

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
الْحَمْدُ لِلَّهِ الَّذِي
خَلَقَ السَّمَوَاتِ وَالْأَرْضَ
وَالَّذِي يُضَوِّبُ الْمَوْتِ
وَالَّذِي يُضَوِّبُ الْمَوْتِ
وَالَّذِي يُضَوِّبُ الْمَوْتِ



دانشگاه شهید باهنر کرمان

دانشکده علوم

گروه شیمی

پایان نامه تحصیلی برای دریافت درجه کارشناسی ارشد شیمی تجزیه

مدل سازی تمایل اتصال بنزودیازپین ها به گیرنده گابا-A
به روش QSAR با استفاده از تکنیک های الگوریتم ژنتیکی
و شبکه عصبی مصنوعی

استاد راهنما:

دکتر مهدی موسوی

استاد مشاور:

دکتر حسین نظام آبادی پور

مؤلف:

بهنام محسنی

شهریور ۱۳۸۷

۱۳۸۸ / ۴ / ۲۱

از کلیه اطلاعات درازک علمی بزرگ
تمسک به درازک

ب

۱۱۵۱۷۱



دانشگاه شهید باهنر کرمان

این پایان نامه به عنوان یکی از شرایط احراز درجه کارشناسی ارشد به

بخش شیمی

دانشکده علوم

دانشگاه شهید باهنر کرمان

تسلیم شده است و هیچ گونه مدرکی به عنوان فراغت از تحصیل دوره مزبور شناخته نمی شود.

دانشجو: بهنام محسنی

استاد راهنما: دکتر مهدی موسوی

استاد مشاور پروژه: دکتر حسین نظام آبادی پور

داور ۱: دکتر زهرا گرکانی نژاد

داور ۲: دکتر ابراهیم نوروزیان

نماینده تحصیلات تکمیلی دانشکده: دکتر محمد رضا اسلامی

حق چاپ محفوظ و متعلق به دانشگاه است



تقدیم به

روح بلند مهندس علیرضا افضل پور

و

همسر فداکارشان بانو فاخره صبا

تقدیر و تشکر

از زحمات بی شائبه و بذل لطف اساتید خویش جناب آقای دکتر مهدی موسوی و جناب آقای دکتر حسین نظام آبادی پور که افتخار شاگردی شان را داشتم قدردانی نموده و از صبر ایشان در علم آموزی بنده تشکر می کنم. بی شک جهد و تلاش و بردباری آن ها بود که توانستم این مقطع تحصیلی را به خوبی بگذرانم، صمیمانه برایشان آرزوی موفقیت دارم.

از اساتید بزرگوار جناب آقای دکتر نوروزیان و سرکار خانم دکتر گرکانی نژاد که زحمت مطالعه، تصحیح و داوری این پایان نامه را منقبل شدند، تشکر و قدردانی می کنم.

چکیده

بنزودیازپین ها دسته ای از داروهای آرام بخش-خواب آور هستند که ساختار شیمیایی شان بر مبنای گروه های دیازپین و فنیل می باشد. این ترکیبات با اتصال به گیرنده گابا-A در درون مغز، اثرات خود را به وجود می آورند. بنزودیازپین ها با اتصال به گیرنده گابا-A، آن را به ساختاری که تمایل بیشتری به نوروترانسمیتر گابا دارد، در می آورند. این عمل فرکانس باز شدن کانال یون کلرید مربوطه و فرا قطبیده شدن غشاء نرون را افزایش می دهد. به دلیل تفاوت های ساختاری، بنزودیازپین های مختلف تمایل متفاوتی برای اتصال به گیرنده گابا-A دارند. IC_{50} ، ۵۰٪ حداکثر غلظت بازدارندگی، این ترکیبات معیار مناسبی از تمایل اتصال آنها به گیرنده گابا-A است.

