

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده شیمی

پایان نامه کارشناسی ارشد شیمی تجزیه

ارائه یک مدل پیش بینی میزان حذف رنگ از پساب‌ها با استفاده از پارامترهای تجربی و توصیف

گرهای ساختاری محاسبه ای

نگارنده:

داوود نادعلی

استاد راهنما:

دکتر منصور عرب چم جنگلی

استاد مشاور:

دکتر ناصر گودرزی

شهریور ماه ۱۳۹۷

تقدیم بہ:

روح پاک پدرم کہ عالمانہ بہ من آموخت تا چگونه در عرصہ زندگی، ایستادگی را تجربہ نمایم.

و

مادرم با بوسہ بردنش کہ وجودش برایم ہمہ مہر است.

و

خواہران نازنینم کہ عطر حضورشان تکرار خوشی ہای من است.

و

دخوشی ہمیشگی ام؛ برادر عزیزم کہ صفایش مایہ می آرامشم است

پاسکزاری

پاس خدای را که سخوران، در ستون او بنامد و شمارندگان، شمردن نعمت های او ندانند و گوشندگان، حق او را گزارش ندهند. و سلام و دور بر محمد و خاندان پاک او، طاهران معصوم، هم آمان که وجودمان و امدار وجودشان است؛ و نفرین سوسه بردستان ایشان تا روز رستاخیز...

بدون شک جایگاه و منزلت معلم، اتمل از آن است که در مقام قدر دانی از زحمت بی ثباتی او، بازبان قاصد دست ناتوان، چیزی بخاریم.

اما از آنجایی که تجلیل از معلم، پاس از انسانی است که هدف و غایت آفرینش را تائین می کند و سلامت امانت بی را که بدستش سپرده اند، تقصیر؛ بر حسب وظیفه و از باب ”من لم یسکر المنعم من المخلوقین لم یسکر الله عزوجل“:

از پدر و مادر عزیزم، این دو معلم بزرگوارم... که همواره بر کوه تاهی و درشتی من، قلم عنقوشیده و گریانه از کنار غفلت هایم گذشته اند و در تمام عرصه های زندگی یار و یاور بی چشم داشت برای من بوده اند؛

از استاد باکالات و شایسته؛ جناب آقای دکتر منصور عرب چم جملگی که در کمال سعادت، با حسن خلق و فروتنی، از هیچ کجی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این پیمان نامه را بر عهده گرفتند؛

از استاد ارجمندم، جناب آقای دکتر ناصر کوردزی که زحمت مشاوره این پیمان نامه را منتقل شدند و همچنین از اساتید فرزانه و دلسوز؛ سرکار خانم دکتر مصدر الامور و جناب آقای دکتر باقریان؛

زحمت دآوری این پیمان نامه را منتقل شدند؛ کمال تشکر و قدر دانی را دارم.

باشد که این خردترین، بخشی از زحمت آمان را پاس گوید.

تعهد نامه

اینجانب **داوود نادعلی** دانشجوی دوره کارشناسی ارشد رشته شیمی تجزیه دانشکده شیمی دانشگاه شاهرود نویسنده پایان نامه ارائه یک مدل پیش بینی میزان حذف رنگ از پساب ها با استفاده از پارامترهای تجربی و توصیف گره‌های ساختاری محاسبه‌ای تحت راهنمایی **دکتر منصور عرب چم جنگلی** و مشاوره **دکتر ناصر گودرزی** متعهد می‌شوم :

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

امضای دانشجو

تاریخ

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .

چکیده

فاضلاب‌های نساجی شامل انواع مختلفی از رنگ‌ها هستند که بدون انجام فرایند مناسب برای دفع آن‌ها باعث ایجاد سمیت در میکروارگانیسم‌ها و حیات آبریان می‌شود. بی توجهی به پخش رنگ‌ها تاثیر منفی روی محیط زیست و زندگی انسان‌ها می‌گذارد.

در این پروژه از مدل QSPR مبتنی بر شبکه عصبی تنظیم بایزین برای مدل‌سازی و پیش‌بینی دقیق راندمان رنگبری بر اساس ساختار رنگ استفاده شد. برای این منظور رنگبری فتوکاتالیستی ۳۱ رنگ مختلف با استفاده از TiO_2 تحت تابش UV مورد بررسی قرار گرفت. راندمان رنگبری تحت تابش دهی اشعه‌ی ماوراء بنفش به مدت ۵، ۱۰ و ۱۵ دقیقه در شرایط ثابت شامل مقدار کاتالیست (۰/۰۱۰ گرم)، حجم نمونه (۵۰/۰ میلی لیتر) و در $\text{pH} = 7/0$ اندازه گیری شد. ساختار رنگ‌ها با استفاده از نرم‌افزار دراگون به توصیف‌کننده‌های مولکولی تبدیل شد و موثرترین توصیف‌کننده‌ها انتخاب شدند و به عنوان ورودی شبکه عصبی مورد استفاده قرار گرفتند. پس از به‌دست آوردن مدل BR-ANN به عنوان مدل بهینه، توانایی پیشگویی مدل با استفاده از سری آزمون و روش رد مرحله‌ای تک‌تک مورد ارزیابی قرار گرفت. مجذور ضریب همبستگی برای سری آزمون و روش رد مرحله‌ای تک‌تک به ترتیب ۰/۹۵۵۳ و ۰/۹۳۲۰ به‌دست آمد. نتایج به‌دست آمده توانایی مدل پیشنهادی را در پیش‌بینی میزان رنگبری به خوبی نشان داد.

کلمات کلیدی: شبکه عصبی مصنوعی، TiO_2/UV ، فرایندهای اکسیداسیون پیشرفته، درصد رنگبری، QSPR

نتایج حاصل از این پایان نامه در یک پوستر تحت عنوان:

1. “Application of artificial neural network for prediction of dyes photodegradation efficiency in UV/TiO₂ condition based on structural information”

در ششمین سمینار دوسالانه کمومتریکس ایران در آبان ماه ۱۳۹۶ پذیرفته و ارائه شد.

فهرست مطالب:

۱- فصل اول	۱
۱-۱ - آلودگی‌های آب	۳
۱-۲ - پساب و انواع آن	۳
۱-۲-۱ - آلاینده‌های رنگی	۴
۱-۲-۲ - تقسیم‌بندی رنگ‌ها	۵
۱-۲-۳ - طبقه‌بندی رنگ‌ها بر اساس ساختار شیمیایی	۵
۱-۲-۴ - طبقه‌بندی مواد رنگزا بر اساس نوع کاربرد	۶
۱-۳ - مروری بر کارهای انجام شده و ضرورت تحقیق	۷
۲- فصل دوم	۱۱
۲-۱ - حذف رنگ	۱۲
۲-۲ - روش‌های تصفیه پساب‌های رنگی	۱۲
۲-۲-۱ - تصفیه فیزیکی	۱۳
۲-۲-۲ - تصفیه شیمیایی	۱۳
۲-۲-۳ - روش‌های زیستی	۱۳
۲-۲-۴ - روش‌های اکسیداسیون شیمیایی	۱۴
۲-۲-۴-۱ - انواع روش‌های اکسیداسیون پیشرفته	۱۶
۲-۳ - فتوکاتالیست‌های نیمه هادی	۱۹
۲-۳-۱ - فرایند انتقال الکترونی بر روی سطح کاتالیست	۱۹
۲-۳-۲ - مکانیسم تخریب فتوکاتالیستی در حضور نور فرابنفش	۲۱

- ۲۲.....TiO₂ فتوکاتالیست ۴-۲
- ۲۳.....TiO₂ ویژگی‌های ۱-۴-۲
- ۲۴.....TiO₂ کاربردهای ۲-۴-۲
- ۲۴.....کموتریکس ۵-۲
- ۲۷.....(QSPR) خاصیت ساختار-ارتباط کمی ۶-۲
- ۲۷.....فراهم کردن سری داده‌ها ۱-۶-۲
- ۲۸.....بهینه‌سازی ساختار هندسی ترکیبات ۲-۶-۲
- ۳۰.....محاسبه توصیف‌کننده‌ها ۳-۶-۲
- ۳۱.....انتخاب توصیف‌کننده‌های مناسب ۴-۶-۲
- ۳۱.....انتخاب متغیر به روش رگرسیون خطی ۱-۴-۶-۲
- ۳۳.....ایجاد مدل‌های آماری ۵-۶-۲
- ۳۴.....مقدمه ای بر شبکه عصبی ۶-۶-۲
- ۳۵.....تاریخچه ۱-۶-۶-۲
- ۳۶.....ساختمان و عملکرد نرون طبیعی ۲-۶-۶-۲
- ۳۷.....ساختمان و عملکرد نرون مصنوعی ۳-۶-۶-۲
- ۳۷.....تابع انتقال ۴-۶-۶-۲
- ۳۹.....ساختارهای شبکه ۵-۶-۶-۲
- ۴۰.....آموزش شبکه‌های جلو سو با تکنیک پس‌انتشار ۶-۶-۶-۲
- ۴۱.....ارزیابی مدل ۷-۲
- ۴۱.....استفاده از پارامترهای آماری ۱-۷-۲

- ۲-۷-۲ - استفاده از نمودار برگشتی ۴۷
- ۲-۷-۳ - استفاده از نمودار خطای باقی مانده ها ۴۷
- ۲-۷-۴ - استفاده از سری آزمون ۴۷
- ۳- فصل سوم ۴۹
- ۳-۱ - مواد شیمیایی به کار رفته در پروژه ۵۰
- ۳-۲ - تهیه محلول ها ۵۷
- ۳-۳ - دستگاهوری ۵۷
- ۳-۴ - نرم افزارهای مورد استفاده ۵۸
- ۳-۴-۱ - بسته نرم افزاری Hyperchem 8.0.5 ۵۸
- ۳-۴-۲ - بسته نرم افزاری Dragon 5.5 ۵۹
- ۳-۴-۳ - بسته نرم افزاری SPSS 21 ۵۹
- ۳-۴-۴ - نرم افزار MATLAB ۵۹
- ۳-۵ - روش کار در آزمایش رنگبری فتوکاتالیستی ۶۰
- ۳-۶ - محاسبات و مدل سازی QSPR ۶۶
- ۳-۶-۱ - سری داده ها ۶۶
- ۳-۶-۲ - رسم و بهینه سازی ساختار مولکول ها ۶۶
- ۳-۶-۳ - محاسبه توصیف کننده ها ۶۷
- ۳-۶-۴ - انتخاب بهترین توصیف کننده ها با توجه درصد مشارکت در مدل های تصادفی حاصل از رد گروهی ۶۸
- ۳-۶-۵ - بررسی وجود هم خطی بین بهترین توصیف کننده ها ۷۱
- ۳-۶-۶ - مدل سازی شبکه عصبی مصنوعی با توصیف کننده های انتخاب شده توسط روش فراوانی توصیف کننده ها ۷۲

