

## ۱) مقدمه

### ۱-۱) زمان بازداری

یکی از پارامترهای مهم در شناسایی ترکیبات مختلف از جمله ترکیبات آلی، زمان بازداری ترکیبات می باشد که بطور معمول توسط اطلاعات حاصل از کروماتوگرافی گازی بدست می آیند. زمان بازداری یک پارامتر بسیار مهم برای شناسایی ترکیبات آلی است. در این پروژه سعی خواهد شد تا با استفاده از روش های کمومتری مدلی ارائه شود که بتوان توسط آن پارامتر زمان بازداری را پیش بینی نمود.

### ۲-۱) مروری بر روش های کمومتریکس

برای درک مکانیسم فرآیندهای مختلف شیمیایی، کشف و توسعه مواد جدید، حفظ محیط زیست و زمینه های دیگر شیمی هنوز توانایی حل مسائل به طور کامل وجود ندارد و برای عملی کردن بعضی از مسائل نیاز به سیستم های بسیار پیچیده ای می باشد که انجام آنها در گرو صرف هزینه های بسیار و مطالعات گسترده است. در جهت حل این مشکل، روشهای محاسباتی کمومتریکس می توانند مفید باشند. تجزیه و تحلیل آماری و ریاضی داده های شیمیایی، تحت عنوان کمومتریکس یاد می شود [۱]. به عبارتی کمومتریکس یک روش کارآمد برای خلاصه کردن اطلاعات مفید از یک سری داده مشخص و پیش بینی سری دیگر داده هاست. هدف کمومتریکس، بهبود بخشیدن فرآیندهای اندازه گیری و استخراج اطلاعات شیمیایی مفیدتر از داده های اندازه گیری شده فیزیکی و شیمیایی می باشد [۲]. کمومتریکس اولین بار توسط دانشمند سوئدی به نام ولد<sup>۱</sup> [۳] در سال ۱۹۷۲ به کار گرفته شد و توسط کووالسکی<sup>۲</sup> توسعه داده شد [۴].

---

<sup>۱</sup> – Wold

<sup>۲</sup> – Kowalski

چندین تعریف برای کمومتریکس بیان شده است که غالباً در متن های تجزیه ای به کار می روند. یکی از جامع ترین تعاریف به صورت زیر است:

کمومتریکس شاخه ای از شیمی است که از ریاضی و آمار و منطق استفاده می کند برای اینکه: الف) فرآیندهای تجربی بهینه را طراحی و انتخاب کند، ب) حداکثر اطلاعات شیمیایی قابل حصول را از تحلیل اطلاعات شیمیایی فراهم کند و پ) بتوان اطلاعات بیشتری در مورد سیستم های شیمیایی بدست آورد [۵].

کمومتریکس مجموعه ای از یک سری قواعد شیمیایی است که با استفاده از ریاضی، آمار و کامپیوتر در عرصه های ارزیابی و تفسیر اطلاعات، بهینه کردن و مدلسازی فرآیندها و آزمایشات و استخراج حداکثر اطلاعات شیمیایی از داده های تجربی به ما کمک می کند. به طور کلی می توان گفت کمومتریکس، علم استفاده از کامپیوتر، ریاضی و آمار در شیمی می باشد.

کمومتریکس به عنوان یک شاخه علمی جوان در دو دهه اخیر به سرعت توسعه پیدا کرده است. این رشد سریع مدیون پیشرفت دستگاههای هوشمند و خودکار آزمایشگاهی و همچنین امکان استفاده از کامپیوترهای قدرتمند و نرم افزارهای ساده است. بنابراین کمومتریکس به عنوان یک وسیله در همه قسمتهای شیمی و به طور زیادی در زمینه شیمی تجزیه مورد استفاده قرار گرفته است. امروزه یک شیمیدان تجزیه به صورت فزاینده ای به استفاده از روشهای آماری و ریاضی در کارهای روزمره اش به کمومتریکس نیاز پیدا می کند [۶].

امروزه کمومتریکس در شاخه های مختلف شیمی مورد استفاده قرار گرفته است که برخی از آنها عبارتند از: کنترل فرآیندها، تجزیه و تحلیل و شناخت الگوها، پردازش علائم، بهینه کردن شرایط [۷]. یکی از زمینه های مهم کاربرد کمومتریکس در مطالعاتی است که خواص مولکول ها را به ویژگی های ساختاری آنها نسبت می دهد. از نظر شیمیدانان فعالیت ها و خواص یک ترکیب ناشی از ویژگی های ساختاری آن است. این نوع از مطالعات به بررسی کمی ارتباط ساختمان با فعالیت، QSAR، و همچنین بررسی کمی ارتباط ساختمان با ویژگی (QSPR)، معروف می باشد. هدف از

مطالعات QSAR پیدا کردن رابطه ای است که بین رفتار فیزیکی شیمیایی یک مولکول با پارامترهای ساختاری آن وجود دارد.

نتایج این مطالعات علاوه بر شفاف سازی نحوه ارتباط بین خواص ملکول ها و ویژگی های ساختمانی آنها به پژوهشگران در پیش بینی رفتار مولکول های جدید براساس رفتار مولکول های مشابه کمک می کند. به مجموعه ابزارها و روشهایی که به این منظور مورد استفاده قرار می گیرند روش های پارامتری گویند. در روش های پارامتری سعی می شود بین یک سری توصیف کننده های مولکولی با فعالیت یا خاصیت مورد نظر ارتباط منطقی برقرار نمایند. توصیف کننده های مولکولی که به این منظور استفاده می شوند، مقادیر عددی می باشند که جنبه های مختلف ساختاری مولکول را به طور کمی نشان می دهند. مقاله های فراوانی در مورد استفاده از روش های پارامتری در مطالعات QSAR تا کنون منتشر شده اند. برخی از مقالات منتشر شده در این زمینه به قرار زیر است:

- پیش بینی خواص سمی برخی از ترکیبات با کروماتوگرافی مایسلی توسط گیلبرت<sup>۱</sup> و همکارانش در سال ۲۰۰۱ [۸].

- مطالعه یک سری از آنالوگ های فنوکسی تک استخلافی ارتو توسط فوماگالی<sup>۲</sup> و همکارانش [۹].

- مدلسازی و پیش بینی تحرک الکتروفوریتیک مشتقات بنزوئیک اسید در حلالهای آبی - الکلی مختلف در الکتروفورز موئینه توسط فاطمی<sup>۳</sup> و همکارانش [۱۰].

---

1. Gillabert    2. Fumagalli  
3. Fatemi

- پیش بینی اندیس بازداری یک سری از ترکیبات سمی در کروماتوگرافی گازی توسط گرکانی نژاد و همکارانش [۱۱].

### ۳-۱) روش های پارامتری

روشهای پارامتری، مجموعه تکنیکهایی است که برای مدل سازی و پیش بینی فعالیتهای مولکولی مورد استفاده قرار می گیرند. با استفاده از این روش ها می توان ارتباط منطقی بین فعالیت یا خاصیت موردنظر با تعدادی از توصیف کننده های مولکولی برقرار نمود. توصیف کننده های مولکولی، متغیرهای مستقلی هستند که ویژگی های مولکول را به طور کمی بیان می کنند. معمولاً از روش رگرسیون خطی چندگانه جهت ایجاد این روابط استفاده می شود.

در رگرسیون خطی چندگانه، ارتباط بین توصیف کننده ها به عنوان متغیر مستقل و فعالیت یا ویژگی به صورت یک متغیر وابسته در قالب یک معادله ریاضی طرح ریزی می شود. روش های آماری در مورد ارتباط بین متغیرهای وابسته و مستقل کاربرد وسیعی دارند و نرم افزارهای آماری متفاوتی جهت تسریع و سهولت انجام عملیات محاسباتی مورد استفاده قرار می گیرند. بعد از معرفی متغیرهای مستقل و وابسته، براساس روش های رگرسیون خطی چندگانه، معادلات خطی ارتباط دهنده بین این دو متغیر ایجاد می گردد. شکل کلی این معادلات خطی به صورت زیر می باشد [۱۲].

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (1)$$

که در این معادله،  $y$ ، فعالیت یا ویژگی،  $a_0$  مقدار ثابت (عرض از مبدا)،  $x_i$  مقادیر عددی توصیف کننده ها، ضرایب  $a_0$  تا  $a_n$  ضرایب رگرسیون و  $n$  تعداد توصیف کننده ها می باشد.

قابل ذکر است که متغیرهای مستقل براساس اصول علمی که انتظار می رود با متغیر وابسته در

ارتباط باشند، انتخاب می شوند. مراحل کلی مدل سازی به روش پارامتری به ترتیب زیر می باشد:

۱- انتخاب سری داده ها

۲- انتخاب و محاسبه توصیف کننده ها

۳- تجزیه و تحلیل آماری توصیف کننده ها

۴- آنالیز مدل های آماری و انتخاب مدل مناسب

۵- ارزیابی اعتبار مدل های انتخاب شده

### ۱-۴) انتخاب سری داده ها

اولین مرحله ، جمع آوری و انتخاب یک سری مولکولی است که مقادیر تجربی خاصیت موردنظر آنها در دسترس باشد. در صورت زیاد بودن تعداد آنها، سری داده ها به سه قسمت سری آموزشی<sup>۱</sup> و سری آزمون<sup>۲</sup> و سری ارزیابی<sup>۳</sup> تقسیم می شوند. عملیات مدل سازی بر روی سری آموزشی که اکثر مولکول ها را در بر می گیرد، انجام می شود. سری آزمون و سری ارزیابی به نحوی انتخاب می شوند که نماینده جمعیت مولکول های سری آموزشی باشند. لازم به ذکر است که مولکول های سری ارزیابی در هیچ یک از مراحل مدل سازی شرکت ندارند. در واقع سری ارزیابی برای ارزیابی اعتبار مدل های حاصله است. هر قدر که سری داده ها بزرگتر و متنوع تر باشد مدل حاصل از آن معتبرتر بوده و قدرت پیش بینی بالاتری خواهد داشت. البته لازم به یادآوری است که این تقسیم بندی برای زمانی است که تعداد ترکیبات در سری داده ها زیاد باشند. اگر تعداد ترکیبات در سری داده ها کم باشد، برای ارزیابی مدل از روش ارزیابی تقاطعی<sup>۴</sup> استفاده می شود.

### ۱-۵) محاسبه توصیف کننده ها

توصیف کننده ها پارامترهای عددی هستند که ویژگی های متفاوتی از مولکول را بیان می کنند. به عبارت دیگر، هر یک از این توصیف کننده ها، اطلاعات خاصی از مولکول را در اختیار می گذارند. تعیین و انتخاب توصیف کننده ها از اهمیت به سزایی برخوردار است. در صورتی که توصیف کننده های انتخاب شده مناسب نباشند، عمل مدل سازی با مشکل مواجه شده و مدل مناسب

---

<sup>1</sup> – Training set

<sup>2</sup> – Test set

<sup>3</sup> – Validation set

<sup>4</sup> – Cross- validation

به دست نخواهد آمد. بنابراین در انتخاب توصیف کننده ها باید دقت کرد و توصیف کننده هایی می بایست انتخاب شوند که در ارتباط با ویژگی خاص مولکول باشند.

توصیف کننده ها از نظر بیان چگونگی خصوصیات مولکول به دسته های متفاوتی تقسیم می شوند که معروف ترین آنها در اینجا آورده شده است:

- توصیف کننده های توپولوژیکی<sup>۱</sup>
- توصیف کننده های هندسی<sup>۲</sup>
- توصیف کننده های الکترونی<sup>۳</sup>
- توصیف کننده های فیزیکوشیمیایی<sup>۴</sup>
- توصیف کننده های الکتروتوپولوژیکی<sup>۵</sup>

### ۱-۵-الف) توصیف کننده های توپولوژیکی

این توصیف کننده ها جزء ساده ترین نوع توصیف کننده ها می باشند و به ساختار فضایی مولکول ارتباطی نداشته و تنها به نوع اتم، نوع پیوندها و نحوه ارتباط اتم ها به یکدیگر وابسته است. بنابراین این پارامترها را بدون بهینه کردن ساختار ملکول می توان محاسبه نمود. تعداد اتمها، تعداد هالوژن ها، تعداد پیوندهای ساده یا دوگانه، اندیس های ارتباطی مثال هایی از این نوع توصیف کننده ها هستند. توصیف کننده های توپولوژیکی را به چهار دسته فرعی تقسیم می کنند. این چهار گروه عبارتند از: توصیف کننده های جزء، اندیس های توپولوژی، توصیف کننده های زیرساختاری و توصیف کننده های محیطی.

