(8), pp: 4527-1554.
- [50]-V.Consonni and R.Todeschini. (2001)"Rational approaches to drug design ".Porous Science, Barcelona.
- [51]-V.Consonni, R.Todeschini, and M.Pavan. (2002)"Structure/response correlation and Similarity/Diversity Analysis by GETAWAY descriptors.1.Theory of the novel 3D molecular descriptors". *J.Chem.Inf.Comput.Sci* , 42(3), pp: 682-692.
- [52]-V.Consonni, R.Todeschini, M.Pvan, and P.Grammatica. (2002)"Structure/Response Correlation and Similarity/Diversity Analysis by GETAWAY Descriptors.2.Application of the novel 3D molecular descriptors to QSAR/QSPR studies". *J.Chem.Inf.Comput.Sci*, 42(3), pp: 693-705.
- [53]-Lapin, L.L. (1990)"probability and statistics for modern engineering".PWQ-KENIT, New York.
- [55]-Pitts, W., MacCullouch, S. (1947), *Bull.Math.Biophysic*.9, pp: 65.
- [56]-Hebb, D.O., (1949): the organization of behavior ".Wiley, New York.
- [57]-Hassain, A.S., Yu, X.Q., Johnson, R.D., (1991), *Pharm.Res.*,8,pp:1248.
- [59]-Zupan, J., (1994)"Introduction to artificial neural network method: What they are and how to use them". *Acta Chemical Slovenica*, 41(3), pp: 327-352.
- [60]-<http://www.doc.ic.ac.uk/~nd/surprise-96/journal/vol4/csll/report.html>
- [61]-MAT LAB Tool Box
- [62]-Rumelhart, D.E., Hinton, G.E., Williams, R.J., (1986)"Learning internal representations by error backpropagation , in distributed parallel processing :Explorations in the microstructures of Cognition, Eds.Rumelhart ,D.E., MacClelland, J.L" ., Vol.1.MIT Press, Cambridge, MA, USA, PP:318-362.
- [63]-Lipmann, R.P., April (1987)"An introduction to computing with neural nets", *IEEE ASSP Magazin*.pp:155-162.
- [64]-Svovil, D., Kvasnicka, V., Pospichal. (1997)"Introduction to multi-layer feed-forward neural networks". *J.Chem.Intell.Lab.Sys*, 39, pp: 43.
- [65]-Agatonovic Kustrin, S., Ling, H.L, Tham, S.Y., R.G., Alany. (2002)"molecular descriptors that influence the amount of drugs transfer into human breast milk, *J.Pharmac.Biomed*". *Anal*.29, pp:103.

[66]-Hagan M. T., and Menhaj M. (1994), "Training feed-forward networks with the Marquardt algorithm," IEEE Transaction on Neural Network, Vol. 5, No. 6, 1999, pp:989-993.

[67]- K. Levenberg, "A method for the solution of certain problems in least squares, *Quart. Appl. Math.*, 1944, Vol. 2, pp. 164-168.

[68] -D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, 1963, Vol. 11, pp. 431-441.

[69] - Zhendi Wang\*, K. Li, P. Lambert, Chun Yang "Identification, characterization and quantification pyrogenic PAHs and other organic compounds in tire fire products" *Jurnal of chromatography A* .1139(2007)14-26.

[70]-Roberto Todeschini, Consonni,(2002), "Hand book of Molecular Descriptor", WILEY-VCH.

۵۴-خراسانی زاده، ع.، ۱۳۷۵، مقدمه ای بر نرم افزار SPSS در پژوهش های آماری، تهران، انتشارات قائم.

۵۸-منهاج، م.، ۱۳۸۴، مبانی شبکه های عصبی، تهران، انتشارات دانشگاه امیر کبیر.

### Abstract

In this work, quantitative structure-property relation ship (QSPR) has been applied for modeling and predicting the retention index of a series of polycyclic aromatic hydrocarbons (PAHs). The Polycyclic Aromatic Hydrocarbons (PAHs) have been for a long time, a focus of greater attention by the scientific community due to their impact on public health and the environment. Usually, the PAHs are introduced into the environment as a result of anthropogenic activities, which have increased dramatically in the last 20 years. PAHs are a class of several hundred individual compounds containing at least two condensed ring. PAHs are product of incomplete combustion and

the pyrolysis of fossil fuel and other organic material from natural and anthropogenic source. Chromatographic techniques such as gas chromatography (GC) and high performance liquid chromatography (HPLC) coupled with UV, FID and MS detectors are the most frequently used. Retention index is a parameter that used in chromatography for separation of hydrocarbon mixture and provides the elution order of each compound. This order due of interaction between the mobile and stationary phase. This interaction related to the molecular structure of both phases and the variations in RI result from the structural variation of the compounds. Quantitative structure-properties relationships (QSPR) have extensively been used to explain separation mechanisms and prediction of retention behavior in chromatography. In the present work, we developed a QSPR model using artificial neural network (ANN) and multiple linear regressions (MLR) to predict the retention index of some PAHs compounds. The results of this study also demonstrate the underlying factors governing on the retention index of test molecules. The experimental retention indexes of 17 polycyclic aromatic hydrocarbons were taken from the published experimental paper. All molecules were drawn into HyperChem and pre-optimized using semi-empirical AM1 method. The resulted geometry was transferred into software Dragon that can calculate the different descriptors.. The result obtained reveals the superiority of ANN over the MLR model.

**Keywords:** QSPR, Retention Index, ANN, MLR, PAHs