- ۳-۶-۶-۱ - انتخاب تعداد لایه‌های پنهان..... ۷۳
- ۳-۶-۶-۲ - انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع آموزش و نوع تابع انتقال..... ۷۳
- ۳-۶-۶-۳ - ساختار شبکه عصبی مصنوعی بهینه شده..... ۷۷
- ۳-۷-۷ - ارزیابی مدل..... ۷۹
- ۳-۷-۱ - ارزیابی مدل با استفاده از داده‌های سری آزمون..... ۷۹
- ۳-۷-۲ - ارزیابی مدل غیر خطی به روش رد مرحله‌ای تک‌تک..... ۸۱
- ۳-۸-۸ - ارزیابی مدل با استفاده از پارامترهای آماری..... ۸۶
- ۳-۹-۹ - ارزیابی مدل ارائه شده با استفاده از آزمون χ^2 -تصادفی..... ۸۷
- ۳-۱۰-۱۰ - بررسی ارتباط توصیف‌کننده‌های وارد شده در مدل با درصد رنگبری..... ۸۸
- ۳-۱۱-۱۱ - بررسی میزان مشارکت توصیف‌کننده‌های منتخب در شبکه عصبی..... ۹۴
- ۴-۴-۴ - فصل چهارم..... ۹۷
- ۴-۱-۱ - نتیجه‌گیری..... ۹۸
- ۴-۲-۲ - آینده نگری..... ۱۰۰

فهرست شکل‌ها:

- شکل (۱-۲) دسته بندی واکنش‌های فوتونی کاتالیست [۲۳]..... ۲۰
- شکل (۲-۲) مسیرهای مختلف بازیابی الکترون و حفره در یک نیمه هادی [۲۲, ۲۴, ۲۵]..... ۲۲
- شکل (۳-۲) ساختارهای بلوری TiO_2 ۲۲
- شکل (۴-۲) ارتباط کمومتریکس و علوم [۳۲]..... ۲۵
- شکل (۵-۲) ساختار یک نرون زیستی و شبکه عصبی مصنوعی ۳۷
- شکل (۶-۲) تابع انتقال لگاریتم سیگموئید ۳۸
- شکل (۷-۲) تابع انتقال تانژانت سیگموئید ۳۹
- شکل (۱-۳) نمودارهای الف) تعداد توصیف‌کننده، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین به همراه تابع انتقال لگاریتم سیگموئید ۷۵
- شکل (۲-۳) نمودارهای الف) تعداد توصیف‌کننده، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE برای تابع آموزش تنظیم بایزین به همراه تابع انتقال تانژانت سیگموئید ۷۶
- شکل (۳-۳) ساختار شبکه عصبی بهینه شده ۷۸
- شکل (۴-۳) نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی برای سری تست ۸۱
- شکل (۵-۳) نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی به روش رد مرحله‌ای تک‌تک برای داده‌ها ۸۵
- شکل (۶-۳) نمودار باقیمانده‌ها بر حسب مقادیر تجربی برای کل داده‌ها ۸۶
- شکل (۷-۳) مشارکت توصیف‌کننده‌ها در شبکه عصبی بهینه ۹۵

فهرست جدول‌ها:

- جدول (۱-۲) اکسید کننده‌های متداول [۱۵] ۱۵
- جدول (۱-۳) مواد شیمیایی مورد استفاده در این پروژه ۵۰
- جدول (۲-۳) میانگین درصد رنگبری در زمان‌های مختلف ۶۱
- جدول (۳-۳) فهرست توصیف کننده‌های محاسبه شده توسط نرم‌فزار Dragon ۶۷
- جدول (۴-۳) توصیف کننده‌های به دست آمده از تقسیم‌بندی مختلف داده‌ها ۶۹
- جدول (۵-۳) توصیف کننده‌های انتخاب شده بر اساس فراوانی توصیف کننده‌ها ۷۰
- جدول (۶-۳) ماتریس همبستگی کل توصیف کننده‌های انتخاب شده بر اساس فراوانی توصیف کننده‌ها ۷۱
- جدول (۷-۳) مقادیر عامل افزایش واریانس توصیف کننده‌های انتخاب شده توسط روش فراوانی توصیف کننده‌ها ۷۲
- جدول (۸-۳) تابع و پارامترهای شبکه بهینه توصیف کننده‌های انتخاب شده بر اساس فراوانی توصیف کننده‌ها ۷۷
- جدول (۹-۳) توابع و مقادیر پارامترهای بهینه شده شبکه عصبی ۷۸
- جدول (۱۰-۳) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون ۸۰
- جدول (۱۱-۳) نتایج حاصل از ارزیابی مدل با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها ۸۲
- جدول (۱۲-۳) پارامترهای آماری برای مدل برتر طراحی شده توسط شبکه عصبی مصنوعی ۸۷
- جدول (۱۳-۳) مقادیر R^2 برای سری آزمون با استفاده از آزمون Y-تصادفی ۸۸
- جدول (۱۴-۳) نمایش ارتباط خاصیت با مقدار توصیف کننده MATS1m برای بعضی از ترکیبات مورد مطالعه ۹۱
- جدول (۱۵-۳) نمایش ارتباط خاصیت با مقدار توصیف کننده CH_3-X برای بعضی از ترکیبات مورد مطالعه ۹۳

فصل اول

مقدمه

مقدمه

آب مهمترین منبع طبیعی در جهان است، به طوری که بدون آن حیات نمی تواند وجود داشته باشد و اکثر صنایع نیز قادر به ادامه ی کار نخواهند بود. منابع طبیعی آب محدود هستند و لزوم حفاظت از منابع و تأمین آب شرب امری ضروری است. آب به عنوان یکی از اصلی ترین نیازهای بشری همواره نقش مهمی ایفا نموده و یکی از مشکلاتی که همواره گریبان گیر بسیاری از جوامع می باشد، مشکل تأمین آب سالم و بهداشتی است. در کشور ما نیز مشکل آلودگی آب وجود داشته و ارائه ی راه کارهای مناسب تصفیه آب جزء اولویتهای تحقیقاتی می باشد. آب یک سرمایه ی ملی است. برای روشن شدن این موضوع کفایت توجه کنیم که منابع ارزان و فراوان تولید انرژی به منظور جانشینی با سوخت های فسیلی وجود دارند اما در حال حاضر هیچ منبعی به عنوان جانشین برای منابع آبی وجود ندارد و امروزه استفاده بهینه از منابع آبی مرتباً تأکید می شود [۱].

به طوری که روشن است آب های مصرف شده در زندگی ما به نحوی به منابع اولیه ی خود برگشت داده می شوند ولی در غالب موارد آب برگشتی همان آب اولیه نیست بلکه به همراه خود مقادیر فراوانی از ناخالصی مختلف حمل می کند که تا اندازه ای در آب محلول هستند. این حلالیت به دما، فشار، pH، پتانسیل شیمیایی و غلظت نسبی دیگر مواد در آب بستگی دارد. در طبیعت این عوامل چنان بهم مربوط هستند که کمتر می توان حلالیت ماده ای را در آب به طور دقیق پیش بینی کرد. با توجه به اینکه آب به صورت یک موهبت الهی تقریباً به طور رایگان در اختیار جوامع بشری قرار گرفته است، لذا غالباً در نگهداری و آلوده نکردن آن سعی و توجه کافی مبذول نمی شود و تنها وقتی که برای تهیه ی آن با ویژگی های معین با مشکلاتی مواجه می شویم به اهمیت و ارزش وجودی آن پی می بریم.

۱-۱ - آلودگی‌های آب

آلودگی آب عبارت است از افزایش مقدار هر معرف اعم از شیمیایی، فیزیکی یا بیولوژیکی که موجب تغییر خواص و نقش اساسی آن در مصارف ویژه‌اش شود. تمام فاضلاب‌های صنعتی و خانگی در تغییر کیفی جریان‌های آب مؤثرند. وقتی این تغییر به حدی رسید که دیگر نتوانیم از آب در مصارف عادی استفاده کنیم، آب را آلوده می‌گویند. همه‌ی آب‌های طبیعی حاوی آلاینده‌های متنوعی هستند که از فرایندهای فرسایش، شستشو و تماس با هوا سرچشمه می‌گیرند.

۱-۲ - پساب و انواع آن

با توسعه شهرها و افزایش جمعیت آن‌ها و گسترش صنایع، روز به روز بر اهمیت کنترل آلودگی محیط زیست افزوده می‌شود. فاضلاب‌ها به‌عنوان یکی از عوامل آلودگی محیط زیست، حاصل فعالیت‌های انسان و صنایع هستند که بسته به منبع تولید، دارای کیفیت و محتوای متفاوتی از مواد آلی و معدنی محلول و معلق و... می‌باشند. آن‌ها باید قبل از تخلیه به طریق بهداشتی جمع‌آوری، تصفیه و مجدداً به گردش آب در طبیعت برگردانده شوند [۲].

یکی از مهم‌ترین آلاینده‌های آب، رنگ‌ها هستند. به گونه‌ای که می‌توان گفت، آب‌های آلوده شیمیایی شامل دو نوع اصلی از آلوده‌کننده‌ها هستند:

- یون‌های فلزات واسطه و سنگین مانند سرب، جیوه، روی، نیکل، کروم و کادمیوم
- ترکیبات آلی مانند رنگدانه‌ها

با توجه به توسعه شهرنشینی و صنایع که منجر به افزایش میزان فاضلاب و پساب صنایع تولیدی گردیده است. فلزات سنگین عمدتاً از طریق دفع نادرست و غیربهداشتی فاضلاب شهری و پساب صنعتی وارد محیط زیست می‌گردند. مرگ و میرهای آبیان در اثر تخلیه پساب‌های محتوی فلزات سنگین در دنیا و ایران بی‌سابقه نیست. وجود فلزات سنگین در غلظت بیش از استاندارد در آب شرب باعث عوارض مختلف نظیر مسمومیت، حساسیت شدید، ضایعات کروموزومی، سنگ کلیه و انواع سرطان می‌گردد. یکی از کارشناسان محیط زیست، آلودگی محیط (مخصوصاً آب) با فلزات سنگین را به عنوان بزرگ‌ترین گناهی که بشر در طبیعت انجام می‌دهد ارزیابی نموده است [۳].

۱-۲-۱ - آلاینده‌های رنگی

یکی دیگر از عوامل آلوده کننده‌ی آب، مواد رنگی (به ویژه صنعتی) می‌باشند که در صنایع مختلف نظیر نساجی و ... مورد استفاده قرار می‌گیرند. بخش بزرگی از ترکیبات آلی که باعث ایجاد آلودگی در آب‌های طبیعی می‌گردند، مواد رنگزای شیمیایی هستند که به صورت صنعتی و خانگی مورد استفاده قرار می‌گیرند. حضور مواد رنگزای شیمیایی علاوه بر آنکه بر آلودگی منابع آبی تاثیر می‌گذارند، با متوقف کردن تولید اکسیژن و جلوگیری از نفوذ نور خورشید موجب مرگ موجودات زنده و وارد آمدن صدمات جدی به محیط زیست می‌گردند [۴].