---

<sup>1</sup> – Topological Descriptors

<sup>2</sup> – Geometric Descriptors

<sup>3</sup> – Electronic Descriptors

<sup>4</sup> – Physico - chemical Descriptors

<sup>5</sup> – Electrotopological Descriptors

## ۱-۵-ب) توصیف کننده های جزء

این گروه از توصیف کننده ها مشخص کننده اطلاعات ساده مولکولی بوده و از نظر محاسبه ساده ترین نوع توصیف کننده ها می باشند. مقدار توصیف کننده های فوق به سادگی از فرمول مولکولی ترکیب مورد نظر قابل محاسبه می باشد. این نوع از توصیف کننده ها می توانند جرم مولکولی، حضور یا عدم حضور گروههای عاملی خاص و یا اتمهای معینی را در مولکول نشان دهند. در جدول ۱ نام برخی از توصیف کننده های فوق آمده است.

جدول ۱: نام برخی از توصیف کننده های جزء

نشانه	توصیف کننده ها
MW	جرم ملکولی
NSB	تعداد پیوندهای ساده
NDB	تعداد پیوندهای دوگانه
NA	تعداد اتمها
NC	تعداد اتمهای کربن

## ۱-۵-پ) اندیس های توپولوژی

این اندیس ها دسته ای از توصیف کننده های توپولوژیکی هستند که اطلاعاتی راجع به ساختمان، اندازه و میزان شاخه ای شدن ملکول، نحوه اتصال اتم ها و نوع اتم های موجود در یک مولکول در اختیار ما قرار می دهند. جهت محاسبه این نوع از توصیف کننده ها ابتدا باید ساختار مولکول را به صورت یک گراف شیمیایی<sup>۱</sup> رسم نمود. تئوری گراف شیمیایی، شاخه ای از علم شیمی - ریاضی<sup>۲</sup> می باشد که عموماً با شیمی نظری در ارتباط است. براساس این تئوری یک گراف نشان دهنده مجموعه ای از عناصر یک گروه و نحوه ارتباط بین آنها می باشد. هر گراف از تعدادی

<sup>1</sup> - Chemical Graph

<sup>2</sup> - Mathematical Chemistry

راس<sup>۱</sup> و یال<sup>۲</sup> تشکیل شده است. یال ها خطوطی می باشند که راس ها را به هم متصل کرده اند. یک دسته از گراف ها، گراف های مولکولی هستند. در این گراف ها هر نقطه نشان دهنده یک اتم بوده و خطوط بین نقاط نیز نشان دهنده پیوند شیمیایی بین اتم هاست. در این نوع از گراف ها منظور از مسیر<sup>۳</sup>، فاصله ای است که در طول آن هیچ نقطه ای دو بار تکرار نشود. معمولاً در گراف های مولکولی، اتم هیدروژن را نشان نمی دهند. در جدول ۲ نام انواع توصیف کننده های توپولوژیکی که با استفاده از مفاهیم گراف مولکولی قابل محاسبه می باشند آمده است.

جدول ۲: نام برخی از توصیف کننده های توپولوژیکی

نشانه	نام اندیس
X	اندیس ارتباط ملکولی
Z	اندیس هوزویا
J	اندیس بالابان
W	عدد واینر
V	اندیس شولتز
H	اندیس هارای

### ۱-۵-۵ (ت) توصیف کننده های زیرساختاری

این توصیف کننده ها اطلاعاتی را در مورد وجود یا عدم وجود گروههای عاملی خاص یا زیرساختارهای معین در مولکول فراهم می سازند. اگر زیرساختار مشخصی در مولکول وجود داشته

- 
1. Vertex
  2. Edge
  3. Path



باشد، توصیف کننده مقداری معادل با تعداد دفعات تکرار آن زیر ساختار در مولکول را خواهد داشت و در غیر این صورت مقدار آن صفر خواهد شد.

### 1-5-ج) توصیف کننده های محیطی

این نوع توصیف کننده ها نحوه قرار گرفتن یک زیرساختار را در مولکول نشان می دهند. اگر یک زیرساختار خاص در مولکول وجود داشته باشد، آنگاه یک شبه مولکول توسط این زیرساختار و اولین اتم مجاور آن تشکیل می شود. مقدار توصیف کننده محیطی برای این زیرساختار خاص برابر با اندیس ارتباطی مولکول مسیر درجه اول برای شبه مولکول حاصل می باشد. در صورتی که یک زیرساختار معین چندبار در مولکول تکرار شده باشد، مقدار متوسط اندیس ارتباطی آنها در نظر گرفته می شود [۱۳].

### 1-5-چ) توصیف کننده های هندسی

توصیف کننده های هندسی نسبت به نوع توپولوژیک از اهمیت بیشتری برخوردارند. یکی از خصوصیات بارز این توصیف کننده ها، ارتباط آنها با ساختار سه بعدی مولکول ها می باشد. بنابراین قبل از محاسبه این توصیف کننده ها باید پایدارترین ساختار فضایی مولکول تعیین شود که بهینه کردن مولکول توسط روش های مکانیک کوانتومی انجام می شود. حجم واندروالس، مساحت سایه، فاکتور شکل<sup>۱</sup> از این نوع توصیف کننده ها می باشند. در اینجا فقط به اختصار نحوه محاسبه حجم واندروالس آمده است.

---

1- Shape factor

## 1-0-5 ح) محاسبه حجم واندروالس

الگوریتم محاسبه حجم واندروالس ترکیبات شیمیایی در سال ۱۹۸۶ توسط جرز<sup>۱</sup> ارائه شد که در آن هر ملکول در درون یک مکعب مستطیل قرار می گیرد بطوری که هر یک از وجوه این مکعب مستطیل با دانسیته ای مشخص شبکه بندی می شوند (شکل ۱). به عبارتی هر نقطه با مختصات  $x$  و  $y$  و  $z$  نمایانگر یک مکعب به ابعاد عکس دانسیته موردنظر است. تعداد نقاطی از مکعب مستطیل که در داخل مولکول موردنظر گرفته اند، متناسب با حجم مولکول می باشد. بنابراین باید تمامی نقاط موجود در شبکه بررسی شوند، به نحوی که مشخص گردد آیا این نقاط در داخل مولکول هستند یا در خارج آن. برای رسیدن به این هدف فاصله بین هر نقطه از شبکه و مرکز هر اتم محاسبه می شود. اگر این فاصله کمتر یا مساوی با شعاع واندروالسی اتم مذکور باشد نقطه در درون مولکول قرار دارد و اصطلاحاً یک نقطه on به حساب می آید و در غیر این صورت نقطه مذکور off تلقی می شود. فاصله موردنظر با استفاده از رابطه (۲) محاسبه می شود:

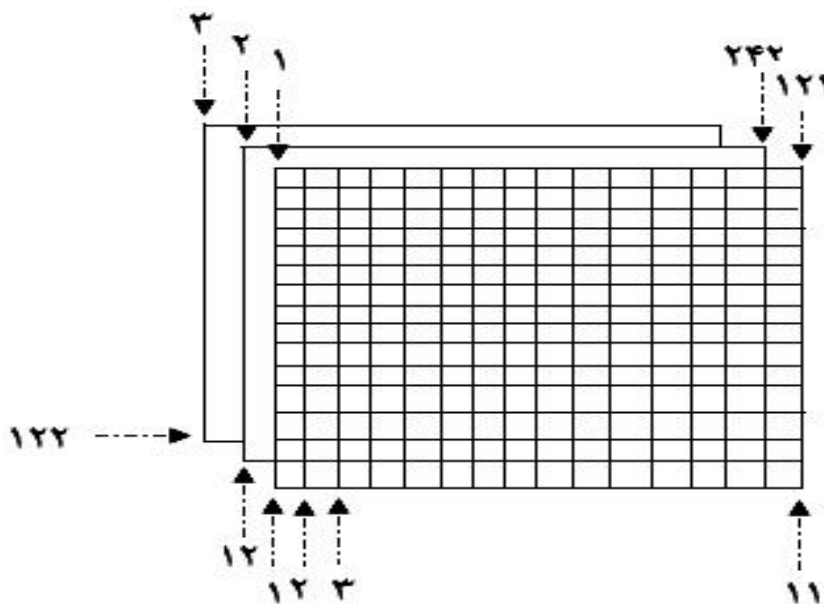
$$d^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 \quad (2)$$

$x$  و  $y$  و  $z$  مختصات یک نقطه از شبکه و  $x_i$  و  $y_i$  و  $z_i$  مختصات مرکز  $i$  امین اتم می باشد. در محاسبه حجم، همپوشانی کره های سخت با شعاع واندروالس مربوطه باید درنظر گرفته شود. برای این منظور چنانچه فاصله نقطه ای از شبکه تا مرکز یک اتم کمتر یا مساوی با شعاع واندروالس بود دیگر فاصله این نقطه تا مرکز سایر اتمها محاسبه نمی شوند. دقت محاسبه با دانسیته بکار رفته متناسب می باشد به طوری که هر چه دانسیته بیشتر باشد حجم محاسبه شده دقیق تر خواهد بود [۱۴].

## 1-0-5 خ) توصیف کننده های الکترونی

توصیف کننده های الکترونی بیانگر خواص الکترونی مولکول ها می باشند و اطلاعات مفیدی راجع به برهم کنش های الکترونی بین مولکولی و درون مولکولی ارائه می دهند. این نوع از

برهم کنش ها خواص فیزیکی و شیمیایی مولکول ها را شدیداً تحت تاثیر قرار می دهند. اهمیت این



شکل ۱: تصویر یک شبکه سه بعدی جهت محاسبه حجم

برهم کنش ها به ویژه هنگامی که مولکول های عمل کننده قطبی باشند، بیشتر می شود. با توجه به اینکه مولکول ها از تعدادی الکترون، پروتون و نوترون تشکیل شده اند، بنابراین بسیاری از رفتارهای مولکول را می توان با توجه به تاثیرات متقابل الکتریکی درک و توجیه نمود. اندازه و قدرت نیروهای حاصل از برهم کنش های الکترونی، تابعی از بار جزئی اتم ها، نحوه توزیع بار در مولکول، فاصله بین بارهای مخالف، گشتاور دوقطبی و قطبش پذیری مولکول ها می باشد. بنابراین دانستن مقادیر فوق برای یک مولکول می تواند در پیش بینی و توجیه رفتار آن در محیط مفید باشد.

هر نوع توصیف کننده ای که بتواند به نحوی یکی از عوامل موثر در برهم کنش های الکترونی را بیان کند، در گروه توصیف کننده های الکترونی قرار می گیرد. برای محاسبه این توصیف کننده ها عموماً از روش های اوربیتال ملکولی استفاده می شود. این روشها کلاً به دو گروه روش های نظری آغازین و نیمه تجربی<sup>۱</sup> تقسیم می شوند [۱۵]. در روش های نظری آغازین که امروزه

<sup>۱</sup> – Semi-empirical

به روشهای ab initio نیز معروفند، جهت محاسبه پارامترهای الکترونی ساختاری از حل دقیق معادله شرودینگر استفاده می کنند. روشن است که برای حل دقیق معادله شرودینگر ناگزیر از حل تعداد زیادی انتگرال پیچیده و دشوار هستیم. این کار علاوه بر پیچیدگی، زمان زیادی از وقت کامپیوتر را می گیرد. به طور مثال برای حل تابع موج یک مولکول متوسط باید بیش از  $10^6$  انتگرال و معادله درجه بالا را حل نمود. البته بسیاری از انتگرال های فوق مقادیری برابر یا نزدیک به صفر دارند. مشخص است که چنانچه روشی قادر به استفاده از تقریب های مناسب جهت حذف این انتگرال ها باشد بسیار مطلوب بوده و باعث صرفه جویی در زمان لازم برای محاسبات خواهد شد. بعلاوه حذف این انتگرال ها باعث می شوند که بتوان مولکول های بزرگتر را بررسی کرد.

تئوری های تقریب اوربیتال مولکولی اغلب از داده های تجربی مربوط به اتم ها و سیستم های مولکولی جهت تخمین کمیت های قراردادی که پارامترهای محاسباتی محسوب می شوند، استفاده می کنند. از این رو به این تکنیک ها، روشهای نیمه تجربی گفته می شود. یک روش نیمه تجربی در صورتی موفق خواهد بود که حائز شرایط زیر باشد:

- روش مورد استفاده باید حتی الامکان ساده باشد تا به خوبی بر روی تمام توابع موجی اوربیتال مولکولی موجود عمل نماید.

- تقریب های اعمال شده نباید نیروهای فیزیکی اولیه را که تعیین کننده ساختار مولکول هستند حذف کنند.

- تئوری مورد استفاده باید بتواند به خوبی نتایج را تفسیر نموده و برای رد یا قبول فرضیات کیفی کارآیی داشته باشد.

- تقریب مورد استفاده باید به دور از هر گونه تعصب و پیش داوری باشد. همچنین برای محاسبه یک کمیت نباید از مباحث کیفی قراردادی که در حین محاسبه بدست می آیند استفاده شود.