هدف تحقیق حاضر مدل سازی تمایل اتصال بنزودیازپین ها به گیرنده شان می باشد. به منظور انجام این کار، روش رابطه کمی ساختار فعالیت (QSAR) برای ایجاد یک مدل ریاضی جهت مرتبط کردن $logIC_{50}$ و ۱ و ۴ بنزودیازپین ۲-ان ها به عنوان تمایل آنها به گیرنده گابا-A، به توصیف کننده های ساختاری شان مورد استفاده قرار گرفت. در ابتدا، یک سری شامل ۸۰ ترکیب بنزودیازپین با ساختارهای مولکولی متفاوت به عنوان سری داده ها انتخاب گردید. سپس ۶۴ ترکیب از ۸۰ ترکیب به عنوان سری مرجع جهت مدل سازی $logIC_{50}$ و ۱۶ ترکیب باقی مانده به عنوان سری پیش بینی شونده به منظور ارزیابی مدل های محاسبه شده انتخاب شدند. در مرحله محاسبه توصیف کننده، ۱۵۱۸ پارامتر ساختاری برای هر مولکول بنزودیازپین محاسبه شد. پس از ارزیابی، ۸۷۹ توصیف کننده چون که همبستگی بالا و مقادیر ثابت داشتند کنار گذاشته شدند. در مرحله بعد، مدل سازی با استفاده از روش مرحله ای

رگرسیون خطی چندگانه (MLR) انجام شد. در واقع، در این قسمت تحقیق، MLR به عنوان روش انتخاب ویژگی استفاده شد. محاسبه و ایجاد مدل های MLR معایب مبنی بر روابط خطی بین توصیف کننده ها و $\log IC_{50}$ ، وقت گیر و خسته کننده بودن را دارد. به منظور در نظر گرفتن هم روابط خطی و هم روابط غیرخطی بین توصیف کننده ها و $\log IC_{50}$ ، شبکه های عصبی مصنوعی (ANN) به کار گرفته شدند. در این مرحله، همچنین الگوریتم ژنتیکی (GA) به عنوان روش انتخاب ویژگی برای انتخاب توصیف کننده های دارای بیشترین اطلاعات، به کار گرفته شد. در این جفت شدگی ANN و GA نقش های متفاوتی ایفا می کنند. ANN برای انطباق تمایل اتصال بنزودیازپین ها به گیرنده گابا-A با توصیف کننده هایی که قبلاً توسط الگوریتم ژنتیکی به عنوان انتخاب گر توصیف کننده انتخاب شده بودند، به کار گرفته شد. نتایج به دست آمده از جفت شدگی الگوریتم ژنتیکی و شبکه عصبی مصنوعی نسبت به نتایج روش MLR برتری نشان می دهند. بر طبق محاسبات شبکه عصبی ژنتیکی (GNN)، ۸ توصیف کننده از ۶۳۹ توصیف کننده در بهترین مدل حضور دارند. این توصیف کننده ها Mor18v, G2v, TW-5, MATS5m, BEHm7, C-028, C-8, $\log P-7$ هستند. از بین آنها Mor18v, G2v, TW-5 توصیف کننده های هندسی، MATS5m, BEHm7, C-028 توصیف کننده های توپولوژیکی، C-8 و $\log P-7$ به ترتیب توصیف کننده الکترونی و فیزیکوشیمیایی هستند. ساختار بهینه شبکه عصبی مصنوعی پس انتشار جفت شده با الگوریتم ژنتیکی ۱-۳-۵-۸ بود.

نهایتاً، مقایسه $\log IC_{50}$ محاسباتی و تجربی ترکیبات بنزودیازپین برای سری های مرجع و پیش بینی شونده همبستگی خوب و توانایی پیش بینی بالای مدل GNN بدست آمده را نشان داد. پارامترهای آماری نتایج بهترین مدل محاسبه شده روی سری مرجع $R=0/982$ ،

همچنین R ، SE و $F=20.06$ می باشند. روی سری پیش بینی شونده به ترتیب 0.955 ، 0.28 و 144 هستند. شایسته است که اشاره شود سری داده های به کار گرفته شده در این تحقیق کاملاً متنوع بوده و نتایج ما عمومیت و برتری نسبت به کارهای تقریباً مشابه دیگر نشان می دهند.