از دیدگاه شیمی آلی ترکیبات اشباع شده و متشکل از پیوندهای sp^3 بی‌رنگ هستند، زیرا جذب آن‌ها خارج از منطقه‌ی قابل رؤیت طیف می‌باشد. سیستم‌های اتصالی دوگانه در ترکیبات حاوی گروه‌های $C=O$ ، $C=N$ ، $N=O$ و $C=C$ به‌ویژه زمانی که این نوع پیوندها با یکدیگر در تبادل باشند باندهای جذبی کم و بیش به منطقه‌ی قابل رؤیت سوق داده می‌شود و باعث ظهور رنگ می‌شوند [۵].

۱-۲-۲ - تقسیم‌بندی رنگ‌ها

مواد رنگزای می‌توان براساس ساختار شیمیایی و نوع کاربرد طبقه‌بندی کرد. در طبقه‌بندی بر اساس نوع ساختار شیمیایی، ماده رنگزا از نظر گروه‌های رنگ‌ساز تقسیم‌بندی می‌شود. در نوع دیگری از دسته بندی مواد رنگزا بر طبق نوع کاربرد و براساس اینکه چگونه در فرایند رنگرزی به کار می‌روند طبقه بندی می‌شوند.

۱-۲-۳ - طبقه‌بندی رنگ‌ها بر اساس ساختار شیمیایی

فتالوسیانین‌ها: گروه مهمی از مواد رنگزا با رنگ‌های مختلف از آبی تا سبز- زرد را شامل می‌شوند. آنتراکینون‌ها: کروموژن همه مواد رنگزای قرمز طبیعی براساس بنیان آنتراکینون می‌باشد. آنتراکینون بعد از مواد رنگزای آزو مهم‌ترین طبقه از مواد رنگزا را تشکیل می‌دهند.

تری آریل متان: سیستم کروموفوری رنگ‌های تری آریل متان از کربن متصل به سه حلقه آروماتیک تشکیل شده است.

نیترو- نیتروزو: گروه نیترو در هر رنگی با هر گونه ساختار شیمیایی می‌تواند وجود داشته باشند ولی در اغلب رنگ‌های نیترو این عامل به‌عنوان تنها سیستم رنگزا محسوب می‌شود. که به کمک گروه‌های آکسو کروم نظیر هیدروکسیل یا آمینو و در نتیجه رزونانس حاصله بین دو یا چند فرم توتومری رنگ نهایی ایجاد می‌شود.

آزوها: مواد رنگزای آزو از لحاظ تجاری مهم‌ترین طبقه مواد رنگزا هستند که تقریباً بیش از ۵۰ درصد کل مجموع مواد رنگزای تولید شده در جهان را تشکیل می‌دهند. گروه $-N=N-$ گروه آزو یا دی ایمید

نامیده می‌شود. به ترکیب‌هایی آزو گفته می‌شود که دارای شکل مولکولی $R-N=N-R'$ باشند. که در آن R و R' هر کدام می‌توانند گروه‌های آلیفاتیک یا آروماتیک باشند.

۱-۲-۴ - طبقه‌بندی مواد رنگزا بر اساس نوع کاربرد

رنگزاهای اسیدی: رنگینه‌های اسیدی ترکیبات آنیونی هستند که برای رنگ کردن فیبرهای نیتروژن دار مانند ابریشم، پشم، پلی‌آمیدهای سنتزی به کار می‌روند. نام رنگینه‌های اسیدی به خاطر فرایند رنگری است. محیط رنگری این دسته از رنگینه‌ها شامل سولفوریک اسید، فرمیک اسید و استیک اسید می‌باشد. از انواع این رنگ‌ها می‌توان به آبی اسیدی^۱ اشاره کرد.

رنگزاهای مستقیم: این نوع رنگ‌ها معمولاً در حمام رنگری خنثی یا کمی قلیایی در نزدیک نقطه جوش با افزایش سدیم کلرید یا سدیم سولفات به کار گرفته می‌شوند. این رنگینه‌ها برای کتان، کاغذ، چرم، پشم ابریشم و نایلون استفاده می‌شوند و همچنین به‌عنوان شناساگر pH به کار می‌روند. نارنجی مستقیم^۲ از انواع این رنگ‌هاست.

رنگزاهای بازی: یک طبقه از رنگ‌ها با گروه‌های بازی در ساختار شیمیایی می‌باشند. این رنگ‌ها برای رنگری فیبرهای اسید دار به کار می‌روند. رنگ‌های بازی عموماً در صنایع کاغذ، صنعت چاپ، صنایع آرایشی و صنایع غذایی به کار برده می‌شوند. از انواع شناخته شده رنگزاهای بازی می‌توان ملاکیت سبز^۳ و آبی متیلن^۴ را نام برد.

رنگزاهای دندان‌های: رنگینه‌های دندان‌های رنگینه‌هایی هستند که به طور طبیعی میل ترکیبی به فیبرها ندارند، بنابراین دندان‌های یک ترکیب شیمیایی مورد نیاز است تا با فیبر در طول فرایند رنگری پیوند برقرار

^۱ -Acid Blue

^۲ -Direct orange

^۳ -Malachite green

^۴ -Methylen Blue

کند. بسیاری از این رنگ‌ها شامل گروه‌های اسید سولفوریک و اتم فلزی مانند کروم، کبالت یا نیکل هستند. رنگزاهای دندانه‌ای برای رنگرزی پشم، ابریشم، نایلون و چرم به کار می‌روند. اریکوکروم سیاه^۱ از انواع این رنگ‌هاست [۶،۷].

۱-۳ - مروری بر کارهای انجام شده و ضرورت تحقیق

یکی از مسائل نگران‌کننده در سال‌های اخیر بروز بحران آب در جهان است. ایران هم جزء کشورهای است که با بحران کمبود آب مواجه شده است و همین امر ضرورت تصفیه پساب‌ها را ایجاب می‌کند. در سال‌های اخیر تلاش‌های فراوانی برای یافتن روش‌هایی برای تصفیه و حتی استفاده دوباره از این پساب‌ها صورت گرفته است. اخیراً برای رنگبری از پساب‌ها از روش‌های AOPs^۲ استفاده می‌شود به این دلیل که این روش‌ها توانایی زیادی در تبدیل آلاینده‌های آلی به مواد معدنی دارند. محققان مختلف در طول سال‌های ۲۰۱۸-۲۰۰۰ توانسته‌اند با استفاده از مطالعات QSPR مدل‌های خطی و غیر خطی متفاوتی را برای رنگبری و تخریب رنگ‌ها ارائه دهند که این امر باعث کاهش هزینه‌ها و اتلاف زمان می‌شود. تعدادی از این مطالعات در ادامه مرور می‌شود.

چارچلی و همکارانش^۳ در سال ۲۰۰۰ با استفاده از مطالعات QSPR رنگبری و تخریب رنگ‌های آزو محلول در آب را بررسی نمودند. آن‌ها در این مطالعه از روش رگرسیون خطی چندگانه^۴ (MLR) استفاده نمودند. مدل مستخرج نشان می‌دهد که هرچه نسبت اندازه مولکول به بار یونی آن بیشتر باشد تجزیه زیستی رنگ‌های آزو بیشتر است [۸].

^۱ Eriochrome black

^۲ Advanced Oxidation Processes

^۳ Churchley and et al.

^۴ Multiple Linear Regression

سوزوکی و همکارانش^۱ در سال ۲۰۰۱ به منظور کشف رابطه کمی بین تخریب زیستی رنگ‌های آزو با ساختار شیمیایی آن‌ها، ۲۵ رنگ آزو سولفوناتی را با استفاده از روش آنالیز مولفه‌های اصلی^۲ مورد بررسی قرار دادند. تخریب رنگ‌ها با استفاده از تکنیک تشخیص آنالیز رگرسیون دوگانه^۳ بر پایه توصیف‌کننده‌های ساختاری مدل شد. آن‌ها تاثیر موقعیت گروه هیدروکسی را در تخریب زیستی رنگ‌های آزو نشان دادند [۹].

کیوزیک و همکارانش^۴ در سال ۲۰۱۳ در چارچوب مطالعات QSPR برای ایجاد مدل‌هایی که ارتباط دهنده‌ی پارامترهای اکولوژیک با ساختار رنگ‌های مورد نظر باشد مطالعاتی را انجام دادند. آن‌ها مطالعاتی را بر روی ۹ رنگ راکتیو تری آزین انجام دادند و نشان دادند که فوتولیز رنگ‌های تری آزین مورد مطالعه به شدت به نوع کروموفور وابسته است [۱۰].

لی و همکارانش^۵ در سال ۲۰۱۳ با استفاده از مطالعات QSPR رنگبری کاتالیستی ۲۸ رنگ آزوی آنیونی با استفاده از کاتالیست فیبر آمید اکسیماتید پلی آکرلیونیتریل^۶ (PAN) کمپلکس شده با آهن را بررسی کردند. در این کار از روش MLR استفاده شد. M_w/S (نسبت وزن مولکولی به تعداد گروه‌های سولفونات) و $N_{N=N}$ (تعداد پیوند آزو) به‌عنوان مهم‌ترین عامل رنگبری در نظر گرفته شد آن‌ها همچنین نشان دادند که همبستگی منفی معنی‌دار بین M_w/S یا $N_{N=N}$ و درصد رنگبری یا حذف کل کربن آلی (TOC) وجود دارد. توانایی پیش‌بینی معادلات QSPR با استفاده از روش اعتبارسنجی متقاطع^۷ (CV) مورد ارزیابی قرار گرفت [۱۱].

¹ Suzuki and et al.

² principal components analysis

³ discriminant technique of two-value regression analysis

⁴ Kusic and et al.

⁵ Li and et al.