- تمام الکترونهاى موثر در عوامل شیمیایی مثل الکترونهاى پیوندی و الکترونهاى II در سیستم های مزدوج مسطح باید در تئوری منظور شوند.

پاپل<sup>۱</sup>، سانتری<sup>۲</sup> و سگال<sup>۳</sup> در سال ۱۹۶۵ با وارد کردن تقریب دیفرانسیل همپوشانی صفر به صورت منطقی، عناصر ماتریس و عملگر فوک<sup>۴</sup> (F) را توسط داده های تجربی جایگزین کردند. این محققین تقریب دیفرانسیل همپوشانی صفر را در سطوح مختلف بکار بردند و در نتیجه روش های نیمه تجربی متفاوتی مثل صرف نظر کامل از دیفرانسیل همپوشانی (MNDO) را ارائه کردند [۱۶].

در سال ۱۹۷۷ دوار<sup>۵</sup> و تیل<sup>۶</sup> در ادامه روند فوق روش MINDO را معرفی کردند [۱۷]. تجربه نشان داده است که این روش در برخی از موارد با نواقصی همراه است. مهمترین نقص این روش در عدم توانایی آن در ایجاد پیوندهای هیدروژنی است که مانع از کاربرد آن در موارد بیولوژیکی شده است [۱۶]. یکی از جدیدترین روش های نیمه تجربی روش AM1 است که امروزه به نحوه گسترده ای توسط پژوهشگران استفاده می شود. این روش اولین بار توسط دوار و همکاران در سال ۱۹۸۵ جهت رفع نواقص MNDO معرفی شد [۱۶]. در این روش انتگرال های دافعه الکترونی مشابه با روش MNDO بوده، جاذبه هسته - الکترون فرموله شده و انتگرال های دافعه الکترونی یک مرکزی از مقادیر اولری<sup>۷</sup> بدست می آیند. سایر پارامترهای مورد استفاده در این روش از داده های تجربی استخراج می گردند. در واقع جهت پارامتری کردن توابع مورد استفاده از یک مجموعه بزرگ از مولکول ها استفاده می شود. هر چه این مجموعه بزرگتر بوده و حاوی مولکول های متنوع تری باشد، نتایج حاصل بهتر خواهد بود. در جدول ۳ نام برخی از توصیف کننده های الکترونی که قابل محاسبه به روش AM1 می باشد و با استفاده از بسته نرم افزاری MOPAC بدست آمده اند، نشان داده شده است.

---

1 – Pople  
2 – Santry  
3 – Segal  
4 – Fock  
5 – Dewar  
6 – Tiel  
7 – Oleris Value

جدول ۳: نام برخی از توصیف کننده های الکترونی قابل محاسبه توسط MOPAC

نام توصیف کننده	نشانه
بار جزئی روی مثبت ترین اتم	PCPA
بار جزئی روی منفی ترین اتم	PCNA
قطبش پذیری	POL
گرمای تشکیل	HF
انرژی الکترونی کل	TE
انرژی دافعه هسته- هسته	CCR
انرژی بالاترین تراز اشغال شده	HOMO
انرژی پایین ترین تراز اشغال نشده	LUMO
پتانسیل یونیزاسیون	IONP
دانسیته الکترونی منفی ترین اتم	EDNA
دانسیته الکترونی مثبت ترین اتم	EDPA

### ۱-۵-۵) توصیف کننده های فیزیکو شیمیایی

این توصیف کننده ها از جمله مهم ترین توصیف کننده ها هستند و با برهم کنش های حیاتی، همبستگی بالایی نشان می دهند. این دسته از توصیف کننده ها بیانگر بعضی از خواص فیزیکی و شیمیایی مولکول ها می باشند که به ساختار مولکول وابستگی شدیدی نشان می دهند. لگاریتم ضریب تقسیم آب - اکتانول، ویسکوزیته و میزان حلالیت ترکیبات در آب از این سری توصیف کننده ها می باشند.

### ۱-۶) تجزیه، تحلیل و ارزیابی آماری توصیف کننده ها

تا این مرحله انتخاب توصیف کننده ها براساس احتمالات و تشخیص نظری استوار است. بنابراین احتمال اینکه برخی از توصیف کننده ها برای مدل سازی مناسب نباشند وجود دارد و باید آنها را حذف کرد. توصیف کننده هایی که حذف می شوند معمولاً یک یا چند ویژگی زیر را دارند:

۱- توصیف کننده هایی که کمتر از ۱۰٪ مقادیر غیرصفر دارند یا دارای بیش از ۹۰٪ مقادیر یکسان باشند.

۲- توصیف کننده هایی که با توصیف کننده دیگر همبستگی بالایی دارند ( $r > 0.9$ )

۳- توصیف کننده هایی که با متغیر وابسته، همبستگی کمی دارند و محاسبه آنها مشکل است. با توجه به موارد فوق، تعدادی از توصیف کننده ها حذف می شوند و از پیچیدگی محاسبات جلوگیری می شود.

## ۷-۱) آنالیز مدل های آماری و انتخاب مدل مناسب

در این مرحله می توان با استفاده از روشهای آماری مختلف به جستجوی مدل پرداخت. مدل، در واقع یک رابطه ریاضی است که بیان کننده رابطه بین متغیر وابسته و مستقل می باشد و به کمک آن می توان با داشتن مقادیر متغیرهای مستقل، متغیر وابسته را ارزیابی کرد. برای این کار چند روش رگرسیون متفاوت وجود دارد.

۱ - **روش ورود اجباری**: در این روش اجباراً همه متغیرهای مستقل وارد مدل می شوند. در این روش اثرات مثبت و منفی متغیرهای مستقل برای ورود در مدل نادیده گرفته می شود. بنابراین روش مطلوبی نیست.

۲ - **روش انتخاب جلورونده**: در این روش ابتدا متغیری که بیشترین همبستگی را با متغیر وابسته دارد وارد مدل می شود و ابتدا یک خط یک متغیره ایجاد می شود. با انجام آزمون F روی مدل و معنی دار بودن آزمون، آن متغیر وارد شده در مدل باقی می ماند. بقیه متغیرهایی که بزرگترین آماره F را داشته و معنی دار باشد وارد مدل می شوند. این کار تا زمانی که دیگر آماره F حاصل از متغیرهای خارج از مدل معنی دار بنامند تکرار می شود و در نهایت، مدل نهایی بدست می آید.

۳ - **روش حذف عقب رونده**: در این روش ابتدا یک آماره F برای هر متغیر موجود در مدل تعیین می شود. با توجه به اثر متغیر بر روی F کلی و ضریب خطی بودن مدل، متغیر حذف یا نگه داشته می

شود. حذف شدن متغیرها تا زمانی که دیگر متغیری قابل حذف نباشد تکرار می شود و نهایتاً مدل باقی مانده به عنوان مدل مناسب انتخاب می شود.

**۴ - روش رگرسیون چند مرحله ای:** این روش مشابه روش انتخاب جلوبرنده است، با این تفاوت که بر روی متغیرهای وارد شده به مدل در مرحله قبل، آزمون مجدد انجام می گیرد. در این روش در هر مرحله برای متغیر وارد شده در مدل، آماره  $F$  محاسبه می شود و متغیر با کوچکترین آماره  $F$  بدون معنی حذف می شود و مدل مناسب بدست می آید [۱۸].

### ۸-۱) ارزیابی مدل بدست آمده

برای اطمینان از اینکه مدل بدست آمده، مدلی مناسب می باشد باید یک سری پارامترها را در نظر بگیریم، به طور کلی می توان گفت مدلی مناسب تر است که دارای ویژگی های زیر باشد:

- تعداد متغیرهای مستقل آن از سایر مدل های دیگر کمتر باشد.
- محاسبه متغیرهای مستقل آن ساده تر از سایر مدل ها باشد.
- حتی الامکان توصیف کننده هایی که مقدار آن ها باید به روش تجربی اندازه گیری شود در مدل وارد نشده باشد.
- ضریب همبستگی بین توصیف کننده هایی که در مدل شرکت دارد کم باشد.
- خطای استاندارد آن کم باشد.
- آماره  $F$  بزرگ بوده و ضریب همبستگی آن به یک نزدیک باشد.
- قدرت پیش بینی بالایی داشته باشد.

### ۹-۱) نرم افزارهای مورد استفاده در این نوع پژوهش ها

در این بخش مختصراً به شرح بسته نرم افزارهایی که در این پروژه مورد استفاده قرار گرفته اند پرداخته می شود.



## ۹-۱- الف ) بسته نرم افزاری Hyperchem

با این نرم افزار، ابتدا شکل مولکول به طور تقریبی بر روی صفحه نمایشگر کامپیوتر رسم می گردد، سپس توسط روش های مکانیک کوانتومی، ساختار مولکول بهینه می شود. با استفاده از این نرم افزار می توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچشی و ... را به دست آورد. این بسته نرم افزاری، قابلیت ترسیم ساختار کلاسترها و ترکیبات پیچیده را دارد و در کتابخانه آن ساختار اسیدهای آمینه و نوکلئیک اسیدها موجود است.

## ۹-۱- ب ) بسته نرم افزاری MOPAC

MOPAC یک نرم افزار اوربیتال ملکولی نیمه تجربی برای مطالعه ساختارهای مولکولی و واکنش های شیمیایی است. این نرم افزار با استفاده از خروجی Hyperchem که شامل زوایای پیچشی و زوایای پیوندی و طول پیوندها است محاسبات را انجام می دهد. از نتایج حاصل از چنین محاسباتی می توان اطلاعاتی از قبیل گرمای تشکیل، پتانسیل یونیزاسیون، انرژی الکترونی، انرژی دافعه هسته - هسته، گشتاور دو قطبی و بسیاری از فاکتورهای دیگر را در مورد مولکول ها بدست آورد. با وجود اینکه این برنامه براساس مفاهیم تئوری کوانتومی و مفاهیم ترمودینامیک عمل کرده و از محاسبات پیشرفته ریاضی کمک می گیرد، کاربران نیازی به دانستن عمیق مفاهیم فوق ندارند. داده های ورودی و خروجی این برنامه در حد ممکن ساده هستند تا جایی که می توان به راحتی توجه خود را به سایر جنبه های شیمیایی مسئله معطوف کرد [۱۹].

## ۹-۱- پ ) بسته نرم افزاری SPSS/PC

این بسته نرم افزاری، توانایی تحلیل کلی اطلاعات را داراست و عمومی ترین نرم افزار آماری است. امروزه آمار، مجموعه ای از مفاهیم و روش هایی است که در هر زمینه پژوهشی، برای گردآوری و تعبیر اطلاعات مربوط به آن و انجام نتیجه گیری ها، در شرایطی که عدم قطعیت

وجود دارد به کار می رود. SPSS/PC دارای توانایی بسیار بالایی است. در اینجا از این نرم افزار، برای بدست آوردن بهترین مدل، حذف متغیرهای با همبستگی بالا و ارزیابی و اعتبار مدل استفاده شده است.

### ۱-۹-ت) نرم افزار Dragon

دراگون نرم افزاری است که با استفاده از یک فایل ورودی، یک سری از توصیف کننده های مولکولی را که محاسبه آنها مشکل است، محاسبه می کند. فایل ورودی نرم افزار دراگون می تواند فایل (\*.hin) مربوط به نرم افزار Hyperchem باشد. این نرم افزار توسط گروه کمومتریکس دانشگاه میلانو طراحی شده است و قابلیت محاسبه ۱۴۹۷ توصیف کننده برای هر ملکول که کمتر از ۱۵۰ اتم داشته باشد را دارا است [۲۰].

### ۱-۱۰) شبکه عصبی مصنوعی

پیاده سازی ویژگی های شگفت انگیز مغز در یک سیستم مصنوعی (سیستم دینامیکی ساخته بشر) همیشه وسوسه انگیز و مطلوب بوده است. محققینی که طی سالها در این زمینه فعالیت کرده اند بسیارند لیکن نتیجه این تلاشها، صرف نظر از یافته های ارزشمند، باور هر چه بیشتر این اصل بوده است که مغز بشر دست نیافتنی است. با تاکید بر این نکته که گذشته از متافیزیک، دور از دسترس بودن ایده آل هوش طبیعی را می توان با عدم کفایت دانش موجود بشر از فیزیولوژی عصبی پذیرفت، باید اذعان داشت که عالی بودن هدف و کافی نبودن دانش موجود، خود سبب انگیزش پژوهش های بیشتر و بیشتر در این زمینه بوده و خواهد بود، همچنان که امروزه شاهد بروز چنین فعالیت هایی در قالب شبکه های عصبی مصنوعی هستیم.