فهرست مطالب

صفحه	عنوان
	فصل اول: مقدمه
۲	۱-۱ مقدمه
۲	۲-۱ کمومتریکس
۴	۳-۱ رابطه کمی ساختار- فعالیت
۶	۱-۳-۱ تاریخچه پیدایش و تکامل QSAR
۷	۲-۳-۱ روش و اهداف QSAR
	فصل دوم: داروها و مکانیسم عمل آنها
۱۲	۱-۲ دارو و داروشناسی
۱۴	۲-۲ اصول کلی اثرگذاری دارو
۱۵	۱-۲-۲ فارماسیوتکس
۱۶	۲-۲-۲ فارماکوسینتیک
۱۸	۱-۲-۲-۲ جذب
۱۹	۲-۲-۲-۲ توزیع
۲۱	۳-۲-۲-۲ متابولیسم
۲۳	۴-۲-۲-۲ دفع
۲۳	۳-۲-۲ فارماکودینامیک
۲۴	۱-۳-۲-۲ تداخل دارو- گیرنده
۲۷	۲-۳-۲-۲ مکانیسم های مولکولی در اثرگذاری داروها
۲۸	۳-۳-۲-۲ ضریب درمانی
۳۰	۴-۳-۲-۲ پاسخ های مضر داروها
۳۱	۳-۲ داروهای آرام بخش - خواب آور

۳۲	۱-۳-۲ سیستم عصبی بدن انسان
۳۳	۱-۱-۳-۲ نرون بیولوژیکی
۳۵	۲-۱-۳-۲ گابا و گیرنده گابا-A
۳۷	۲-۳-۲ بنزودیازپین ها
۳۸	۱-۳-۳-۲ تاریخچه بنزودیازپین ها
۳۹	۲-۲-۳-۲ مکانیسم اثر بنزودیازپین ها
۴۰	۳-۲-۳-۲ آثار فارماکولوژیک بنزودیازپین ها
۴۱	۴-۲ کشف و طراحی دارو
۴۶	۱-۴-۱ پارامترهای بیولوژیکی در QSAR

فصل سوم: روش QSAR

۵۰	۱-۳ روش QSAR
۵۱	۱-۱-۳ انتخاب سری داده ها
۵۲	۲-۱-۳ تقسیم بندی سری داده ها به سری های مرجع و پیش بینی شونده
۵۲	۳-۱-۳ انتخاب و محاسبه توصیف کننده ها
۵۴	۱-۳-۱-۳ توصیف کننده های توپولوژیکی
۵۷	۲-۳-۱-۳ توصیف کننده های هندسی
۵۸	۳-۳-۱-۳ توصیف کننده های الکترونی
۵۹	۴-۳-۱-۳ توصیف کننده های فیزیکوشیمیایی
۵۹	۵-۳-۱-۳ توصیف کننده های توسعه یافته
۶۰	۴-۱-۳ ارزیابی، تجزیه و تحلیل آماری توصیف کننده ها
۶۰	۵-۱-۳ آنالیز آماری مدل ها و انتخاب مدل مناسب
۶۲	۶-۱-۳ ارزیابی اعتبار مدل های انتخاب شده
۶۳	۲-۳ رگرسیون خطی چندگانه
۶۴	۳-۳ شبکه های عصبی مصنوعی
۶۴	۱-۳-۳ تاریخچه شبکه های عصبی مصنوعی
۶۵	۲-۳-۳ مدل یک نرون مصنوعی
۶۷	۳-۳-۳ تابع فعالسازی