⁶ Amidoximated polyacrylonitrile (PAN) fiber

⁷ Cross-Validation

هر چند مدل‌های گزارش شده، مدل‌های پیشگویی می‌باشند ولی هیچ‌کدام از مدل‌ها توانایی پیشگویی میزان رنگبری از رنگ‌های مختلف را در شرایط تجربی ندارد و به عبارت بهتر پارامترهای ساختاری رنگ‌ها در فرایندهای مدل‌سازی در نظر گرفته نشده است. اخیراً لی و همکارانش فرایند رنگبری کاتالیستی رنگ‌های آزو آنیونی را بر اساس ساختار رنگ‌ها، مدل‌سازی و گزارش نموده‌اند. هر چند در این گزارش یک مدل MLR ساده برای پیش‌گویی رنگبری کاتالیستی ارائه شده است ولی مدل فقط محدود به رنگ‌های آزو آنیونی بوده و نمی‌توان آن را برای طیف وسیعی از رنگ‌ها استفاده کرد. بنابراین تا کنون مدل QSPR پیشگو که بتواند میزان رنگبری فتوکاتالیستی رنگ‌ها، با ساختارهای متنوع و گروه‌های مختلف را پیش‌بینی نماید ارائه نگردیده است. در تحقیق حاضر ارتباط کمی ساختار-خاصیت برای پیش‌بینی میزان رنگبری از رنگ‌های مختلف به کار گرفته شد. ۳۱ رنگ از گروه‌های مختلف با استفاده از روش فتوکاتالیست، رنگبری شدند. سپس با استفاده از پارامترهای تجربی و توصیف‌کننده‌های ساختاری، مدل شبکه عصبی مصنوعی برای پیش‌بینی درصد رنگبری رنگ‌های مختلف ارائه گردید. در تحقیقات صورت گرفته مدلی برای پیش‌بینی درصد رنگبری رنگ‌های مختلف ارائه نشده است و این کار به‌عنوان تحقیقی نوین در ارتباط با رنگ‌های مورد مطالعه است. در این پژوهش از مدل شبکه عصبی مصنوعی استفاده شد زیرا توانایی مناسبی در برقراری ارتباط غیر خطی بین ساختار-خاصیت و نیز پیش‌بینی موارد مشابه نشان داد.

فصل دوم

تئوری

۱-۲ - حذف رنگ

فاضلاب‌های نساجی دارای انواع مختلفی از رنگ‌ها هستند که با ورود به محیط زیست در دسترس جانداران قرار می‌گیرند. که سلامت انسان‌ها و محیط زیست را به خطر می‌اندازند و همچنین بر فعالیت فتوسنتزی گیاهان اثر می‌گذارند.

در حذف رنگ‌ها فرایندهایی مانند اکسیداسیون، هیدرولیز و واکنش‌های شیمیایی موجب تولید محصولات جانبی در فاضلاب‌ها می‌شوند. تقریباً ۲۰٪ از کل رنگ‌های تولید شده در جهان از طریق فرایند رنگرزی هدر می‌روند که به فاضلاب‌های نساجی وارد می‌شوند. حذف مواد شیمیایی و آلاینده‌های آلی از پساب‌ها که سلامت انسان‌ها و محیط زیست را به خطر می‌اندازند امری ضروری است [۱۲].

۲-۲ - روش‌های تصفیه پساب‌های رنگی

روش‌های حذف آلودگی آب در چهار دسته کلی طبقه‌بندی می‌شوند [۱۳]:

۱- تصفیه فیزیکی

۲- تصفیه شیمیایی

۳- روش‌های زیستی

۴- روش‌های اکسیداسیون پیشرفته (AOPs)

۲-۲-۱ - تصفیه فیزیکی

تصفیه فیزیکی مقدماتی یا مکانیکی به فرایندهایی اطلاق می‌شود که در آن‌ها از خواص فیزیکی برای جداسازی مواد جامد با اندازه‌های مختلف در فاضلاب استفاده می‌گردد. در این فرایندها با استفاده از اصول حاکم بر عملیات فیزیکی بخش عمده‌ای از آلودگی موجود در پساب، به خصوص ذرات معلق و کلوییدها مجتمع و جدا می‌شوند. فرایندهایی مانند غربال کردن و ته نشینی در زمره‌ی این روش قرار می‌گیرد [۲].

۲-۲-۲ - تصفیه شیمیایی

تصفیه شیمیایی فاضلاب بر اساس کاربرد مواد شیمیایی برای حذف و تبدیل مواد آلاینده استوار است. تصفیه شیمیایی به طور عمده در تصفیه فاضلاب‌های صنعتی به کار می‌رود. از مهم‌ترین فرایندهای شیمیایی برای حذف مواد محلول، خنثی‌سازی، اکسیداسیون، و تعویض یونی و برای حذف مواد نامحلول، انعقاد و لخته سازی و شناور سازی با استفاده از مواد منعقد کننده می‌باشد. از مهم‌ترین معایب تصفیه شیمیایی، افزایش مواد محلول در فاضلاب به دلیل افزودن مواد شیمیایی و هزینه بالای آن‌ها می‌باشد [۲].

۲-۲-۳ - روش‌های زیستی^۱

تصفیه زیستی فرایندی است که در آن از میکرو ارگانیسم‌ها (عوامل بیولوژیکی) برای تبدیل و تجزیه مواد آلاینده موجود در فاضلاب استفاده می‌شود.

اهداف عمده تصفیه بیولوژیکی عبارتند از [۲]:

۱- لخته سازی و جداسازی مواد جامد کلوییدی غیر قابل ته نشینی

¹ Biological method

۲- تثبیت و کاهش محتوای مواد آلی

۳- تجزیه‌ی مواد مغذی مانند نیتروژن و فسفر

۴- جداسازی ترکیبات آلی نادری که ممکن است سمی نیز باشند.

۵- جداسازی یا کاهش غلظت ترکیبات آلی و غیر آلی موجود

گرچه تصفیه بیولوژیکی فناوری مقرون به صرفه و سازگار با محیط زیست است اکثر رنگ‌های سنتزی دارای ماهیت آروماتیکی پایدار و پیچیده‌ای هستند که در برابر تجزیه بیولوژیکی مقاومند بدین منظور تحقیقات به سوی یافتن راه‌هایی جهت ترکیب روش‌های زیستی با روش‌های دیگری نظیر فرایندهای اکسیداسیون پیشرفته، جذب بر روی جامدها (کربن فعال، خاک و غیره) و غیره متمرکز شده است [۱۴].

۲-۲-۴ - روش‌های اکسیداسیون شیمیایی

طی فرایند اکسیداسیون، انتقال الکترون از یک ماده به ماده دیگر صورت می‌گیرد، که در جدول (۲-۱) برخی از گونه‌های اکسید کننده متداول آورده شده است همان‌طور که مشاهده می‌شود OH از قدرت اکسیدکنندگی بالایی برخوردار است.

جدول (۲-۱) اکسید کننده‌های متداول [۱۵]

عامل اکسید کننده	قدرت اکسیدکنندگی (V)
فلوئور	۳/۰۳
هیدروکسیل رادیکال	۲/۸۰
اکسیژن اتمی	۲/۴۲
ازون	۲/۰۷
هیدروژن پراکسید	۱/۷۷
پرمنگنات	۱/۶۷
هیپو برمواکسید	۱/۵۹
دی کلرو اکسید	۱/۵۰
اسید هیپو کلرو	۱/۴۹
اسید هیپو یدید	۱/۴۵
کلر	۱/۳۶
برم	۱/۰۹
ید	۰/۵۴

به‌طور کلی روش‌های اکسیداسیون شیمیایی به دو دسته زیر تقسیم بندی می‌شوند:

الف- روش‌های اکسیداسیون کلاسیک

در این روش یک عامل اکسید کننده به آب حاوی آلاینده اضافه می‌شود و عمل اکسیداسیون صورت می‌گیرد، کلر، پتاسیم پرمنگنات، اکسیژن، هیدروژن پراکسید و ازون از عامل‌های اکسید کننده متداول مورد استفاده در این روش می‌باشد.

ب- روش‌های اکسیداسیون پیشرفته

به روش‌هایی اطلاق می‌شود که طی آن اکسیداسیون شیمیایی فاز آبی حاوی آلاینده آلی یا معدنی، در حضور رادیکال OH انجام می‌شود. این روش‌ها بر اساس تولید گونه‌های واکنش پذیر مانند رادیکال OH

استوار هستند. این گونه‌های رادیکالی بسیار ناپایدار بوده و دارای قدرت اکسیداسیون بالایی هستند در نتیجه به صورت غیر انتخابی^۱ به مولکول‌های آلی موجود در آب آلوده حمله کرده و مسیر اکسیداسیونی را ایجاد می‌کند که در نهایت منجر به معدنی‌سازی آلاینده آلی می‌شود [۱۵].

روش‌های اکسیداسیون پیشرفته مزایایی در مقایسه با سایر روش‌ها دارند که عبارت هستند از [۱۶، ۱۷]:

- ✓ مشکل ضایعات دور ریختنی را ندارد.
- ✓ اکثر ترکیبات آلی را به مواد معدنی مانند آب و دی‌اکسید کربن تبدیل می‌کند.
- ✓ قابلیت کاربرد برای طیف وسیعی از آلاینده‌ها را دارند.
- ✓ تخریب و تجزیه‌ی آلاینده در درجه حرارت محیط و فشار اتمسفر صورت می‌گیرد.

۲-۲-۴-۱ - انواع روش‌های اکسیداسیون پیشرفته

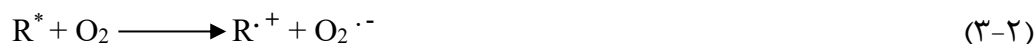
۱- فتولیز:

ساده‌ترین روش اکسیداسیون پیشرفته فتولیز است. در طی این فرایند آلاینده‌های رنگی به طور مستقیم نور UV را جذب کرده و برانگیخته می‌شوند، سپس این گونه‌های برانگیخته به طور همولیتیک^۲ شکسته شده و گونه‌های رادیکالی را ایجاد می‌کنند. این گونه‌های رادیکالی طی واکنش‌های زنجیری محصول نهایی واکنش (محصولات تخریب) را تولید می‌کنند. همچنین طی یک فرایند انتقال الکترون سوپر اکسید رادیکال ($O_2^{\cdot-}$) تولید می‌شود که نقش مهمی در تخریب ترکیبات آروماتیک حلقوی دارد [۱۸].



¹ Non-selective

² Homolytic



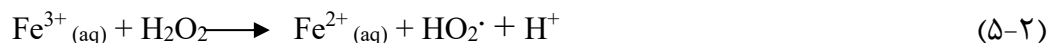
۲- فنتون و فتوفنتون:

واکنش فنتون و فتوفنتون جزء واکنش‌های همگن^۱ محسوب می‌شوند. فنتون شامل واکنش بین

Fe²⁺ و هیدروژن پراکسید (H₂O₂) در تاریکی طبق معادله‌های زیر می‌باشد:



دو مسیر واکنشی زیر برای Fe³⁺ تولید شده وجود دارد:



واکنشگرهای مورد استفاده در فرایند فتوفنتون شامل Fe³⁺ یا H₂O₂/UV/Fe²⁺ می‌باشد. طبق

معادله‌ی زیر واکنش Fe(OH)²⁺ با نور، باعث تولید مجدد یون Fe²⁺ و شروع دوباره چرخه‌ی تولید OH[·]

می‌شود. این امر باعث افزایش کارایی واکنش فتوفنتون نسبت به فنتون می‌شود.