## ۱-۱۰-الف) انگیزه های بیولوژیکی

تحقیقات و علاقه مندی به شبکه های عصبی از زمانی شروع شد که مغز به عنوان یک سیستم دینامیکی با ساختار موازی و پردازشگرهای کاملاً مغایر با پردازشگرهای متداول شناخته شد. نگرش نوین در مورد کارکرد مغز نتیجه تفکراتی بود که در اوایل قرن بیستم توسط رامون سگال در مورد ساختار مغز به عنوان اجتماعی از اجزای محاسباتی کوچک به نام نرون شکل گرفت. در این بخش، هدف آن است که بطور مختصر آن دسته از ویژگی ها، کارکرد و ساختار مغز را که انگیزه ای جهت توسعه شبکه های عصبی مصنوعی به حساب می آیند، تشریح شوند.

مغز به عنوان یک سیستم پردازش اطلاعات با ساختار موازی، از  $10^{11}$  نرون به هم مرتبط با تعداد  $10^{16}$  ارتباط تشکیل شده است. نرون ها ساده ترین واحد ساختاری سیستم های عصبی هستند. بافت هایی که عصب نامیده می شوند اجتماعی از نرون ها هستند که اطلاعات و پیام ها را از یک قسمت بدن به قسمت دیگر منتقل می کنند.

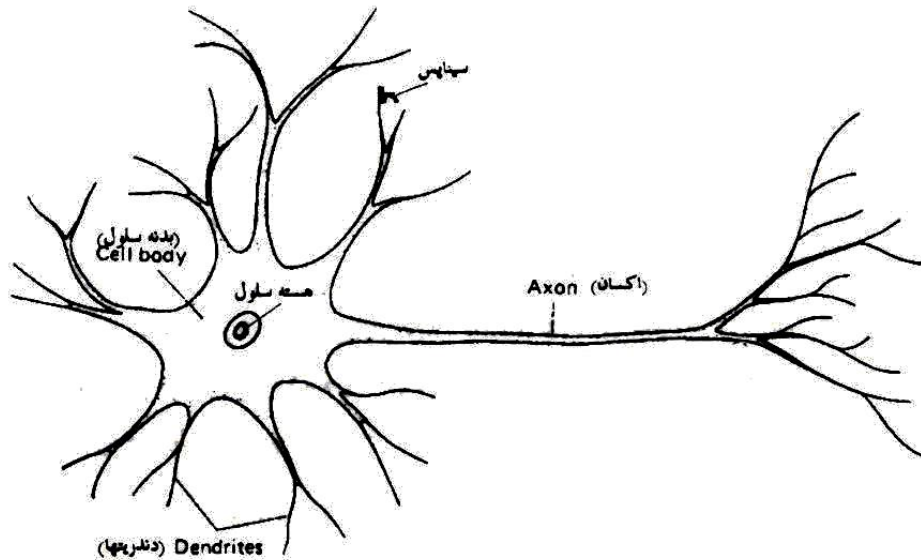
میلیونها نرون در بدن انسان وجود دارند. حتی ساده ترین کارهای روزمره انسان از قبیل پلک زدن تنها از همکاری همه جانبه این نرون ها میسر است. بیشترین تعداد نرون ها در مغز و باقی در نخاع و سیستم های عصبی جانبی تمرکز یافته اند. گرچه همه نرون ها کارکرد یکسانی دارند، ولی اندازه و شکل آنها بستگی به محل استقرارشان در سیستم عصبی دارد. با وجود این همه تنوع، بیشتر نرون ها از سه قسمت اساسی تشکیل شده اند:

- بدنه سلول (که شامل هسته و قسمت های حفاظتی دیگر می باشد).

- دندریت

- آکسون

که دو قسمت آخر عناصر ارتباطی نرون را تشکیل می دهند. شکل ۲ ساختمان سلول عصبی را نشان می دهد.



شکل ۲: نواحی اصلی یک سلول عصبی بیولوژیکی

دندریت ها به عنوان مناطق دریافت سیگنال های الکتریکی شبکه هایی تشکیل شده از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه های انشعابی بی شمار می باشند، به همین علت آنها را شبکه های دریافتی درخت گونه گویند. دندریت ها سیگنال های الکتریکی را به هسته سلول منتقل می کنند. بدنه سلول، انرژی لازم را برای فعالیت نرون فراهم کرده و بر روی سیگنال های دریافتی عمل می کند که با یک عمل ساده جمع و مقایسه با یک سطح آستانه انجام می شود [۲۱]. آکسون برخلاف دندریت ها از سطحی هموارتر (تعداد شاخه های کمتر) برخوردار می باشد. آکسون طول بیشتری داشته و سیگنال الکتروشیمیایی دریافتی از هسته سلول را به نرون های دیگر منتقل می کند. محل تلاقی یک آکسون از یک سلول به دندریت های سلول دیگر را سیناپس می گویند. سیناپس ها واحد ساختاری کوچک

تابعی<sup>۱</sup> هستند که ارتباط بین نرونها را برقرار می سازند. سیناپسها انواع مختلفی دارند که یکی از مهمترین آنها، سیناپس های شیمیایی است.

پیامهای عصبی تنها به صورت یک طرفه حرکت می کنند، از دندریت ها به بدنه سلول و سپس به آکسون. سالها محققان بیولوژی فکر می کردند که سیگنال های عصبی تنها از نوع الکتریکی می باشند. جزئیات فعالیت های عصبی همچنان نامشکوف بود، تا این که محققان با دستیابی به تکنولوژی معاصر توانستند پاسخ های یک سلول عصبی منفرد را اندازه بگیرند. نتیجه این بود که این پالس های عصبی توسط یک فرآیند الکتروشیمیایی به انجام می رسند.

#### ۱-۱۰-ب) تشابهات و انتظارات

با توجه به مقدمات فوق، می توان گفت که با تمام اغراق ها در مورد شبکه عصبی مصنوعی، این شبکه ها اصلاً سعی در حفظ پیچیدگی مغز ندارند. اکنون موضوع یاد شده را تحت دو عنوان تشابهات و انتظارات مورد بررسی قرار می دهیم.

#### ۱-۱۰-پ) تشابهات

۱- بلوکهای ساختاری در هر دو شبکه مصنوعی و بیولوژیکی، دستگاه های محاسباتی خیلی ساده ای هستند و علاوه بر این، نرون های مصنوعی از سادگی بیشتری برخوردار می باشند.

۲- ارتباط های بین نرون ها، عملکرد شبکه را تعیین می کند.

گرچه نرون های بیولوژیکی از نرون های مصنوعی، بسیار کندتر هستند (یک میلیون بار)، اما عملکرد مغز، خیلی سریعتر از عملکرد یک کامپیوتر معمولی است. علت این پدیده، بیشتر به خاطر

---

<sup>1</sup> - Functional

ساختار کاملاً موازی نرونها می باشند و این یعنی این که همه نرون ها معمولاً به طور همزمان کار می کنند و پاسخ می دهند [۲۲]. شبکه های عصبی مصنوعی نیز دارای ساختار موازی هستند.

### ۱-۱۰-ب) انتظارات

شبکه های عصبی مصنوعی با وجود اینکه با سیستم عصبی طبیعی قابل مقایسه نیستند ویژگی هایی دارند که آنها را در بعضی از کاربردها مانند تفکیک الگو، رباتیک، کنترل و به طور کلی در هر جا که نیاز به یادگیری خطی یا غیرخطی باشد، ممتاز می نمایند. این ویژگی ها به شرح زیر هستند:

- قابلیت یادگیری

- پراکندگی اطلاعات

- قابلیت تعمیم<sup>۱</sup>

- پردازش موازی

- مقاوم بودن (قابلیت تحمل خطا)

### ۴-۳- مدل سازی سیستم های عصبی

پس از روشن شدن ساختار و نحوه اتصال سلول های عصبی کوشش های زیادی برای ایجاد یک مدل ریاضی با عملکرد مشابه سلول های عصبی صورت گرفت. در سال ۱۹۴۳، مک کالوچ<sup>۲</sup> یک عصب شناس و یک ریاضی دان به نام پیتز<sup>۳</sup> برای نخستین بار یک مدل ریاضی برای بهینه سازی عملکرد یک سلول عصبی پیشنهاد کردند [۲۳]. این مدل کاملاً ساده بوده و امروزه به نام خود آنها معروف است. آن ها در بررسی های خود یک تابع منطقی را به طور دلخواه جهت نشان دادن حد آستانه نرون های ریاضی خود پیشنهاد کردند که خاصیت همه یا هیچ دانست اما نتوانستند راهی برای

<sup>1</sup> - Generalization

<sup>2</sup> - McCulloch

<sup>3</sup> - Pitz

آموزش سلولهای عصبی پیشنهادی ارائه دهند. در سال ۱۹۴۹ هب<sup>۱</sup> الگوی یادگیری موثری برای آموزش سلول های عصبی بهینه سازی شده ارائه داد که امروزه به الگوی آموزشی هیبان<sup>۲</sup> معروف است [۲۴]. این الگو راه حل عملی را جهت آموزش سلول های عصبی مصنوعی و شبکه های حاصل از آنها در پیش روی محققین قرار دارد. به این ترتیب سلول های عصبی بهینه سازی شده به عنوان اجزاء اصلی سازنده شبکه های عصبی مصنوعی مورد استفاده قرار گرفت.

### ۱-۱۰-۱) عنصر پردازش<sup>۳</sup>

عنصر پردازش در شبکه های عصبی مصنوعی نقش مشابه سلول عصبی را در سیستم های عصبی موجودات زنده دارد. در واقع واحد اصلی سازنده شبکه های عصبی مصنوعی عنصر پردازش نام داد. شکل ۳ یک عنصر پردازش را نشان می دهد. هر عنصر پردازش دارای  $x$  ورودی  $x_{ij}$  می باشد. زیرنویس  $i$  می تواند مقداری بین ۱ و  $n$  داشته باشد و نشان دهنده سیگنال ورودی به عنصر پردازش است. هر ورودی  $x_{ij}$  قبل از ورود به عنصر پردازش وزن می گیرد. وزن را با  $W_{ij}$  نشان می دهند و مقدار آن متناظر با قدرت اتصالات سیناپسی است. هر وزن در ورودی مربوط به خودش ضرب می شود. هر عنصر پردازش دارای یک ترم بایاس<sup>۴</sup> و یک حد آستانه است. جمع سیگنال های درونی عنصر پردازش باید از یک حد آستانه تجاوز کند تا سیگنال خروجی ایجاد شود. یک تابع انتقال نیز مسئول ایجاد سیگنال های خروجی،  $O_i$  است. از این سیگنال خروجی می توان به عنوان ورودی برای سایر عناصر پردازش استفاده کرد. معمولاً به هر عنصر پردازش که جزئی از یک شبکه با تعداد زیادی

---

<sup>1</sup> – Hebb

<sup>2</sup> – Hebbian

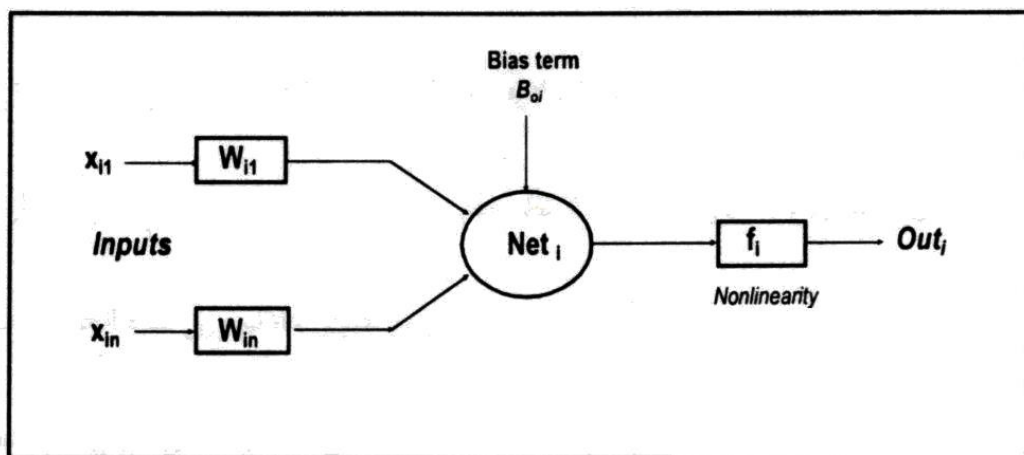
<sup>3</sup> – Processing element

<sup>4</sup> – Bias

عناصر پردازش است، گره<sup>۱</sup> می گویند. معادله ریاضی مربوط به عملکرد عنصر پردازش  $i$  به صورت زیر نوشته می شود:

$$o_i = f_i \left[ \left( \sum_{j=1}^n w_{ij} \cdot x_{ij} \right) + B_{0i} \right] \quad (1-4)$$

اندیس  $i$  نشان دهنده نرون موردنظر و  $j$  نشان دهنده ورودی هایی است که به عنصر پردازش موردنظر وارد شده است.  $f_i$  نیز نشان دهنده تابع انتقال مورد استفاده است.