۶۸	۴-۳-۳ ساختار شبکه های عصبی مصنوعی
۶۹	۵-۳-۳ یادگیری و یادآوری در شبکه های عصبی مصنوعی
۷۱	۱-۵-۳-۳ روش یادگیری پس انتشار
۷۴	۲-۵-۳-۳ ارزیابی الگوریتم آموزش پس انتشار
۷۸	۶-۳-۳ کاربردهای شبکه های عصبی مصنوعی
۷۸	۴-۳ الگوریتم ژنتیکی
۷۹	۱-۴-۳ وراثت و انتخاب طبیعی
۸۱	۲-۴-۳ الگوریتم ژنتیکی
۸۶	۳-۴-۳ ساختار الگوریتم ژنتیکی
۹۱	۴-۴-۳ ارزیابی کارایی الگوریتم ژنتیکی

فصل چهارم: بخش تجربی

۹۴	۱-۴ انتخاب و تقسیم بندی سری داده ها
۹۵	۲-۴ محاسبه توصیف کننده ها و ارزیابی و تجزیه و تحلیل آنها
۱۰۱	۳-۴ محاسبه مدلها و ارزیابی آنها
۱۰۱	۱-۳-۴ مدلسازی خطی و انتخاب مدل مناسب
۱۰۴	۲-۳-۴ ارزیابی اعتبار مدلهاى خطی انتخاب شده
۱۰۷	۳-۳-۴ استفاده از شبکه های عصبی مصنوعی جهت مدلسازی
۱۱۶	۴-۳-۴ استفاده از الگوریتم ژنتیکی بعنوان روش انتخاب متغیر
۱۲۱	۴-۴ بحث و نتیجه گیری

۱۳۸

مراجع

فصل اول

مقدمه

۱-۱ مقدمه

بهبود میزان آگاهی بشر همواره نیازمند اطلاعات بیشتر است. بدین منظور دستگاهها و وسایل پیشرفته به کار گرفته می شوند که داده های فراوان حاصل از اندازه گیری های زیاد روی نمونه های متعدد را با صرف هزینه های سنگین، در اختیار قرار می دهند. تفسیر و ارزیابی این داده ها با روشهای معمول کاری بس مشکل و در اغلب اوقات غیر ممکن می نماید. به کارگیری روشهایی که دستیابی ارزان و سریع اطلاعات را فراهم آورد، بسیار مورد توجه است. علم شیمی نیز با توجه به گستردگی فراوان و وجود مسائل پیچیده و حل نشده بسیار به کارگیری چنین روشهایی را طلب می کند. پیشرفت استراتژی های محاسباتی و الگوریتم ها، اشتیاق دانشمندان برای مواجه شدن با حجم انبوه داده ها را دو چندان کرده است [۱]. استفاده از علوم ریاضی، آمار و کامپیوتر در زمینه های مختلف علمی همچون زیست شناسی، داروسازی و شیمی منجر به پیدایش علوم بیومتریک، آمارداری و کمومتریکس^۱ گردیده است [۲].

۱-۲ کمومتریکس

کمومتریکس شاخه ای از شیمی است که با هدف تهیه حداکثر اطلاعات از داده های شیمیایی، بوجود آمده است. کاربرد عملی ریاضی، آمار و کامپیوتر در شیمی را کمومتریکس یا شیمی سنجی گویند. در واقع چنین تلفیقی با سه هدف عمده صورت می پذیرد که عبارتند از:

۱- طراحی یا بهبود فرآیندهای اندازه گیری شیمیایی

۲- استخراج اطلاعات شیمیایی مفیدتر از داده های فیزیکی و شیمیایی

۳- پیدا کردن دانشی در مورد سیستم های شیمیایی [۳, ۴].