رادیکال هیدروکسیل تولید شده طی دو فرایند فنتون و فتوفنتون به عنوان یک اکسنده‌ی قوی

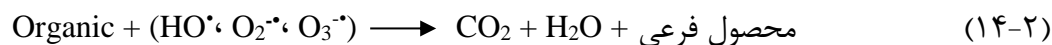
عمل می‌کند و در واکنش با آلاینده طبق معادله‌ی (۸-۲) آن را تخریب می‌کند [۱۹]:



¹ Homogeneous

۳- ازوناسیون و فتوازوناسیون:

در فرایند ازوناسیون با استفاده از ازون، رادیکال هیدروکسیل تولید می‌شود و این رادیکال به ترکیبات آلی حمله می‌کند. ازون یک اکسید کننده قوی است و قابلیت بسیار بالایی برای انجام واکنش‌های اکسیداسیون - احیا با بسیاری از ترکیبات آلی را دارد. حلالیت ازون در آب به عواملی مانند دمای آب و pH بستگی دارد. ازون طبق واکنش‌های زیر در محلول آبی، رادیکال هیدروکسیل تولید می‌کند [۱۳، ۲۰]:



با ترکیب ازون و انرژی UV تشکیل رادیکال هیدروکسیل تسریع می‌گردد:



۴- تخریب فتوکاتالیستی:

تجزیه فتوکاتالیستی به عنوان یک تکنیک برتر برای رنگبری و تخریب رنگ‌های آلی پیشنهاد شده است. این روش در مقایسه با دیگر تکنیک‌های اکسیداسیون پیشرفته مؤثرتر است چون نیم‌رسانایی که به عنوان کاتالیست استفاده می‌شود، ارزان است و اکثر آلاینده‌های آلی را به مواد معدنی تجزیه می‌کند. هدف

از انجام واکنش فتوکاتالیستی، معدنی‌سازی کامل و یا تبدیل آن‌ها به موادی با خطر کمتر می‌باشد که با روش‌های بیولوژیکی قابل تخریب نمی‌باشند.

یک فتوکاتالیست ایده آل بایستی دارای ویژگی‌های مطلوبی نظیر: پایداری نوری، سازگاری با محیط زیست از نظر شیمیایی و زیستی، انجام واکنش فتوکاتالیستی در دما و فشار محیط باشد، همچنین در دسترس و ارزان باشد [۲۱، ۱۲].

۲-۳ - فتوکاتالیست‌های نیمه هادی

برای رنگبری و تخریب مواد رنگی توسط تکنیک فتوکاتالیستی از نیمه رساناهای مختلفی مانند ZnO و TiO_2 , Fe_2O_3 , ZnS , CdS , Cu_2O به عنوان فتوکاتالیست استفاده می‌شود. ساختار الکترونی اتم‌های نیم رسانا در ترکیب‌های شیمیایی، مهم‌ترین دلیل استفاده از آن‌ها در فرایند فتوکاتالیستی است. زیرا این عناصر دارای باند ظرفیت پر شده و باند هدایت خالی از الکترون می‌باشند. باند ظرفیت (VB)^۱ و باند هدایت (CB)^۲ توسط گاف انرژی از هم جدا شده است. هنگامی که یک مولکول نیم رسانا فوتونی با انرژی برابر یا بیشتر از گاف انرژی دریافت کند، الکترون از لایه ظرفیت به لایه هدایت برانگیخته می‌شود و یک حفره در باند ظرفیت و یک الکترون در باند هدایت ایجاد می‌گردد که تشکیل زوج الکترون-حفره، لازمه‌ی شروع واکنش رنگبری و تخریب فتوکاتالیستی است [۲۲].

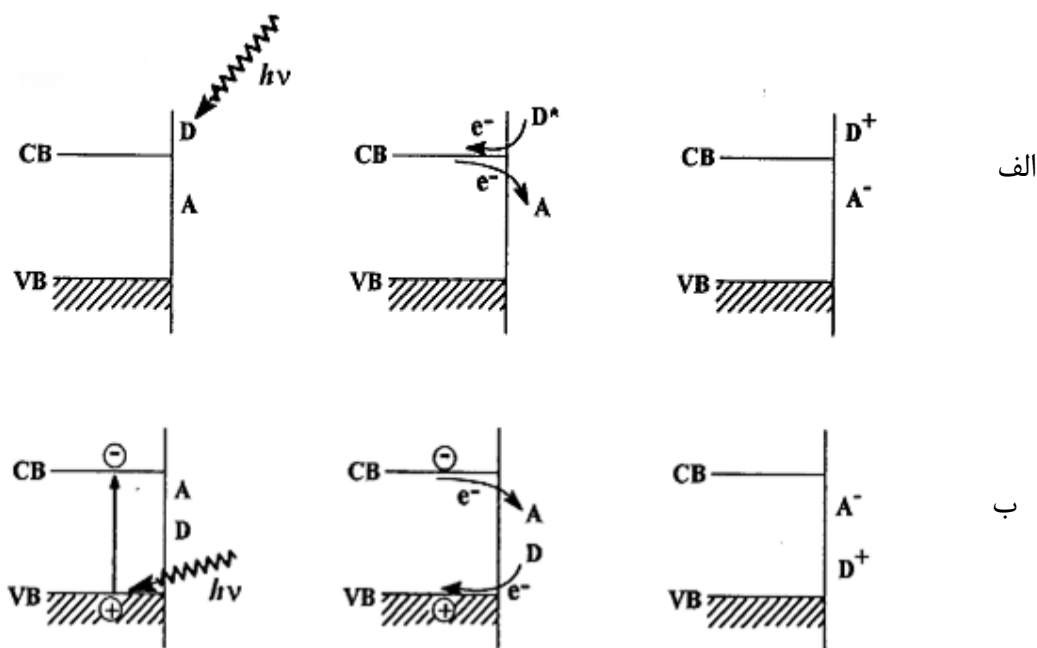
۲-۳-۱ - فرایند انتقال الکترونی بر روی سطح کاتالیست

در حالت برانگیخته الکترونی، مولکول یا نیمه‌هادی بسیار فعال می‌باشد. انتقال الکترون بین مولکول‌های واقع در سطح کاتالیست و یا بین سطح کاتالیست و یک مولکول جذب شده بر روی سطح آن

¹ Valence Band

² Conduction Band

انجام می‌شود. به صورت مشابه با فرایندهای فتوکاتالیستی، فرایندهای انتقال الکترون در نیمه‌هادی نیز می‌تواند به دو دسته الف و ب نشان داده شده در شکل (۱-۲) طبقه‌بندی شود.



شکل (۱-۲) دسته بندی واکنش‌های فوتونی کاتالیست [۲۳]

در این طبقه‌بندی در حالت الف نیمه‌هادی به عنوان واسطه عمل کرده و الکترون از دهنده به آن منتقل شده و سپس از آن جا به گیرنده منتقل می‌شود. این طرح فرایند انجام گرفته توسط ذرات نیمه هادی حساس شده با رنگدانه‌ها است که یک مورد از کاربرد آن‌ها در پیل‌های خورشیدی برای تولید الکتریسیته می‌باشد. در حالت ب برانگیخته شدن الکترون از نیمه‌هادی اتفاق می‌افتد و به نوار هدایت آن می‌رود و حفره‌ای در نوار ظرفیت نیمه‌هادی به جای می‌گذارد. الکترون می‌تواند از نوار هدایت فتوکاتالیست به اوربیتال خالی پذیرنده منتقل شده و به‌طور هم‌زمان یک الکترون از یک اوربیتال دهنده پر به نوار ظرفیت وارد شود که اغلب فرایندهای فتوکاتالیستی در نیمه‌هادی‌ها این مسیر را می‌پیمایند [۲۲, ۲۳].

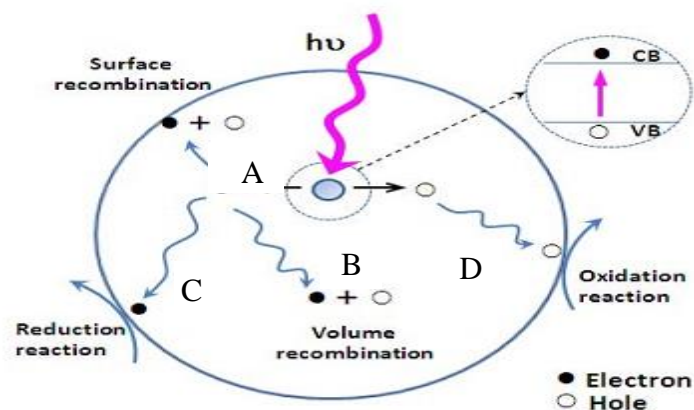
۲-۳-۲ - مکانیسم تخریب فتوکاتالیستی در حضور نور فرابنفش

اولین مرحله فرایند تخریب و رنگبری فتوکاتالیستی آلاینده به وسیله نیمه هادی‌ها، تشکیل جفت الکترون - حفره بر روی نیمه هادی می‌باشد. قسمت بزرگ شده در شکل (۲-۲) بر انگیخته شدن یک الکترون از لایه ظرفیت به لایه هدایت توسط نور با انرژی برابر یا بزرگ‌تر از گاف انرژی^۱ را نشان می‌دهد. بعد از مرحله برانگیخته شدن، الکترون و حفره‌های ایجاد شده می‌توانند چندین مسیر را طی کنند. شکل (۲-۲) مسیرهای مختلفی را که الکترون‌ها و حفره‌ها در آن بازیابی می‌شوند را نشان می‌دهد.

در اثر مهاجرت الکترون‌ها و حفره‌ها به سطح نیمه هادی، الکترون‌ها به وسیله ذرات مواد آلی یا معدنی جذب شده و یا به حلال منتقل می‌شوند، بدین طریق بر روی سطح نیمه هادی الکترون‌ها می‌توانند ذرات پذیرنده الکترون را احیا کنند (در محلول‌های هوا داده شده معمولاً اکسیژن گونه پذیرنده الکترون می‌باشد) (مسیر C)، در مقابل حفره می‌تواند به سطح مهاجرت کرده به طوری که یک الکترون از ذرات الکترون دهنده با حفره اکسید کننده روی سطح ترکیب می‌شود (مسیر D). بازیابی الکترون و حفره در رقابت با انتقال بار به ذرات جذب شده می‌باشد. بازیابی الکترون و حفره جدا شده می‌تواند در داخل ذرات نیمه هادی (مسیر B) و یا بر روی سطح نیمه هادی (مسیر A) همراه با آزاد شدن گرما انجام شود [۲۵].