شکل ۳: نمای کلی یک عنصر پردازش

### ۱-۱۰-۱) توابع انتقال<sup>۲</sup>

تابع انتقال  $f$  می تواند خطی یا غیرخطی باشد. یک تابع انتقال براساس نیاز خاص حل یک مساله، مساله ای که قرار است به وسیله شبکه عصبی حل شود، انتخاب می شود. در عمل تعداد محدودی از توابع انتقال مورد استفاده قرار می گیرند. توابع انتقال باید دو شرط اصلی داشته باشند:

۱- مقدار تابع باید در محدوده مشخصی تعریف شده باشد.

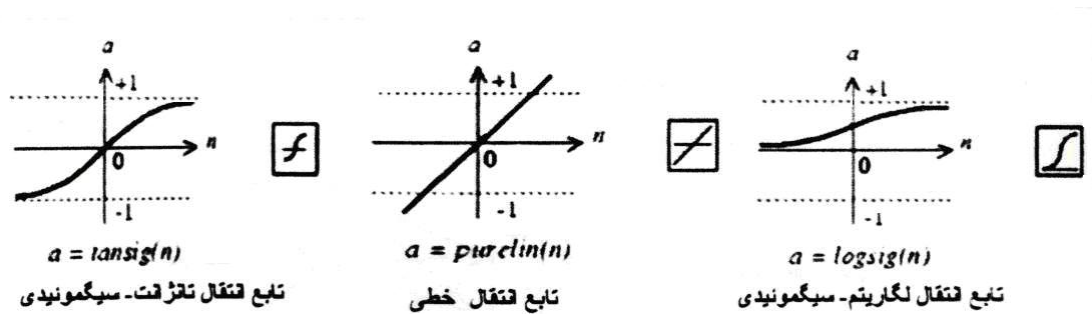
۲- مقدار تابع باید به صورت یکنواختی افزایش یابد.

<sup>1</sup> – Node

<sup>2</sup> – Transfer function



چند مورد تابع انتقال در شکل ۴ آورده شده است.



شکل ۴: چند مورد از توابع انتقال موجود در ANN

در اغلب موارد از توابع انتقال سیگموئیدی<sup>۱</sup> استفاده می شود. چون این توابع دارای محدوده معینی هستند، مشتق آنها به سادگی قابل محاسبه بوده و غیرخطی می باشند. خروجی این توابع همواره بین صفر و یک است.

### ۱-۱۰-ج) شبکه ها

ساختار توپولوژی یک شبکه توسط تعداد ورودی ها، تعداد خروجی ها و نحوه ارتباط این متغیرها با هم مشخص می شود. ساختار یک شبکه استاندارد به عناصر پردازش (نرون) شبکه و گروه های مختلفی که تحت عنوان لایه از آنها یاد می شود تقسیم می شوند.

به کلیه n عنصر پردازش موجود در یک لایه تعداد m متغیر ورودی با وزن  $W_{ij}$  وارد می شود و پس از عبور از تابع انتقال به سمت لایه بعدی هدایت می شود. این عمل را می توان به صورت معادلات ریاضی زیر نشان داد:

<sup>1</sup> – Sigmoidal

$$Net_i = \sum (W_{ij} \cdot out_j) + Bo_i \quad (2-4)$$

$$out_i = f(Net_i) \quad (3-4)$$

در توابع فوق  $out_i$  خروجی آمین گره در لایه موردنظر بوده و  $out_j$  خروجی زامین گره در لایه قبلی است. همچنانکه قبلاً ذکر شد تابع انتقال مورد استفاده معمولاً از نوع سیگموئیدی بوده و به صورت زیر نشان می دهیم:

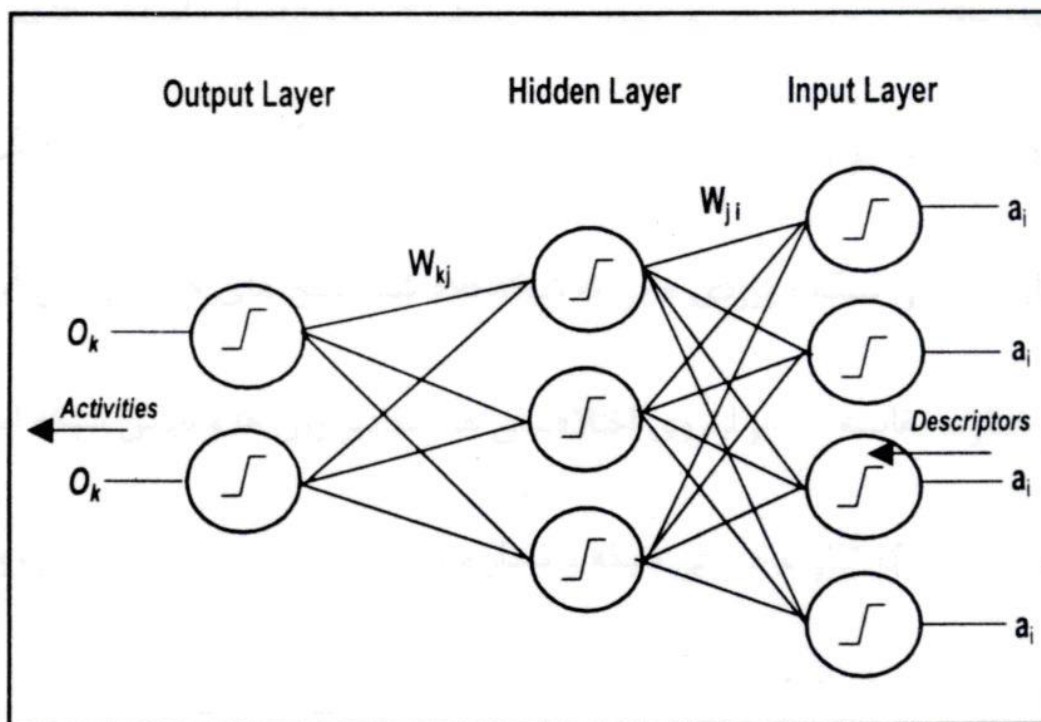
$$f(net_i) = \frac{1}{1 + e^{-net_i / G}} \quad (4-4)$$

در رابطه فوق  $G$  به بهره<sup>۱</sup> معروف بوده و هر چه مقدار آن بزرگتر باشد تغییرات تابع سیگموئیدی آهسته تر خواهد بود.

لازم به ذکر است که  $m$  متغیر ورودی ممکن است متغیرهای ورودی اولیه و یا خروجی نرون های لایه قبلی باشد که به عنوان ورودی برای لایه بعدی محسوب می شود. تعداد نرون های موجود در هر لایه و تعداد لایه ها در هر شبکه بستگی به کاربرد ویژه شبکه عصبی طراحی شده دارد. معمولاً لایه ای که اطلاعات ورودی به آنها داده می شود تحت عنوان لایه ورودی و لایه ای که داده های خروجی از آن دریافت می شود تحت عنوان لایه خروجی نامیده می شود و به لایه های دیگر موجود بین این دو لایه، لایه یا لایه های پنهانی گفته می شود [۲۵] و جهت حرکت سیگنال ها همواره از سمت لایه ورودی به سوی لایه خروجی است (شکل ۵).

---

<sup>1</sup> – Gain



شکل ۵: شمایی از یک شبکه عصبی مصنوعی

#### ۴-۳-۴- انواع شبکه

شبکه پس انتشار (BP)<sup>۱</sup> و شبکه عصبی رگرسیون عمومی (GRNN)<sup>۲</sup> برای مدل سازی، پیش بینی و رده بندی کاملاً مناسب اند. شبکه BP روش بسیار ساده برگشت پذیر غیرخطی چند منظوره ای است که خطا را به حداقل می رساند. به طور نظری هر تابع چند بعدی را با این شبکه می توان حل کرد. این شبکه می تواند مجموعه داده های پیچیده را به طور بسیار فشرده نمایش دهد و قادر است هر نوع مسئله ای را به دقت اداره کند. پژوهشگران بسیاری در توسعه این مدل کمک نموده اند و در حال حاضر این روش کاربرد فراوانی دارد، به همین منظور به شرح این روش اکتفا می کنیم [۲۶].

<sup>۱</sup> – Back-Propagation

<sup>۲</sup> – General Regression Neural Network

## ۱-۱۰-۱) شبکه پس انتشار (BP)

شبکه BP متشکل از ورودی، خروجی و دست کم یک لایه پنهان است. شبکه با محاسبه خطا میان خروجی مطلوب و واقعی و پس فرستادن اطلاعات خطا به هر یک از گره ها در شبکه، آموزش می بیند. خطای پس انتشار یافته برای انجام یادگیری در هر گروه به کار برده می شود. شمایی از یک شبکه پس انتشار در شکل ۶ آورده شده است. فرآیند تغییر وزن های رابط تا دست یابی به نتیجه مطلوب، یادگیری یا انطباق خواننده می شود. چندین قاعده یادگیری وجود دارد. بین این قواعد، پر کاربردترین آن ها یادگیری دلتا و قاعده نزول گرادیان می باشد. قاعده دلتا بر مبنای تفکر تنظیم مداوم وزن ها و بایاس شبکه، در جهت کاهش اختلاف خروجی شبکه با مقدار واقعی استوار است و قاعده نزول گرادیان این جهت را نشان می دهد.

طبق قاعده نزول گرادیان تغییر وزن ها و بایاس شبکه در طی هر چرخه آموزش با مشتق اول خطای شبکه متناسب است. خطای شبکه را معمولاً به صورت میانگین مربع خطا نشان می دهند که از رابطه زیر محاسبه می شود:

$$E_k = \frac{1}{2} \sum_{j=1}^n (t_{kj} - y_{kj})^2 \quad (5-4)$$

در این رابطه،  $E_k$  خطای  $k$ امین لایه،  $y_{kj}$  مقدار مطلوب  $k$ امین گره و  $t_{kj}$  مقدار خروجی واقعی  $k$ امین

گره است که طبق قاعده نزول گرادیان به صورت معادله زیر محاسبه می شود:

$$\Delta k w_{ji} \alpha = -\partial E_k / \partial w_{ji} \quad (6-4)$$

که در اینجا  $\Delta k w_{ji}$  تغییر وزن بین عناصر  $i$  و  $j$  را نشان می دهد که منفی گرادیان تابع خطای مثبت به تغییر وزن ها است و نشان دهنده حرکت به سوی جهتی است که تابع خطا همواره کاهش یابد. این تغییر وزن باعث می شود که در هر معادله، جهت تغییر وزن ها به سمت خطای کمتر باشد. به عبارت دیگر، این معادله نشان دهنده چگونگی تصحیح وزن ها به روش نزول گرادیان در جهت خطای کمتر

می باشد. طبق این الگوریتم، تغییر وزن های شبکه در طی هر چرخه آموزشی از رابطه زیر قابل محاسبه است:

$$\Delta w_{ji}^{new} = \eta(\delta k_j - y_{ji}) + \alpha \Delta w_{ji}^{previous} \quad (7-4)$$

در این رابطه،  $\eta$  سرعت آموزش<sup>۱</sup>،  $\delta k_j$  خطای گره زد در لایه n و  $y_{kj}$  خروجی نرون j در لایه n-1 می باشد. در معادله فوق  $\alpha$  به ممنتوم<sup>۲</sup> معروف است و  $\Delta w_{ji}^{previous}$  میزان تغییر وزن در چرخه آموزشی مرحله قبل است. سرعت یادگیری،  $\eta$ ، ثابتی است که کنترل کننده میزان تغییر در وزن ها، در هر مرحله است. انتخاب مقادیر نسبتاً بزرگ  $\eta$  باعث می شود که شبکه به سرعت به حالت بهینه برسد، ولی مقادیر بسیار بزرگ آن باعث نوسان تابع خطا در اطراف نقطه بهینه می شود. معمولاً مقادیر  $\eta$  بین ۰/۱ و ۰/۹ در نظر گرفته می شود. نحوه محاسبه خطای گره،  $\delta k_j$  برای گره های لایه پنهان و لایه خروجی متفاوت است. این مقدار برای گره های لایه خروجی از رابطه زیر قابل محاسبه است:

$$\delta k_j = (t_{kj} - y_{kj}) y_{kj} (1 - y_{kj}) \quad (8-4)$$

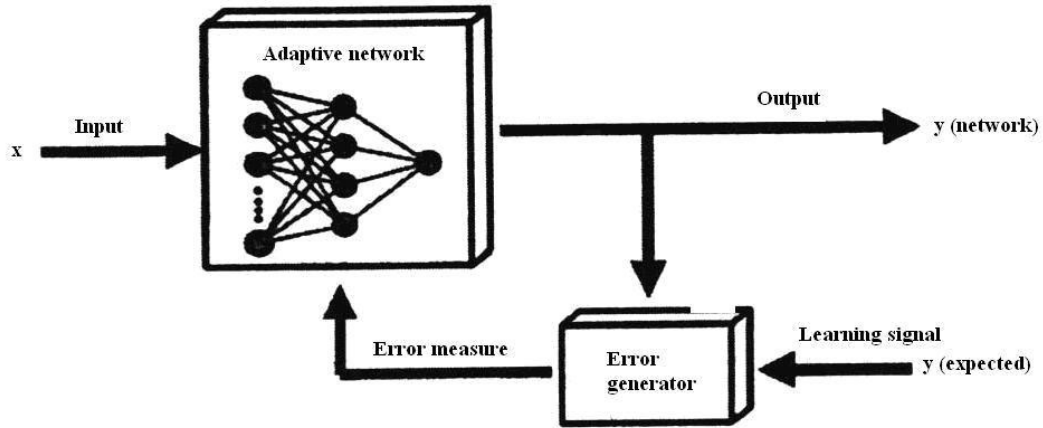
در معادله فوق،  $y_{kj}$  خروجی گره و  $t_{kj}$  مقدار واقعی کمیت مورد نظر است. چون مقدار خروجی برای گره های لایه پنهان مشخص نیست، لذا از معادله ۴-۹ برای محاسبه ترم خطا استفاده می شود.

$$\delta j_i = y_{kj} (1 - y_{kj}) \sum \delta(k+1)_j w(k+1)_j \quad (9-4)$$

در این معادله، ترم خطای  $\delta(k+1)_j$  نشان دهنده میزان خطا در لایه بعد از لایه پنهان مورد نظر است [۷۷ و ۷۸].