پیدایش این زمینه علمی که یکی از شاخه های شیمی به حساب می آید به سال ۱۹۶۹ بر می گردد، زمانی که جرز، کوالسکی، آیزنهور و ریلی^۱ مقاله ای را در زمینه کاربرد ماشین یادگیری جهت طبقه بندی طیف جرمی با تفکیک پایین منتشر کردند [۵]. اما اصطلاح کمومتریکس اولین بار توسط سوانت ولد^۲ سوئدی و بروس کوالسکی^۳ آمریکایی در سال ۱۹۷۲ مطرح شد [۶]. در پی کوشش های فراوان این دو دانشمند، انجمن بین المللی کمومتریکس^۴ (ICS) در سال ۱۹۷۴ تاسیس گردید [۷]. این انجمن واژه کمومتریکس را به این صورت تعریف کرده است: کمومتریکس مجموعه ای از قواعد شیمیایی است که از روشهای ریاضی و آمار جهت طراحی یا انتخاب یک روش مناسب برای آزمایش و یا برای حصول حداکثر اطلاعات شیمیایی از داده های حاصل از یک فرآیند شیمیایی استفاده می کند [۴]. در این تعریف سخنی از کامپیوتر به میان نیامده است زیرا انجام محاسبات پیچیده ریاضی و آماری بدون استفاده از کامپیوتر عملاً کاری غیر ممکن است. استفاده از کامپیوتر و تکنیک های اطلاعاتی برای حل مسائل در شیمی را کموانفورماتیکس^۵ گویند. اصطلاح کموانفورماتیکس اولین بار توسط براون^۶ در سال ۱۹۹۸ عنوان شد و هر چند که اصطلاحات مشابه نیز بدان اطلاق می شود اما عمومیت ندارد [۸, ۹]. کموانفورماتیکس جزو لاینفک کمومتریکس است و امروزه بیشترین کاربرد را در شرکت های دارو سازی در فرآیند طراحی و کشف دارو، ذخیره

1-Jurs, Kowalski, Isenhour and Reilly

2-Svante Wold

3-Bruce Kowalski

4-International Chemometrics Society

5-Chemoinformatics

6-Brown

وبازبایی اطلاعات شیمیایی و آزمایشهای مجازی دارد، هر چند که در صنعت نیز جهت آنالیز داده های شیمیایی به کار گرفته می شود.

کموتریکس از داده های شیمیایی موجود، حداکثر اطلاعات را استخراج کرده و این اطلاعات به منظور اتخاذ تصمیم بهتر و عملکرد سریعتر به دانش تبدیل می شود. با پیشرفت کامپیوتر ها، ریز پردازنده ها و بسته های نرم افزاری، کموتریکس به یک علم توانمند در انجام بسیاری از فعالیتهای شیمیایی، زیستی، دارویی و حتی نانو تکنولوژی تبدیل شده است [۲, ۱۰, ۱۱]. با توجه به پیشرفت و افزایش دستگاههای شیمیایی و در نتیجه آن پیدایش انبوهی از داده ها، پردازش و استخراج اطلاعات از این داده ها کاری وقت گیر و طاقت فرسا می نمود که کموتریکس با ویژگی های خاص خود جهت انجام این کار بیش از پیش مورد توجه و علاقه محققین واقع شد، به طوری که امروزه در زمینه های مختلفی همچون تشخیص الگو، طراحی آزمایش، پردازش سیگنال، رابطه کمی ساختار- فعالیت^۱ (QSAR) و ... به کار می رود [۱].

۱-۳ رابطه کمی ساختار - فعالیت

بررسی خواص بیولوژیکی فراورده های طبیعی و برخی از مشتقات آنها که خواص دارویی دارند، مقدمه ای جهت پیدا کردن وابستگی احتمالی ساختمان شیمیایی با فعالیت بیولوژیکی ارائه نمود. آگاهی از ساختار مولکولی کلید فهم عملکرد و خصوصیات مولکول هاست و این به دلیل ارتباطی