[۲۲، ۲۴]

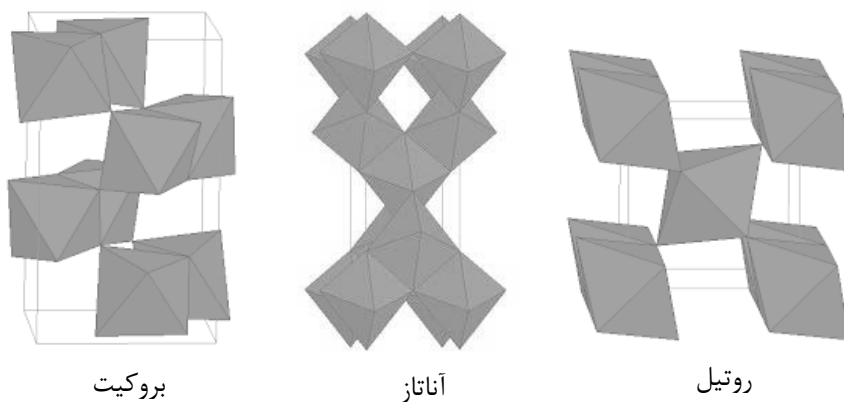
¹ Band Gap



شکل (۲-۲) مسیرهای مختلف بازیابی الکترون و حفره در یک نیمه هادی [۲۲, ۲۴, ۲۵, ۲۶]

۲-۴ - فتوکاتالیست TiO_2

TiO_2 در سه ساختار شبکه‌ای روتیل، آناتاز و بروکیت دیده می‌شود. مهم‌ترین عامل که بر عملکرد فتوکاتالیستی TiO_2 اثر می‌گذارد، شکل کریستال آن است. در مقایسه بین این سه حالت شبکه‌ای، آناتاز بیشترین سهم فعالیت فتوکاتالیستی را دارد و امروزه این ترکیب بیشترین کاربرد را در تولید محصولات صنعتی دارد.



شکل (۳-۲) ساختارهای بلوری TiO_2

با توجه به شکل (۲-۳) از نظر فضایی، فازهای روتیل و آناتاز تتراگونال هستند و بروکیت، اورتورومبیک هستند. آناتاز با داشتن گاف انرژی (۳/۲ الکترون ولت) نتایج بهتری از روتیل (۳ الکترون ولت) در تولید فوتوکاتالیست‌های پودری دارد. یکی از دلایل اصلی آن داشتن پتانسیل احیایی بالاتر برای تولید الکترون‌ها از ابتدای تشکیل تا انتهای نوار هدایت است. در نتیجه آناتاز توانایی بیشتری برای از بین بردن آلاینده‌ها و آلودگی‌ها خواهد داشت.

سرعت تولید الکترون-حفره در روتیل به دلیل پایین تر بودن گاف انرژی در مقایسه با آناتاز بیشتر است، اما فاز آناتاز در مقایسه با روتیل الکترون‌های بیشتری را به دام می‌اندازد، بنابراین ترکیب مجدد الکترون حفره در فاز آناتاز نسبت به روتیل کمتر صورت می‌گیرد که این امر منجر به افزایش تعداد حفرات در دسترس می‌شود [۲۷].

۲-۴-۱ - ویژگی‌های TiO_2

در سال‌های اخیر استفاده از انواع مختلفی از مواد نیمه رسانا مانند Cu_2O ، CdS ، ZnS ، Fe_2O_3 ، TiO_2 و ZnO در فرایند تخریب و رنگبری آلاینده‌ها مورد بررسی قرار گرفته است. در این میان نانو ذرات TiO_2 و ZnO به عنوان بهترین فوتوکاتالیست‌ها در تخریب آلاینده‌های زیست محیطی، به شمار می‌روند. بعضی از ویژگی‌های نانو ذرات که در فرایند فوتوکاتالیستی موثر هستند شامل ابعاد ذرات، ساختمان کریستال، سطح هیدروکسیل شونده، جذب سطحی آلودگی و روش آماده‌سازی می‌باشد [۲۸].

TiO₂ به دلیل ویژگی‌های زیر در مقایسه با سایر نانو ذرات بیشتر مورد استفاده قرار می‌گیرد [۲۹].

[۳۰].

- ✓ بازده فتوکاتالیستی بالا
- ✓ پایداری فیزیکی و شیمیایی
- ✓ ارزان و در دسترس بودن
- ✓ غیر سمی بودن

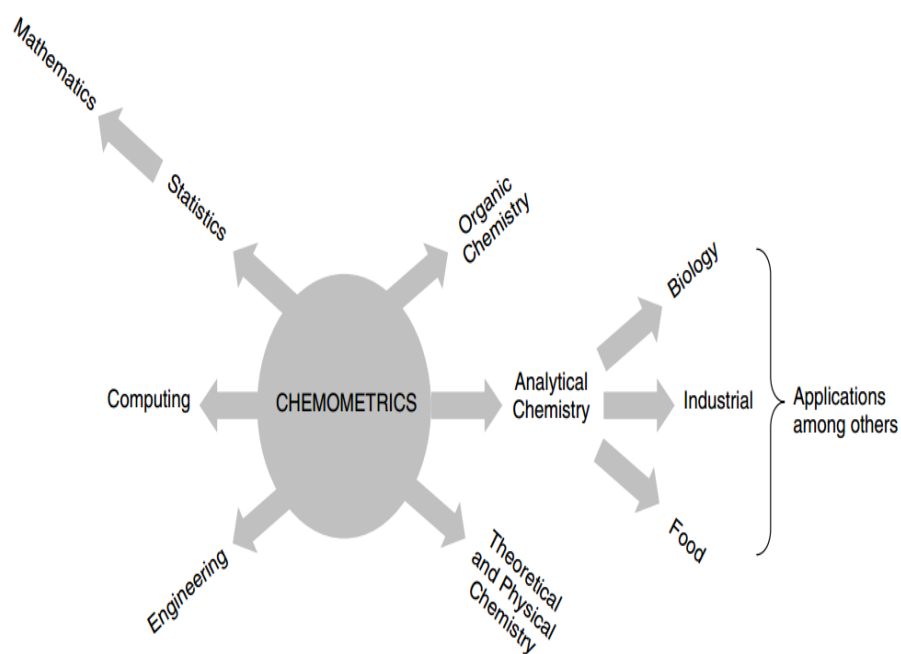
۲-۴-۲ - کاربردهای TiO₂

در سال‌های اخیر به علت ویژگی‌های جالب فتوکاتالیستی کریستال‌های TiO₂ تحقیقات زیادی گزارش شده است. استفاده تجاری از نانو TiO₂ به عنوان فتوکاتالیست در زمینه‌هایی چون تصفیه آب، تصفیه هوا، استرلیزه کردن/ضد عفونی کردن، پنجره‌ها و شیشه‌های خود تمیز شونده و سنسورهای گازی در دنیا فراگیر شده است [۳۱].

۲-۵ - کمومتریکس

هنر استخراج اطلاعات مربوط به شیمی از داده‌های تولید شده از آزمایشات شیمیایی کمومتریکس نام دارد. کمومتریکس یکی از شاخه‌های با اهمیت علمی است که کاربردهای زیادی در بسیاری از شاخه‌های علوم مانند آمار چند متغیره، مدل سازی ریاضی، علوم کامپیوتر و شیمی تجزیه داراست. این موضوع باعث شده است تا در بسیاری از منابع علمی از کمومتریکس به عنوان فرزند ریاضیات، کامپیوتر و شیمی نام برده شود [۳۲، ۳۳]. بنابراین گسترش سریع فناوری به خصوص در قسمت‌های مربوط به کاربرد کامپیوتر در

شیمی تجزیه موجب پیشرفت‌های سریعی در زمینه کمومتریکس در چند سال اخیر شده است. شکل (۲-۴) ارتباط کمومتریکس با سایر علوم را نشان می‌دهد [۳۴].



شکل (۲-۴) ارتباط کمومتریکس و علوم [۳۲]

اصطلاح کمومتریکس برای اولین بار توسط اسوانت ولد^۱ دانشمند جوان سوئدی که در زمینه شیمی فیزیک آلی فعالیت داشت، مطرح گردید. همکاری ولد با بروس آر. کووالسکی^۲ که در دانشگاه واشنگتن بر روی الگوشناسی در شیمی تجزیه مطالعه می‌کرد، منجر به تأسیس انجمن بین‌المللی کمومتریکس (ICS)^۳ در سال ۱۹۷۴ گردید [۳۳]. بنا به تعریف ICS، کمومتریکس عبارت است از کاربرد روش‌های ریاضی، آماری

¹ Svante wold

² Bruce R. Kowaski

³ International Chemometrics Society

برای برقراری ارتباط بین سنجش‌های انجام شده روی یک سیستم یا فرایند شیمیایی به منظور درک بهتر اطلاعات شیمیایی است.

برخی از کاربردهای کمومتریکس عبارتند از:

- (۱) بهینه‌سازی
- (۲) تخمین پارامترهای سینتیکی
- (۳) بهینه‌سازی پارامترهای کنترل کیفی
- (۴) بررسی رفتار کروماتوگرافی غیر خطی
- (۵) پردازش سیگنال
- (۶) طراحی دارو
- (۷) کالیبراسیون غیر خطی و مدل‌سازی [۳۲].

در حال حاضر با توجه به رشد سریع تجهیزات مورد استفاده در شیمی و حجم بسیار زیاد داده‌ها و اطلاعات به منظور جمع‌آوری، پردازش، تفسیر اطلاعات و استخراج نتایج مفید از آنها، نیاز به کامپیوتر امری اجتناب ناپذیر است. از طرف دیگر شیمی‌دانان، گاهی اوقات با موادی سر و کار دارند که بسیار گران، سمی و خطرناک بوده و یا در مواردی به راحتی قابل دسترس نیستند. در این موارد، می‌توان از روش‌های ریاضی و آمار به طور مؤثری جهت توصیف و توجیه نتایج آزمایش‌های مختلف استفاده نمود [۳۵] و می‌توان گفت که موارد فوق مهم‌ترین دلایل افزایش کاربرد روش‌های کمومتریکس نزد شیمی‌دانان است.

۲-۶ - ارتباط کمی ساختار - خاصیت (QSPR)

یکی از کاربردهای ویژه‌ی روش‌های کمومتریکس توسعه مطالعات ارتباط کمی ساختار-خاصیت و ارزیابی داده‌های تجزیه‌ای از طریق آن است. تغییرات کوچک در ساختار مولکول‌های مشابه می‌تواند باعث ایجاد خواص شیمیایی کاملاً متفاوتی شود. نظریه QSPR بر این اساس است که خاصیت یک ترکیب به ویژگی‌های ساختاری آن بستگی دارد [۳۶] و هدف QSPR یافتن رابطه‌ی هماهنگ میان خاصیت شیمیایی و ویژگی‌های مولکولی، به منظور کاربرد این قواعد برای ارزیابی خاصیت ترکیبات جدید می‌باشد.