<sup>1</sup> – Learning rate

<sup>2</sup> – Momentum



شکل ۶: شمایی از یک شبکه پس انتشار

### ۱-۱۰-ح) ویژگی های شبکه عصبی مصنوعی

از نظر کلی شبکه های عصبی مصنوعی شبکه هایی سازش پذیرند که می توانند پس از

آموزش، خروجی های مطلوبی را تولید کنند. از نظر ریاضی، شبکه های فوق یک سیستم

دینامیکی بوده که می توانند بصورت تعدادی از معادلات نشان داده شوند. این شبکه ها قادر به جمع آوری، پردازش و ذخیره اطلاعات می باشند.

ویژگی های مهم شبکه های عصبی مصنوعی عبارتند از:

۱- تطبیق پذیری: الگوریتم قوی آموزش و قواعد خود سازماندهی توانایی سازگاری با تغییرات محیط را به این شبکه ها می دهد.

۲- پردازش غیرخطی: توانایی انجام محاسبات غیرخطی و مصونیت آن از اغتشاش، شبکه های فوق را به کاندیدای مناسبی برای بررسی مسائل مربوط به شناخت الگوها و پیش بینی کمیت ها تبدیل کرده است.

۳- پردازش موازی: معماری خاص این شبکه ها در استفاده از تعداد زیادی عناصر پردازش و ارتباط گسترده بین آن ها، شبکه های فوق را به خوبی برای پردازش موازی و ذخیره غیرمجموع داده ها مناسب می کند.

### ۱-۱۰-خ) مزایا و معایب شبکه های عصبی مصنوعی

برخی از مزایای شبکه های عصبی مصنوعی به قرار زیر می باشد:

- توانایی انجام محاسبات غیرخطی

- توانایی شناسایی و دسته بندی الگوها

- تطبیق پذیری بسیار بالا

- غیرپارامتری بودن

- عمومیت داشتن

- عدم وابستگی به مدلی خاص
- غیرالگوریتمی بودن
- توانایی انجام محاسبات سنگین به صورت موازی
- توانایی یادگیری و آموزش از طریق مثالهایی که به آن داده می شود
- عدم تاثیرپذیری از اغتشاشات
- معایب عمده شبکه های عصبی مصنوعی نیز به صورت زیر می باشد:
- سرعت آهسته آموزش
- مشکلات موجود در توجیه و تفسیر نتایج پیش بینی شده توسط آن
- آموزش اضافی
- وابستگی بالای شبکه به نوع داده های سری آموزشی
- مشکل قرار گرفتن در نقطه بهینه محلی

### ۱-۱۰-۵) کاربرد شبکه های عصبی مصنوعی

تعداد کاربردهای شبکه عصبی مصنوعی در زمینه های مختلف و برای حل مسائل مختلف در شیمی به سرعت در حال رشد است [۲۹-۳۱]. به طور کلی می توان ورودی های شبکه عصبی را به پنج دسته تقسیم کرد.

۱- طیف ها برای تعیین ارتباط بین ساختار و طیف [۳۲ و ۳۳].

۲- توالی اسیدهای آمینه برای تعیین ساختار دوم پروتئین ها [۸۶-۳۴].

۳- مجموعه ای از حسگرهای مختلف برای تشخیص عملکرد تولید در طول کنترل فرآیند

[۳۷ و ۳۸].



۴- آنالیزهای چند عنصری.

۵- مطالعات QSAR [۳۹ و ۴۰].

از کاربردهای دیگر شبکه های عصبی مصنوعی می توان به موارد زیر اشاره نمود:

پیش بینی واکنش پذیری شیمیایی [۴۱ و ۴۲]، سیگنالهایی که از الکترودهای یون گزین برای تعیین

غلظت یونهای مختلف بدست می آیند [۴۳] و طیف سنجی فلوئورسانس اشعه X [۴۴].

## ۲) بخش تجربی

### ۱-۲) روش ها

#### ۱-۱-۲) سری داده ها

داده های زمان بازداری ۹۴ گونه از ترکیبات مختلف آلی از مرجع ۴۵، گرفته شده است. نام

این ترکیبات و مقدار عددی زمان بازداری آنها در جدول ۴، نشان داده شده است. برای مدل سازی،

ابتدا زمان بازداری ترکیبات آلی مورد مطالعه به عنوان پارامتر مورد بحث انتخاب شدند. در مرحله دوم ساختار تمام گونه های مورد نظر با استفاده از نرم افزار hyper chem. رسم و بهینه گردیدند. روش AM1 برای بهینه نمودن ساختار این ترکیبات بکار رفته است.

جدول ۴- نام، فرمول مولکولی و زمان بازداری ترکیبات مورد مطالعه

Compound number	Identification	Molecular formula	Retention time (mine)
۱	Naphthalene	$C_{10}H_8$	11.004
۲	2-Methylnaphthalene	$C_{11}H_{10}$	13.548
۳	1-Methylnaphthalene	$C_{11}H_{10}$	13.910
۴	Biphenyl	$C_{12}H_{10}$	15.410
۵	Acenaphthylene	$C_{12}H_8$	16.822
۶	Acenaphthene	$C_{12}H_{10}$	17.470
۷	Dibenzofuran	$C_{12}H_8O$	18.191
۸	Fluorene	$C_{13}H_{10}$	19.488
۹	2-Methylfluorene	$C_{14}H_{12}$	21.643
۱۰	1-Methylfluorene	$C_{14}H_{12}$	21.759
۱۱	9-Methylene fluorene	$C_{14}H_{10}$	22.126
۱۲	Dibenzothiophene	$C_{12}H_8S$	22.568
۱۳	Phenanthrene	$C_{14}H_{10}$	23.056
۱۴	Anthracene	$C_{14}H_{10}$	23.214

۱۵	thiophene [2,3 – b] Naphtha	$C_{12}H_8S$	23.498
۱۶	Carbazole	$C_{12}H_9N$	24.069
۱۷	1H-Idene,1-phenyl	$C_{15}H_{12}$	24.206
۱۸	1H-Idene,2-phenyl	$C_{15}H_{12}$	24.536
۱۹	3-Methylphenanthrene	$C_{15}H_{12}$	25.003
۲۰	2-Methylphenanthrene	$C_{15}H_{12}$	25.090
۲۱	2-Methylanthrecene	$C_{15}H_{12}$	25.240
۲۲	4-/9-Methylphenanthrene	$C_{15}H_{12}$	25.407
۲۳	1-Methylphenanthrene	$C_{15}H_{12}$	25.490
۲۴	2-Methylcarbazole	$C_{13}H_{11}N$	25.841
۲۵	Methylcarbazole	$C_{13}H_{11}N$	25.940
۲۶	Phenyl naphthalene isomer	$C_{16}H_{12}$	27.433
۲۷	Fluoranthene	$C_{16}H_{10}$	27.704
۲۸	Acephenanthrylene	$C_{16}H_{10}$	28.033
۲۹	thiophene [4,5 – bcd] phenanthro	$C_{14}H_8S$	28.175
۳۰	Pyrene	$C_{16}H_{10}$	28.475
۳۱		$C_{16}H_{10}O$	28.584
۳۲		$C_{16}H_{10}O$	28.801
۳۳	Phenyl methyl naphthalene	$C_{17}H_{14}$	28.850
۳۴	P-Terphenyl	$C_{18}H_{14}$	29.017
۳۵	Benzo	$C_{16}H_{10}O$	29.038
۳۶	2-Methyl fluoranthene	$C_{17}H_{12}$	29.476

۳۷	carbazole [ <i>def</i> ]4h-Benzo	$C_{14}H_9N$	29.551
۳۸	Terphenyl	$C_{18}H_{14}$	29.601
۳۹	fluorine [ <i>a</i> ] Benzo	$C_{17}H_{12}$	29.909
۴۰	4-methylpyrene	$C_{17}H_{12}$	30.164
۴۱	2- Methylpyrene	$C_{17}H_{12}$	30.235
۴۲	1- Methylpyrene	$C_{17}H_{12}$	30.660
۴۳	naphtha thiophene [2,1- <i>d</i> ] [ <i>b</i> ] Benzo	$C_{16}H_{10}S$	32.220
۴۴	fluoranthene Benzo [ <i>ghi</i> ]	$C_{18}H_{10}$	32.380
۴۵	phenanthrene [ <i>c</i> ]Benzo	$C_{18}H_{12}$	32.420
۴۶	naphtho thiophene [1,2- <i>d</i> ] [ <i>b</i> ] Benzo	$C_{16}H_{10}S$	32.540
۴۷	Methylbenzo anthracene [ <i>a</i> ]	$C_{19}H_{14}$	32.640
۴۸	naphtho thiophene [2,3- <i>d</i> ] [ <i>b</i> ] Benzo	$C_{16}H_{10}S$	32.853
۴۹	thracene [ <i>a</i> ] Benzo	$C_{18}H_{12}$	33.225
۵۰	Chrysene/triphenylene	$C_{18}H_{12}$	3.395
۵۱	1,1-Binaphthalene	$C_{20}H_{14}$	33.632
۵۲	2,2- Binaphthalene	$C_{20}H_{14}$	33.775
۵۳	Methyl benzonaphthothiophene	$C_{17}H_{12}S$	33.866
۵۴	9-phenylanthracene	$C_{20}H_{14}$	33.949
۵۵	Methylchrysene	$C_{19}H_{14}$	34.020
۵۶	Benzocarbazole	$C_{16}H_{11}N$	34.192
۵۷	Methylbenzo thiophene	$C_{17}H_{12}S$	34.212

	[2,3- <i>d</i> ] [ <i>b</i> ] naphtha		
۵۸	Methylbenz anthracene [ <i>a</i> ]	$C_{19}H_{14}$	34.274
۵۹	3-Methyl benz anthracene [ <i>a</i> ]	$C_{19}H_{14}$	34.779
۶۰	6-/8-Methylbenz anthracene [ <i>a</i> ]	$C_{19}H_{14}$	34.904
۶۱	2,2-Binaphthalene	$C_{20}H_{14}$	35.245
۶۲	2-phenylphenanthrene	$C_{20}H_{14}$	35.466
۶۳	fluorancene [ <i>b</i> ] Benzo	$C_{20}H_{12}$	37.155
۶۴	fluorancene [ <i>k</i> ] Benzo	$C_{20}H_{12}$	37.217
۶۵	fluorancene [ <i>a</i> ] Benzo	$C_{20}H_{12}$	37.442
۶۶	Benz acephenanthrylene [ <i>e</i> ]	$C_{20}H_{12}$	37.510
۶۷	pyrene [ <i>e</i> ] Benzo	$C_{20}H_{12}$	37.926
۶۸	pyrene [ <i>a</i> ] Benzo	$C_{20}H_{12}$	38.147
۶۹	Perylene	$C_{20}H_{12}$	38.358
۷۰	Phenyl pyrene	$C_{22}H_{14}$	38.446
۷۱	Dibenzophenanthrene	$C_{22}H_{14}$	38.910
۷۲	Dinaphthothiophene	$C_{20}H_{12}S$	40.411
۷۳	triphenylene [ <i>b</i> ] Benzo	$C_{22}H_{14}$	40.523
۷۴	1-phenyl pyrene	$C_{22}H_{14}$	40.786
۷۵	Indeno fluorancene [1,2,3- <i>cd</i> ]	$C_{22}H_{12}$	40.960
۷۶	Indeno pyrene [7,1,2,3- <i>cdef</i> ]	$C_{22}H_{12}$	41.230
۷۷	pyrene [1,2,3- <i>cd</i> ] Indeno	$C_{22}H_{12}$	41.530
۷۸	Dibenz anthracene [ <i>ah</i> ]	$C_{22}H_{14}$	41.603