1-Quantitative Structure-Activity Relationship

است که بین ساختار مولکولی و خواص شیمیایی و یا بیولوژیکی یک ترکیب وجود دارد. این اصل که مولکولهای مشابه فعالیت های مشابه دارند، رابطه ساختار-فعالیت^۱ (SAR) نامیده می شود. وقتی خصوصیات ساختاری گونه ها و فعالیت آنها توسط اعداد و ارقام بیان شوند، می توان رابطه ریاضی یا رابطه کمی بین ساختار و فعالیت گونه ها ایجاد کرد. سپس این رابطه ریاضی می تواند برای پیش بینی پاسخ بیولوژیکی یا شیمیایی دیگر ساختارهای شیمیایی، مورد استفاده قرار گیرد. به این فرایند QSAR گفته می شود که در لغت به معنای رابطه کمی ساختار-فعالیت است. هر چند که اصطلاحات مشابه دیگری همچون QSTR, QSPR و QSRR و ... هم به جای QSAR به کار می رود، اما همگی اشاره به رابطه کمی بین ساختار مولکولی و خواص شیمیایی و یا بیولوژیکی مولکول ها دارند [۱۲].

علیرغم اینکه کمتر از ۵۰ سال از ورود QSAR به عرصه هایی چون شیمی زراعی، شیمی دارویی، سم شناسی و شیمی می گذرد [۱۳]، امروزه این دانش به یک ابزار قدرتمند جهت دستیابی به اطلاعات مورد نظر با صرف حداقل زمان و هزینه، تبدیل شده است. که این خود به دو دلیل است: اول آنکه فرض بنیادینی که فعالیت را تابعی از ساختاری می داند که توسط خواصی همچون خواص الکترونی، هندسی و ... مولکولها توصیف می شود، یک فرض محکم و استوار است، و جنبه دیگر، پیشرفت سریع و ارزشمند تکنیک های محاسباتی است که به منظور تعیین و تصحیح بسیاری از متغیرهای موثر در پیش بینی و مدل سازی این روش بوجود آمده اند، می باشد.

۱-۳-۱ تاریخچه پیدایش و تکامل QSAR

در سال ۱۸۶۸ کرام و فریزر^۱ عنوان کردند که فعالیت بیولوژیکی آلکالوئید های مختلف تابعی از ساختمان مولکولی آنهاست. کمتر از یک دهه بعد کرنر^۲ وجود همبستگی بین ساختار مولکولی و خواص فیزیکوشیمیایی بنزن های دو استخلافی را گزارش کرد و در سال ۱۸۹۳ ریچت^۳ نشان داد که سمیت یک سری متنوع از مولکولهای آلی رابطه عکس با حلالیت آن ها در آب دارد. در ابتدای قرن بیستم، مایر^۴ و اورتون^۵ به صورت جداگانه بیان کردند که اثر تسکین دهنده گی گروهی از ترکیبات آلی با ضریب تقسیم آنها بین آب و روغن زیتون متناسب است و استنتاج ترمودینامیکی مبنی بر ارتباط این اثر در برخی ترکیبات فرار با میزان غیر اشباعی آنها، توسط فرگوسن^۶ در سال ۱۹۳۹ انجام پذیرفت. پس از آن کارهای ارزشمند آلبرت، بل و رابلین^۷ اهمیت یونیزه شدن اسیدها و بازهای ضعیف در فعالیت باکتری ها، و هامت^۸ رابطه استخلاف در ترکیبات آلی با میزان واکنش پذیری آنها، را نشان داد.

هانس و فوجیتا^۹ در سال ۱۹۶۲ برای اولین بار QSAR و اساس تئوری آن را مطرح کردند. این

-
- 1-Crum and Fraser
 - 2-Korner
 - 3-Richet
 - 4-Meyer
 - 5-Overton
 - 6-Ferguson
 - 7-Albert, Bell and Roblin
 - 8-Hammet
 - 9-Hansh and Fujita

دانشمندان از روش رگرسیون خطی چند تایی^۱ (MLR) که تنها قادر به بررسی ارتباط خطی بین ساختار و فعالیت است، استفاده کردند. اولین کار QSAR در زمینه بیولوژیکی توسط فری و ویلسون^۲ در سال ۱۹۶۴ انجام پذیرفت. پیشرفتهای اخیر در QSAR منجر به رویکردهای جدیدی در این روش گردیده است که از آن جمله می توان به QSAR، HQSAR، معکوس^۳، QSAR دودویی^۴، 2D-QSAR، 3D-QSAR و 4D-QSAR اشاره کرد [۱۲، ۱۳].