مطالعات QSPR به طور گسترده در پیش‌بینی خواص فیزیکی و شیمیایی مختلفی مثل نقطه جوش و ذوب، فشاربخار، حلالیت و غیره استفاده شده‌اند. این مطالعات شامل مراحل زیر است:

۱- فراهم کردن سری داده‌ها

۲- رسم و بهینه سازی ساختار هندسی ترکیبات

۳- محاسبه توصیف‌کننده‌ها

۴- انتخاب بهترین توصیف‌کننده‌ها

۵- مدل سازی

۶- ارزیابی قدرت پیش‌بینی مدل [۳۷].

۲-۶-۱ - فراهم کردن سری داده‌ها

اولین مرحله مدل سازی جمع‌آوری و انتخاب یک سری مولکولی است که مقادیر تجربی خاصیت مورد نظر آن‌ها باید با صحت قابل قبول اندازه‌گیری شده باشد. فاکتور محدود کننده در توسعه مدل‌های

QSPR در دسترس بودن داده‌های تجربی با کیفیت بالا است. در واقع اعتبار و صحت هر مدل QSPR به صحت تعیین تجربی خاصیت مورد نظر بستگی دارد. داده‌های تجربی می‌توانند در آزمایشگاه تولید یا از مقادیر منتشر شده در مقالات استخراج شوند. لازم به ذکر است که خاصیت ترکیبات باید در شرایط آزمایشگاهی یکسان تعیین شده باشد.

۲-۶-۲ - بهینه‌سازی ساختار هندسی ترکیبات

ایجاد توصیف‌کننده‌های هندسی و هیبریدی بر مبنای ساختار و هندسه دقیق مولکولی استوار است و اگر ساختارها به صورت صورتبندی^۱ با حداقل انرژی نباشد، مقادیر غیر صحیحی برای این توصیف‌کننده‌ها ایجاد می‌شود [۳۸]. بنابراین باید ساختار هندسی ترکیب بهینه^۲ شود و مختصاتی از ساختار ترکیب به دست آید که در آن مختصات کارترین، انرژی مولکول حداقل باشد. این کار توسط شیمی محاسباتی^۳ انجام می‌شود. روش‌های مختلفی که در شیمی محاسباتی برای بهینه‌سازی ساختار مولکول مورد استفاده قرار می‌گیرند شامل [۳۹]:

۱- روش مکانیک مولکولی

۲- روش مکانیک کوانتومی

روش‌های مکانیک مولکولی بر مبنای روابط کلاسیک بنا شده است و به طور خیلی ساده ساختمان مولکولی به صورت گوی و فنر در نظر گرفته می‌شود. در این روش انرژی مولکول به صورت مجموعه‌ای از انرژی‌های کششی، خمشی، الکتروستاتیکی و ... بیان می‌شود. سپس با استفاده از برنامه‌های کامپیوتری مربوط به روش مکانیک مولکولی طول پیوندها، زوایای پیوندی و صورتبندی تغییر می‌کند تا ساختاری که

¹ Conformation

² Optimize

³ Computational Chemistry

عبارت انرژی مکانیک مولکولی را مینیمم کند، پیدا شود. این روش بهینه سازی بیشتر برای ماکرومولکول‌ها به کار می‌رود [۴۰] ولی در مکانیک کوانتومی از معادله شرودینگر^۱ برای محاسبه انرژی مولکول استفاده می‌شود و نسبت به روش مکانیک مولکولی دارای صحت بیشتری است زیرا اثرات الکترونی اعمال شده روی مولکول را نیز در محاسبات وارد می‌کند در حالی که مکانیک مولکولی الکترون‌ها را در نظر نمی‌گرفت. از طرف دیگر این محاسبات که پیچیده می‌باشند، وقت بیشتری را طلب می‌کند. از جمله روش‌های کوانتومی می‌توان به روش‌های آغازین^۲ و روش‌های نیمه تجربی^۳ اشاره کرد [۳۹].

در روش‌های آغازین از هیچ داده تجربی استفاده نشده است بلکه ثابت‌هایی مانند جرم الکترون، ثابت پلانک، ثابت سرعت نور و بار الکترون به کار گرفته می‌شوند و محاسبه انرژی مولکول از حل معادله شرودینگر با استفاده از یکسری تقریب به دست می‌آید. در این روش چون کل الکترون‌ها را در محاسبات وارد می‌کند، بیشتر منحصر به مولکول‌های کوچک است و برای مولکول‌های بزرگتر محاسبات پیچیده‌تر و نیازمند ابر رایانه است [۳۹, ۴۱].

روش‌های نیمه تجربی روش‌هایی هستند که در برنامه‌هایی نظیر MOPAC و Hyperchem وارد شده‌اند. در این روش‌ها برخلاف روش‌های آغازین، فقط الکترون‌های لایه ظرفیت را در محاسبات وارد می‌کنند. در نتیجه زمان محاسبات در این روش‌ها کوتاه‌تر از روش‌های آغازین است و می‌تواند برای مولکول‌های بزرگتر به کار رود. روش‌های نیمه تجربی مختلفی مانند چشم‌پوشی کامل از انتگرال‌های همپوشانی دیفرانسیلی (CNDO)^۴، چشم‌پوشی از برخی از انتگرال‌های همپوشانی دیفرانسیلی (INDO)^۵،

¹ Schrodinger Equation

² Ab Initio

³ Semi empirical

⁴ Complete Neglect of Differential Overlap

⁵ Intermediate Neglect of Differential Overlap

صرف نظر کردن از انتگرال همپوشانی دو اتمی اصلاح شده (MNDO)^۱ و روش آستین^۱ (AM1)^۲ وجود دارد [۴۲]. روش AM1 [۳۹] جزء دقیق ترین روش های نیمه تجربی است که در این پروژه از آن استفاده شده است.

۲-۶-۳ - محاسبه توصیف کننده ها

توصیف کننده های مولکولی، مقادیری عددی هستند که ساختار یا شکل مولکول را توصیف نموده و ویژگی های ساختاری و الکترونی مولکول را به طور کمی نشان می دهند. برخی از ویژگی های یک توصیف کننده مناسب عبارت است از [۴۳]:

۱- ساده بودن اصلاح شده صرف نظر کردن دو اتمی از انتگرال همپوشانی

۲- مستقل بودن

۳- توانایی تفسیر ساختار مولکول

۴- عدم هم بستگی با سایر توصیف کننده ها

۵- قابلیت تمایز بین ایزومرهای مختلف یک مولکول

۶- قابل کاربرد برای دامنه وسیعی از ساختارهای مولکولی

توصیف کننده های مولکولی به روش های مختلفی طبقه بندی می شوند. ساده ترین طبقه بندی بر اساس ماهیت آنها (تجربی یا تئوری) است. توصیف کننده های تجربی بر اساس اندازه گیری های تجربی به دست می آیند، در حالی که توصیف کننده های تئوری با استفاده از ساختار مولکول و بدون نیاز به داده های

¹ Modified neglected of diatomic overlap

² Austian Method 1

تجربی محاسبه می‌شوند که باعث صرفه‌جویی در وقت، هزینه، مواد و تجهیزات خواهند شد و برای هر مولکولی واقعی و یا فرضی که هنوز سنتز نشده است نیز در دسترس می‌باشد [۴۴, ۳۸].

۲-۶-۴ - انتخاب توصیف‌کننده‌های مناسب

در مطالعات QSPR انتخاب بهترین توصیف‌کننده‌ها (متغیرها) که بیشترین وابستگی با خاصیت مورد نظر را دارند از مهم‌ترین مراحل می‌باشد. وجود توصیف‌کننده‌های زیاد و هم‌چنین وجود هم‌بستگی میان آنها، محاسبات و مدل‌سازی را دشوار و پیچیده می‌کند. لذا باید توصیف‌کننده‌هایی انتخاب شوند که با خاصیت مورد نظر بیشترین وابستگی را داشته و تغییرات آن را به خوبی توصیف کنند. از روش‌های متفاوتی برای انتخاب موثرترین توصیف‌کننده‌ها استفاده می‌شود که می‌توان به روش انتخاب متغیر پیش‌رونده، انتخاب متغیر گام به گام^۱ و الگوریتم ژنتیک^۲ اشاره کرد.

۲-۶-۴-۱ - انتخاب متغیر به روش رگرسیون خطی

رگرسیون معادله‌ای برای ایجاد یک رابطه بین یک متغیر از یک طرف و یک یا چند متغیر از طرف دیگر می‌باشد. اگر تنها یک متغیر مستقل داشته باشد، رگرسیون را ساده و در غیر این صورت، رگرسیون را چندگانه می‌گویند [۴۵]. در بسیاری موارد نمی‌توان تغییرات یک متغیر را فقط به مقادیر یک متغیر دیگر مربوط کرد. به عبارت دیگر برای پیش‌بینی مقادیر یک متغیر، دانستن مقادیر دو یا چند متغیر دیگر لازم است.

اما زمانی که تعداد متغیرهای مستقل مورد مطالعه زیاد باشد لازم است که تکنیکی برای انتخاب متغیرهای مهم و اثرگذار مورد استفاده قرار گیرد تا بتوان متغیرهای معنی‌دار را از متغیرهای بی‌اهمیت تمیز

¹ Stepwise Regression

² Genetic Algorithm

داد. روش‌های رگرسیون متفاوتی برای انتخاب متغیرهای مستقل وجود دارد. از جمله مهم‌ترین آن‌ها می‌توان به انتخاب پیش‌رونده^۱، حذفی پس‌رونده^۲ و روش گام به گام اشاره نمود.

• روش انتخاب متغیر پیش‌رونده

در روش انتخاب متغیر پیش‌رونده متغیرهای مستقل بر اساس این که کدام متغیر بیشترین همبستگی را با متغیر وابسته دارد، انتخاب می‌شود. سپس متغیر مستقلی که آماره‌ی فیشر (F) بزرگتر و خطای استاندارد کمتری را برای مدل ایجاد می‌کند وارد معادله مدل می‌شود.

• روش انتخاب متغیر حذفی پس‌رونده

در روش حذفی پس‌رونده با ایجاد یک رگرسیون خطی ساده که شامل تمام متغیرهای مستقل است شروع شده و سپس سهم هر عبارت در مدل مشخص می‌شود. متغیری که کمترین تاثیر و پایین‌ترین سطح معناداری را دارد کنار گذاشته شده و پارامترهای معادله رگرسیون دوباره محاسبه می‌گردند. به همین ترتیب متغیرهای فاقد سطح معناداری قابل قبول ادامه پیدا می‌کند و این روند زمانی خاتمه می‌یابد که خروج یک متغیر باعث کاهش قدرت پیش‌بینی مدل نسبت به مدل قبلی شود.