۷۹	Pentaphene	$C_{22}H_{14}$	41.670
۸۰	chysene [b] Benzo	$C_{22}H_{14}$	41.857
۸۱	Picene	$C_{22}H_{14}$	41.955
۸۲	perylene [ghi] Benzo	$C_{22}H_{12}$	42.108
۸۳	Dibenzo chrysene [def, mno]	$C_{22}H_{12}$	42.480
۸۴	Dibenzo fluoranthene [b, e]	$C_{24}H_{14}$	44.813
۸۵	Naphtho [1,2 - k] fluoranthene	$C_{24}H_{14}$	45.130
۸۶	Dibenzo fluoranthene [b, k]	$C_{24}H_{14}$	45.359
۸۷	Dibenzo fluoranthene [a, e] isomer	$C_{24}H_{14}$	45.455
۸۸	Naphtho [2,3 - k] fluoranthene	$C_{24}H_{14}$	45.690
۸۹	pyrene [2,3 - e] Naphtho	$C_{24}H_{14}$	46.055
۹۰	pyrene [a, e] Dibenzo	$C_{24}H_{14}$	46.393
۹۱	Naphtho pyrene/di [2,1 - a] benzo pyrene [e, l]	$C_{24}H_{14}$	46.493
۹۲	pyrene [2,3 - a] Naphtho	$C_{24}H_{14}$	46.689
۹۳	pyrene [a, i] Dibenzo	$C_{24}H_{14}$	46.793
۹۴	pyrene [a, h] Dibenzo	$C_{24}H_{14}$	47.014

برای محاسبه پارامترهای موثر بر زمان بازداري و استفاده از آنها در بدست آوردن مدل (که به این پارامترها توصیف کننده یا descriptor می گویند) از چند نرم افزار و برنامه استفاده شده است که

بطور خلاصه به شرح آنها خواهیم پرداخت. از این پارامترهای مولکولی برای بدست آوردن بهترین مدل QSPR زمان بازداری استفاده شده است، که شامل توصیف کننده های الکترونی، هندسی و توپولوژیکی می باشد.

اولین توصیف کننده ای که در مدل حضور دارد قطبش پذیری متوسط است، که یک توصیف کننده الکترونی است. وقتی یک مولکول در یک میدان الکتریکی  $E_0$ ، در خلاء قرار می گیرد، یک ممان دوقطبی در مولکول القا می شود که با رابطه زیر تعریف می شود:

$$\mu_{IND} = \alpha \cdot E_0$$

که ثابت تناسب این معادله را قطبش پذیری می نامند. ممکن است که این قطبش پذیری اسکالر را بصورت مجموع دو قطبش پذیری الکترونی  $\alpha_E$  و قطبش پذیری اتمی  $\alpha_A$ ، در نظر بگیرند. بطور معمول، قطبش پذیری اسکالر  $\alpha$ ، برای توصیف قطبش پذیری القا شده کافی نمی باشد، بنابراین یک بردار قطبش پذیری برای بهتر نشان دادن قطبش پذیری القا شده و نمایش قطبش پذیری مولکولی استفاده شده است، که این یک کمیت اسکالر نبوده بلکه یک پارامتر برداری است. دومین توصیف کننده ای که در مدل وارد شده است معکوس بار نسبی مثبت روی مولکول است که این نیز یک توصیف کننده الکترونی است و مقدار آن بر روی میزان و چگونگی برهم کنش مولکول با فاز ثابت تاثیر گذار می باشد. سومین توصیف کننده موثر بر زمان بازداری که در مدل MLR حضور دارد تابع توزیع شعاعی وزن دار شده توسط الکترون گاتیوی ساندرسون اتمی (RDF060e) می باشد.

جدول ۵: مشخصات مدل MLR بدست آمده

Descriptor	Notation	Coefficient	Mean effect
Average polarizability	DIP	0.183 ( ± 0.005)	38.644
Reverse relative positive charge	RRPCG	-0.149( ± 0.006 )	-2.511
Reversed 3D-MoRSE – signal 22/weighted by atomic Sanderson electronegativities constant	RMor22e - -	-0.005 ( ± 0.001 )  0.691 ( ± 0.002)	-1.049  -

ذکر این نکته ارزشمند است که بر خلاف مدل رگرسیون مرسوم، در اینجا روش اعتبار تقاطعی (cross validation) برای ارزیابی مدل استفاده می شود. در این آنالیز، از روش ارزیابی تقاطعی استفاده شده است، به این ترتیب که یک مدل با  $n-1$  ترکیب ساخته شده و سپس ترکیب  $n$  ام را پیش بینی می کنیم. هر ترکیب از مدل خارج شده و سپس با مدل حاصله از سایر ترکیبات پیش بینی شده است. مقدار  $q^2$  (ضریب همبستگی ارزیابی تقاطعی) و عبارت  $PRESS_{val}$  (خطای پیش بینی مجموع مربعات) مدل بدست آمده نیز از معادلات زیر بدست می آیند:

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

$$q^2 = 1 - [PRESS_{val} / \sum_{i=1}^n (y_i - \hat{y}_i)^2] \quad (8)$$

که  $n$  تعداد ترکیبات در سری آموزشی،  $y_i$  زمان بازداری تجربی برای ترکیب  $i$  ام و  $\hat{y}_i$  نیز مقدار زمان بازداری پیش بینی توسط مدل بدون حضور ترکیب  $i$  است.  $PRESS_{val}$  نیز  $PRESS$  محاسبه شده است وقتی تمام ترکیبات سری آموزشی وارد مدل شده باشند. مقدار محاسبه شده  $q^2$  و  $PRESS$  به ترتیب برابر با  $0/936$  و  $0/54671$  می باشد، که این اعداد به نوعی بیانگر میزان اعتبار مدل بدست آمده می باشند.

در مطالعه انجام شده، تکنیک MLR برای پیش بینی زمان بازداری ترکیبات استفاده شده است. نتایج بدست آمده نشان می دهند که، MLR توانسته است یک ارتباط نسبتاً خوبی بین زمان بازداری و



خواص مولکولی برقرار نماید. بنابراین، توصیف کننده های حاضر در مدل های QSAR قادر به فراهم کردن اطلاعاتی قابل توجه در رابطه با خواص مولکولی (همانند خواص الکترونی، هندسی و توپولوژیکی) می باشند، که این خواص می توانند در فرایندهای فیزیکوشیمیایی موثر بر زمان بازداری این ترکیبات مشارکت داشته باشند.

لازم بذکر است که مدل حاصله دارای  $R = 0.980$  می باشد و مدل نسبتاً ایده آلی می باشد که ادامه این پروژه برای بهبود مقدار  $R$  (ضریب رگرسیون) مدل بوده و همچنین ایجاد یک رابطه غیر خطی از طریق روش شبکه عصبی مصنوعی (ANN) می باشد که در حال انجام است

## جدول :

Compound name	Dip	Rrpcg	Rmor22e
Naphthalene	193.96	12.82	24.39
2-Methylnaphthalene	203.30	13.89	7.69
1-Methylnaphthalene	199.21	13.89	4.67
Biphenyl	199.88	11.76	-58.82
Acenaphthylene	214.33	3.98	-11.49
Acenaphthene	179.68	3.98	-11.49
Dibenzofuran	209.00	11.76	-58.82
Fluorene	204.32	3.97	-11.49
2-Methylfluorene	210.43	11.90	4.26
1-Methylfluorene	202.95	11.90	15.63
9-Methylenefluorene	236.27	3.98	-11.49
Dibenzothiophene	239.11	13.89	4.26
phenanthrene	219.43	4.07	58.82
Anthracene	223.70	13.89	71.43
thiophene [2,3-b] Naphtha	232.56	13.70	-22.22
Carbazole	235.77	3.97	-11.49
1H-Idene,1-phenyl	231.08	3.94	-10.87
1H-Idene,2-phenyl	237.83	3.92	-10.87

3-Methylphenanthrene	231.85	3.97	-11.49
2-Methylphenanthrene	239.14	13.89	4.27
2-Methylanthracene	225.01	3.98	-11.49
4-/9-Methylphenanthrene	228.45	13.89	10.87
1-Methylphenanthrene	219.14	3.98	-11.49
2-Methylcarbazole	241.44	3.98	-11.49
Methylcarbazole	247.09	3.95	-11.49
Phenylanthracene isomer	241.26	3.98	-11.49
Fluoranthene	199.13	4.00	-11.49
Acephenanthrylene	253.92	13.70	-35.71
[4,5- <i>bcd</i> ] thiophene phenanthro	255.49	4.00	-11.49
Pyrene	247.85	3.98	-11.49
	89.30	8.00	4.05
	248.05	3.98	-11.49
Phenyl methyl naphthalene	262.37	3.98	-11.49
P-Terphenyl	272.85	13.89	3.76
Benzo	267.23	13.70	18.18
2-Methyl fluoranthene	272.20	3.94	-11.49
carbazole [ <i>def</i> ] 4h-Benzo	100.03	8.93	8.55
Terphenyl	99.23	8.93	12.20
fluorine [ <i>a</i> ] Benzo	108.74	9.90	27.03
4-methylpyrene	108.01	7.46	-7.69
2- Methylpyrene	105.68	8.77	-28.57
1- Methylpyrene	113.66	7.87	-6.99
thiophene naphtha [2,1- <i>d</i> ] [ <i>b</i> ]	116.50	9.52	-19.61
Benzo			
[ <i>ghi</i> ] fluoranthene Benzo	128.11	10.42	-4.63
phenanthrene [ <i>c</i> ] Benzo	125.82	10.42	-4.08
naphtho thiophene [1,2- <i>d</i> ] [ <i>b</i> ]	124.06	10.53	-11.76
Benzo			
[ <i>a</i> ] Methylbenzo anthracene	125.98	3.42	-11.36
naphtho thiophene [2,3- <i>d</i> ] [ <i>b</i> ]	131.57	10.00	7.19
Benzo			
thracene [ <i>a</i> ] Benzo	139.70	9.90	3.52
Chrysene/triphenylene	132.77	3.58	-111.11
1,1-Binaphthalene	119.70	5.59	-11.49
2,2- Binaphthalene	134.94	11.11	-9.71

Methyl benzonaphthothiophene	145.20	10.75	-6.41
9-phenylanthracene	142.72	10.87	-90.91
Methylchrysene	143.11	10.87	-142.86
Benzocarbazole	151.55	10.75	5.26
Methylbenzo thiophene [2,3-d] [b]	141.77	10.87	-9.43
naphtha			
Methylbenz anthracene [a]	142.26	10.87	21.28
3-Methyl benz anthracene [a]	131.26	6.10	-4.98
6-/8-Methylbenz anthracene [a]	129.64	6.02	-5.52
2,2-Binaphthalene	152.34	9.80	-10.87
2-phenylphenanthrene	149.12	10.75	142.86
fluorancene [b] Benzo	148.09	3.58	-5.26
fluorancene [k] Benzo	156.35	10.00	3.11
fluorancene [a] Benzo	163.31	9.62	90.91
Benz acephenanthrylene [e]	160.16	9.62	-10.53
pyrene [e] Benzo	156.87	13.33	-10.53
pyrene [a] Benzo	179.09	13.89	-30.30
Perylene	157.19	9.52	-13.33
Phenyl pyrene	163.10	10.75	-3.83
Dibenzophenanthrene	140.82	5.78	-2.08
Dinaphthothiophene	159.56	13.89	-1.75
triphenylene [b] Benzo	152.34	9.80	-10.99
1-phenyl pyrene	166.51	10.87	58.82
Indeno fluorancene [1,2,3-cd]	167.77	10.87	9.62
Indeno pyrene [7,1,2,3-cdef]	168.10	10.87	25.64
pyrene [1,2,3-cd] Indeno	172.32	3.82	-32.26
[ah] Dibenz anthracene	170.68	9.80	-1.78
Pentaphene	184.07	11.90	5.10
chysene [b] Benzo	170.70	3.79	-5.78
Picene	194.40	12.82	10.53
perylene [ghi] Benzo	177.18	4.22	50.00
Dibenzo chrysene [def, mno]	184.03	11.90	5.10
[b,e] Dibenzo fluoranthene	178.87	11.90	6.21