امروزه با به کار گیری اصول QSAR و به کمک شبکه های عصبی مصنوعی که قادر به بررسی ارتباطات غیر خطی هستند، مقالات متعددی در زمینه های مختلف علمی ارائه شده است.

۱-۳-۲ روش و اهداف QSAR

هدف اصلی QSAR ایجاد یک مدل ریاضی بین خصوصیت تجربی اندازه گیری شده و خصوصیات شیمیایی و ساختاری و سپس استفاده از آن برای پیش بینی ویژگی مورد نظر در ترکیبات جدید می باشد [۱]. اولین مرحله این روش انتخاب یک سری داده های شیمیایی یا بیوشیمیایی معتبر است. چنین در نظر گرفته می شود که هر خاصیت شیمیایی یا بیوشیمیایی از ساختمان مولکولها و برهمکنش آنها با محیط اطرافشان پدید می آید. در مرحله بعد به جنبه های مختلف ساختمان مولکولها کمیت داده می شود. این مقادیر محاسبه شده ساختاری، توصیف کننده نامیده می شوند. به این مرحله اصطلاحاً رمزی کردن گفته می شود. از آنجایی که در این مرحله معمولاً تعداد نسبتاً

1-Multiple Linear Regression

2-Free and Wilson

3-Inverse Quantitative Structure-Activity Relationship

4-Binary Quantitative Structure-Activity Relationship

زیادی توصیف کننده های مولکولی محاسبه می شوند، از روشهای انتخاب ویژگی برای انتخاب مفید ترین و مناسب ترین توصیف کننده ها جهت مدل سازی استفاده می شود. سپس با استفاده از روشهای ریاضی، توصیف کننده های انتخاب شده به کمیت خصوصیت شیمیایی یا بیوشیمیایی ربط داده می شوند. نتیجه این محاسبات، مدل های ریاضی است که در آن توصیف کننده ها به عنوان متغیرهای مستقل و خاصیت شیمیایی یا بیوشیمیایی مورد نظر به عنوان متغیر وابسته در نظر گرفته می شود. همچنین می توان در این مرحله از ماشین های بردار پایه^۱ (SVM)، مسیرهای تصمیم گیری^۲ یا شبکه های عصبی مصنوعی^۳ (ANN) برای مدلسازی استفاده کرد، در این صورت به این مرحله، مرحله یادگیری گفته می شود [۱۲] که مدل های با قابلیت پیش بینی چشمگیر ارائه می کند. از آنجا که در این پروژه از الگوریتم ژنتیک^۴ (GA) به عنوان روش انتخاب ویژگی و از شبکه های عصبی مصنوعی در مرحله یادگیری استفاده شده است، این دو تکنیک در فصل سوم توضیح داده شده اند.

در نهایت از مدلها و معادلات بدست آمده برای پیش بینی خاصیت شیمیایی یا بیوشیمیایی ترکیبات جدید و ناشناخته، استفاده می شود. از اطلاعات بدست آمده از QSAR می توان در توجیه و درک مکانیسم عمل گونه در سیستم شیمیایی یا بیوشیمیایی تحت مطالعه، پیش بینی فعالیت زیستی برای مجموعه ای از داروها، پیش بینی خواص موجود در یک خانواده شیمیایی و

1-Support Vector Machines

2-Decision trees

3-Artificial Neural Networks

4-Genetic Algorithm

طراحی ساختمان ترکیباتی با فعالیت مطلوب مورد نظر، بهره گرفت. شکل (۱-۱) مراحل و اهداف QSAR را نشان می دهد.

هزاران مدل بدست آمده از روش QSAR در زمینه های مختلف علمی، گواه اعتبار مفاهیم و کاربرد گسترده این روش می باشد. اما ظریف ترین، چشمگیرترین و مهمترین کاربرد QSAR در فرآیند طراحی دارو^۱ است که در فصل دوم توضیح داده شده است.