[1,2 - k] Naphtho fluoranthene	188.79	12.99	4.05
[b, k] Dibenzo fluoranthene	192.00	13.70	125.00
[a, e] Dibenzo fluoranthene isomer	195.00	6.25	-3.58
[2,3 - k] Naphtho fluoranthene	198.68	13.89	-2.65
pyrene [2,3 - e] Naphtho	190.51	12.66	14.08
pyrene [a, e] Dibenzo	163.41	6.62	-250.00
[2,1 - a] Naphtho pyrene/di benzo	189.75	4.52	-6.06
pyrene [e, l]			
pyrene [2,3 - a] Naphtho	194.46	12.82	8.26
pyrene [a, i] Dibenzo	205.40	3.91	-10.87
pyrene [a, h] Dibenzo	193.96	12.82	24.39

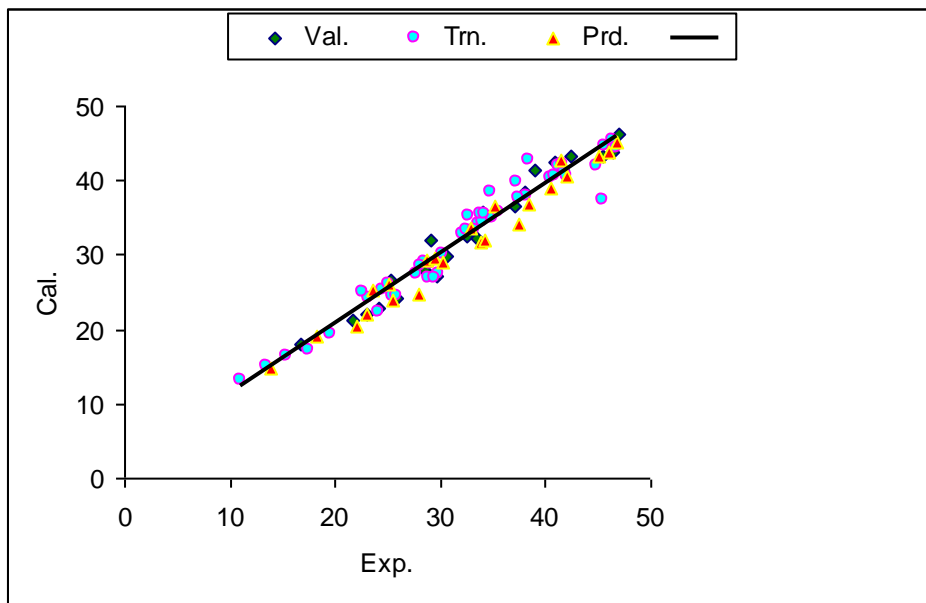
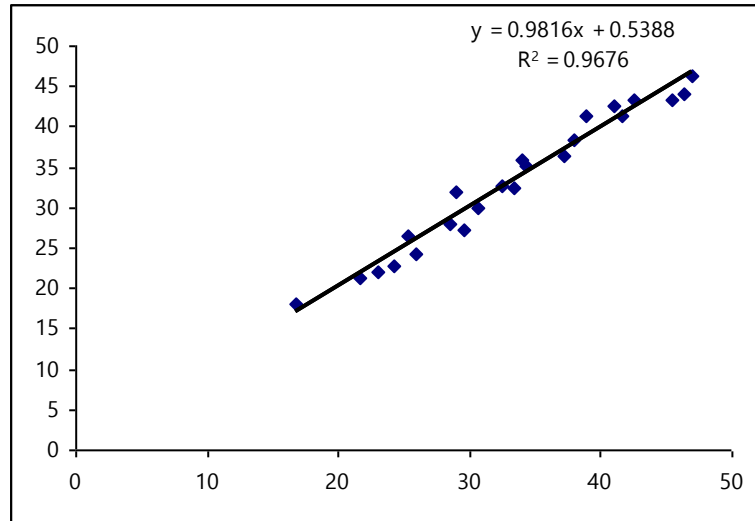
### جدول : مقایسه مقادیر

Compound name	Exp. Retention time (mine)	Calculated MLR Retention time	Calculated ANN Retention time
Naphthalene	11.004		13.148
2-Methylnaphthalene	13.548		15.057
1-Methylnaphthalene	13.910		14.814
Biphenyl	15.410		16.303
Acenaphthylene	16.822		18.072

Acenaphthene	17.470	17.192
Dibenzofuran	18.191	19.161
Fluorene	19.488	19.280
2-Methylfluorene	21.643	21.368
1-Methylfluorene	21.759	20.805
9-Methylene fluorene	22.126	20.525
Dibenzothiophene	22.568	24.883
phenanthrene	23.056	22.099
Anthracene	23.214	24.161
thiophene [2,3 - b] Naphtha	23.498	25.319
Carbazole	24.069	22.272
1H-Idene,1-phenyl	24.206	22.864
1H-Idene,2-phenyl	24.536	25.369
3-Methylphenanthrene	25.003	25.997
2-Methylphenanthrene	25.090	26.197
2-Methylanthracene	25.240	26.583
4-/9-Methylphenanthrene	25.407	24.583
1-Methylphenanthrene	25.490	23.980
2-Methylcarbazole	25.841	24.557
Methylcarbazole	25.940	24.246
Phenyl naphthalene isomer	27.433	27.58
Fluoranthene	27.704	27.473
Acephenanthrylene	28.033	24.618
[4,5 - bcd] thiophene phenanthro	28.175	28.444
Pyrene	28.475	28.011
	28.584	28.923
	28.801	29.277
Phenyl methyl naphthalene	28.850	26.965
P-Terphenyl	29.017	31.870
Benzo	29.038	28.712
2-Methyl fluoranthene	29.476	29.445
carbazole [def] 4h-Benzo	29.551	26.803
Terphenyl	29.601	27.142

fluorine [a] Benzo	29.909	27.475
4-methylpyrene	30.164	28.973
2- Methylpyrene	30.235	30.136
1- Methylpyrene	30.660	29.862
thiophene naphtha [2,1-d] [b]	32.220	32.872
Benzo		
[ghi]fluoranthene Benzo	32.380	31.276
phenanthrene [c] Benzo	32.420	33.453
naphtho thiophene [1,2-d] [b]	32.540	32.560
Benzo		
[a] Methylbenzo anthracene	32.640	35.203
naphtho thiophene [2,3-d] [b]	32.853	33.573
Benzo		
thracene [a] Benzo	33.225	33.444
Chrysene/triphenylene	3.395	32.338
1,1-Binaphthalene	33.632	34.122
2,2- Binaphthalene	33.775	31.760
Methyl benzonaphthothiophene	33.866	35.493
9-phenylanthracene	33.949	35.842
Methylchrysene	34.020	34.394
Benzocarbazole	34.192	32.059
Methylbenzo thiophene [2,3-d] [b]	34.212	35.582
naphtha		
Methylbenz anthracene [a]	34.274	35.257
3-Methyl benz anthracene [a]	34.779	38.517
6-/8-Methylbenz anthracene [a]	34.904	<b>34.83474</b>
2,2-Binaphthalene	35.245	<b>36.64212</b>
2-phenylphenanthrene	35.466	<b>35.87171</b>
fluorancene [b] Benzo	37.155	<b>36.4748</b>
fluorancene [k] Benzo	37.217	<b>39.80468</b>
fluorancene [a] Benzo	37.442	<b>34.12227</b>
Benz acephenanthrylene [e]	37.510	<b>37.59558</b>

pyrene [e] Benzo	37.926	<b>38.31871</b>
pyrene [a] Benzo	38.147	<b>37.79204</b>
Perylene	38.358	<b>36.7026</b>
Phenyl pyrene	38.446	<b>42.64907</b>
Dibenzophenanthrene	38.910	<b>41.4311</b>
Dinaphthothiophene	40.411	<b>40.43411</b>
triphenylene [b] Benzo	40.523	<b>39.10979</b>
1-phenyl pyrene	40.786	<b>40.64637</b>
Indeno fluorancene [1,2,3 - cd]	40.960	<b>42.59463</b>
Indeno pyrene [7,1,2,3 - cdef]	41.230	<b>42.0439</b>
pyrene [1,2,3 - cd] Indeno	41.530	<b>42.85338</b>
[ah] Dibenz anthracene	41.603	<b>42.12587</b>
Pentaphene	41.670	<b>41.43385</b>
chysene [b] Benzo	41.857	<b>41.26065</b>
Picene	41.955	<b>40.49476</b>
perylene [ghi] Benzo	42.108	<b>40.47751</b>
Dibenzo chrysene [def, mno]	42.480	<b>43.23805</b>
[b, e] Dibenzo fluoranthene	44.813	<b>43.85911</b>
[1,2 - k] Naphtho fluoranthene	45.130	<b>43.21804</b>
[b, k] Dibenzo fluoranthene	45.359	<b>37.47954</b>
[a, e] Dibenzo fluoranthene isomer	45.455	<b>43.33135</b>
[2,3 - k] Naphtho fluoranthene	45.690	<b>44.68612</b>
pyrene [2,3 - e] Naphtho	46.055	43.9275
pyrene [a, e] Dibenzo	46.393	43.948
[2,1 - a] Naphtho pyrene/di benzo	46.493	45.331
pyrene [e, l]		
pyrene [2,3 - a] Naphtho	46.689	45.288
pyrene [a, i] Dibenzo	46.793	44.425
pyrene [a, h] Dibenzo	47.014	46.177



مراجع :

[1] R. A. Mayer, *Encyclopedia of analytical chemistry*, WILEY, 2000.



- [2] B. G. M. Vandeginst., *Topics in current chemistry*, 141 ( 1987 ) 1.
- [3] S. Wold, *Analog analysis of chemical system, Kemish Tridskrft*, 3 ( 1972 ) 34.
- [4] B. Koalski, *J. Info. Comput. Sci*, 15 ( 1972 ) 203.
- [5] D. L. Manssnat, B. G. M. Vandeginste, S. N. Deming and L. Kaufman,.  
*Chemometrics, A Text Book*, Elsevier, Amesterdom, 1998.
- [6] P. C. Jurs,., *Computer software Applications in Chemistry*, Sons Inc, New York,  
1996.
- [7] D. L. Manssnat, B. G. M. Vandeginste, S. N. Deming and L. Kaufman,.  
*Chemometrics, A Text Book*, Elesenier, Amesterdom, 1988.
- [8] L. Escuder-Gilabert, Y. Martín-Biosca, S. Sagrado, R. M. Villanueva-Camañas  
and M. J. Medina-Hernández,., *Analytica Chimica Acta*, 448 ( 2001 ) 173.
- [9] Laura Fumagalli, Cristiano Bolchi, Simona Colleoni, Marco Gobbi, Barbara  
Moroni, Marco Pallavicini, Alessandro Pedretti, Luigi Villa, Giulio Vistoli and  
Ermanno Valoti , *Bioorganic and Medicinal Chemistry*, 13 ( 2005 ) 2547.
- [10] M. H. Fatemi and Nasser Goudarzi,., *Electrophoresis* , 26 ( 2005 ) 2968.
- [11] Z. Garkani-Nejad and M. Jalali-Heravi,., *J. Chromatogr. A*. 945 ( 2002 ) 173.
- [12] E. Morgan, *Chemometrics, Experimental Design*, John Eiley and Sons, London,  
1991.
- [13] L. S. Anker, P. C. Jurs and P. A. Edward,., *Anal. Chem.* 62 ( 1990 ) 2627.
- [14] T. R. stouch and P. C. Jurs,., *J. Chem Inf. Comput. Sci.*, 26 ( 1986 ) 1.
- [15] J. A. Pople and D. Beveridge, *Approximate Molecular Orbital Theory*,  
Mc graw-Hill, 1970.
- [16] J. A. Pople, D. P. Santry and G. A. Segal,., *J. Chem. Phys.* 43 ( 1965 ) 129.
- [17] M. J. Dewar and S. Thiel, *J. Am. Chem. Soc.* 99 ( 1977 ) 4899.
- [18] L. Xu and W. J. Zhang, *Anal. Chim. Acta*, 446 ( 2001 ) 477.

[19] MOPAC package, version 6.0, U. S. Air Force Academy, Colorado Spring, Co.  
80840.

[20] R. Todeschini, V. Consonni, A. Mauri and M. Pavan, Software Dragon  
*Calculation of Molecular Descriptors*, Department of Environmental Science,  
University of Milano-Bicocca and Talete, srl., Milan, Italy, 2003.  
<http://disat.unimib.it/chm/Dragon.htm>.

[45] Emilie Massold, *J. Chromatogr. A*, 1154 (2002) 342.