• روش انتخاب متغیر گام به گام

روش رگرسیون گام به گام ترکیبی از روش‌های پیش‌رونده و پس‌رونده است. در این روش متغیرهای مستقل به ترتیب و مشابه با روش پیش‌رونده وارد معادله می‌شوند و با ورود هر متغیر جدید کلیه متغیرهای موجود در مدل بررسی شده و اگر هر کدام از آن‌ها سطح معناداری خود را از دست داده باشند قبل از ورود

¹ Forward

² Backward

متغیر جدید از مدل خارج می‌شود. این روند تا هنگامی که ورود متغیر بعدی تغییر شگرفی در سطح معناداری ایجاد نکند ادامه می‌یابد. روش به کار برده شده برای انتخاب توصیف‌کننده مناسب در این پایان‌نامه، بر اساس فراوانی توصیف‌کننده‌ها است که توسط نرم‌افزار SPSS انجام می‌شود.

۲-۶-۵ - ایجاد مدل‌های آماری

مدل، در واقع یک رابطه‌ی ریاضی است که بیان‌کننده‌ی رابطه بین متغیر وابسته و مستقل می‌باشد و به کمک آن می‌توان با داشتن مقادیر متغیرهای مستقل، متغیر وابسته را ارزیابی کرد. پس از انتخاب مناسب‌ترین توصیف‌کننده‌ها، با استفاده از روش‌های آماری مختلف به جستجوی مدل مناسبی پرداخته می‌شود که بتواند ارتباط بین توصیف‌کننده‌های انتخابی و پارامترهای مورد مدل‌سازی را به درستی بیان کند. برای مدل‌سازی از روش‌های گوناگون خطی و غیر خطی می‌توان استفاده کرد. در زیر به دسته بندی کلی این روش‌ها اشاره می‌شود:

روش‌های آماری چند متغیره (روش‌های خطی)

رگرسیون خطی چندگانه^۱ (MLR)

حداقل مربعات جزئی^۲ (PLS)

رگرسیون اجزای اصلی^۳ (PCR)

^۱ Multiple Linear Regression

^۲ Partial Least Square

^۳ Principle Component Regression

روش‌های غیر خطی

شبکه عصبی مصنوعی^۱ (ANN)

یادگیری بردار پشتیبان^۲ (SVM)

در این پایان نامه از روش غیرخطی شبکه عصبی مصنوعی برای مدل‌سازی استفاده شده است.

۲-۶-۶ - مقدمه ای بر شبکه عصبی

شبکه عصبی مصنوعی یک سامانه پردازشی داده‌ها است که از مغز انسان ایده گرفته و پردازش داده‌ها را به عهده پردازنده‌های کوچک و بسیار زیادی سپرده که به صورت شبکه‌ای به هم پیوسته و موازی با یکدیگر رفتار می‌کنند تا یک مسئله را حل نمایند. در این شبکه‌ها به کمک دانش برنامه‌نویسی، ساختار داده‌ای طراحی می‌شود که می‌تواند همانند نرون^۳ عمل کند. به این ساختار داده نرون گفته می‌شود. با ایجاد شبکه‌ای بین این نرون‌ها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می‌دهند. در این شبکه عصبی نرون‌ها دارای دو حالت فعال (روشن یا یک) و غیرفعال (خاموش یا صفر) می‌باشند و هر یال (سیناپس یا ارتباط بین گره‌ها) دارای یک وزن^۴ می‌باشد. یال‌های با وزن مثبت موجب تحریک یا فعال نمودن گره فعال بعدی می‌شوند و یال‌های با وزن منفی، گره متصل بعدی را غیرفعال یا مهار می‌کنند [۴۶].

¹ Artificial Neural Networks

² Supported Vector Machine

³ Neuron

⁴ Weight

۲-۶-۶-۱ - تاریخچه

پیشرفت عمده شبکه‌های عصبی از سال ۱۹۴۳ با فعالیت‌های وارن مک کلوت^۱ و والتر پیتز^۲ آغاز شد. آن‌ها نشان دادند که شبکه عصبی می‌تواند هر تابع حسابی یا منطقی را محاسبه کند [۴۷]. به دنبال آن دونالد هب^۳ مکانیسمی برای آموزش شبکه ارائه داد. اولین کاربرد عملی شبکه‌های مصنوعی در سال ۱۹۵۸ بود که در آن از شبکه‌ای که روزنبلات^۴ پیشنهاد داده بود، استفاده شد [۴۸]. روزنبلات یک شبکه چند لایه را ارائه کرد و توانایی آن را در انجام شناسایی الگو اثبات نمود. این موفقیت زود هنگام باعث پیدایش زمینه وسیعی در تحقیقات شبکه عصبی شد. تقریباً در همان زمان برنارد ویدرو^۵ الگوریتم آموزشی جدیدی ارائه داد که از لحاظ ساختار به شبکه روزنبلات شبیه بود. قاعده آموزش ویدروهاف امروزه نیز در شبکه‌های عصبی استفاده می‌شود. اما شبکه ویدروهاف و روزنبلات دارای محدودیت‌هایی بود که اولین بار توسط ماروین مینسکی^۶ مطرح شد. محققان تحت تاثیر نظریه مینسکی بر این عقیده بودند که تحقیق در مورد شبکه‌های عصبی پایان یافته است و از آنجایی که کامپیوترهای دیجیتال قوی در دسترس نبود، نزدیک به یک دهه تحقیق در مورد شبکه‌های عصبی معلق ماند. اما در سال ۱۹۸۶ با ظهور کامپیوترهایی با توان محاسباتی سریع، مجدد تحقیق در مورد شبکه‌های عصبی آغاز شد. در این سال الگوریتم آموزشی پس انتشار توسط جیمز مک کللاند^۷ ارائه شد [۴۹ و ۵۰] که نقطه عطفی در تحقیقات شبکه‌های عصبی به شمار می‌آید. از سال ۱۹۸۶ تاکنون محققان زیادی در مورد شبکه‌های عصبی تحقیق نموده‌اند که نتیجه آن صدها مقاله در این زمینه است.

¹ Warren McCulloch

² Walter Pitts

³ Donald Hebb

⁴ Rosenblatt

⁵ Bernard Widrow

⁶ Marvin Minsky

⁷ James McClelland

امروزه از شبکه‌های عصبی به عنوان یک ابزار کارآمد در زمینه‌های مختلف علمی از جمله صنایع الکترونیک، پزشکی، اکتشاف نفت و گاز، رباتیک، شیمی و داروسازی استفاده می‌شود [۵۰].

۲-۶-۲ - ساختمان و عملکرد نرون طبیعی

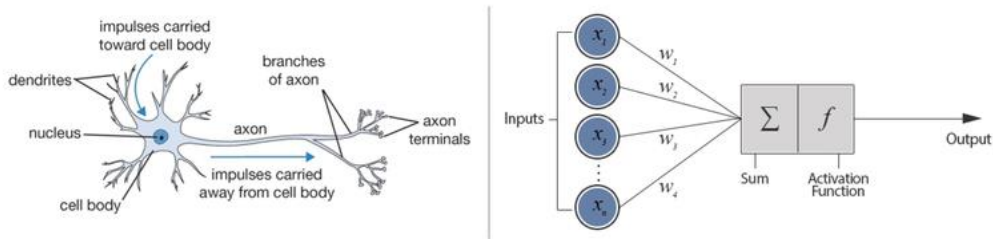
با بررسی‌های انجام شده روی مغز انسان مشخص شده است که مغز انسان از میلیاردها نرون عصبی منحصر به فرد تشکیل شده است. نرون اصلی‌ترین عنصر پردازش^۱ اطلاعات بشمار می‌آید که متشکل از دندریت^۲، بدنه سلول^۳ (شامل هسته و قسمت‌های حفاظتی دیگر) و اکسون^۴ می‌باشد. دندریت‌ها به‌عنوان مناطق دریافت سیگنال‌ها، شبکه‌هایی متشکل از فیبرهای سلولی هستند که دارای سطح نامنظم و شاخه‌های انشعابی بی‌شمار می‌باشند. به همین علت آن‌ها را شبکه‌های دریافتی "درخت گونه" می‌گویند. دندریت‌ها پس از دریافت اطلاعات به شکل سیگنال آن را به هسته‌ی سلول که در بدنه‌ی سلول قرار دارد، هدایت می‌کنند. بدنه‌ی سلول ضمن فراهم کردن انرژی لازم برای فعالیت نرون، بر روی سیگنال‌های دریافتی عملیات لازم را انجام داده که این عملیات شبیه یک عمل جمع ساده بر روی کل سیگنال‌ها و مقایسه با یک سطح آستانه می‌باشد. اکسون که برخلاف دندریت‌ها از سطحی هموارتر و تعداد شاخه‌های کمتر ولی طول بیشتر برخوردار می‌باشد، سیگنال‌های الکتروشیمیایی دریافتی از بدنه‌ی سلول را توسط سیناپس‌ها که ارتباط دهنده نرون‌ها هستند، به نرون‌های دیگر منتقل می‌کند و بدین ترتیب فعالیت‌های مغزی انجام می‌شود [۵۱]. شکل (۲-۵) ساختار یک نرون زیستی و ساختار شبکه عصبی را نشان می‌دهد.

¹ Processing Element

² Dendrite

³ Cell body(soma)

⁴ Axon



شکل (۵-۲) ساختار یک نرون زیستی و شبکه عصبی مصنوعی

۲-۶-۳ - ساختمان و عملکرد نرون مصنوعی

یک نرون مصنوعی در حقیقت مدل محاسباتی است که از نرون‌های عصبی واقعی انسان، الهام گرفته است. در شبکه‌های عصبی مصنوعی نیز بلوک‌های ساختاری و یا نرون‌ها، دستگاه‌های محاسباتی خیلی ساده‌ای هستند و ارتباط بین نرون‌ها عملکرد شبکه را تعیین می‌کند و هدف از آموزش شبکه‌های عصبی مصنوعی تعیین ارتباط مناسب، جهت حل مسائل مختلف است. در واقع شبکه‌های عصبی مصنوعی نیز با ایده گرفتن از رفتار شبکه عصبی بیولوژیکی شکل یافته است.

برای ایجاد یک شبکه عصبی مصنوعی باید سیستمی طراحی شود که دارای تعدادی ورودی باشد و با توجه به اهمیت هر یک، آن‌ها را با یکدیگر جمع جبری کند و توسط یک تابع موسوم به تابع انتقال^۱، آن‌ها را به نرون‌های دیگر ارسال نماید.

۲-۶-۴ - تابع انتقال

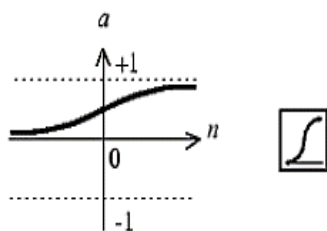
زمانی که ورودی‌ها با توجه به اهمیت آن‌ها با یکدیگر جمع جبری شوند توسط تابع انتقال به نرون بعدی منتقل می‌شوند. در واقع تابع انتقال یکی از اجزای شبکه عصبی می‌باشد که برحسب کاربرد می‌تواند

^۱ Transfer function

توابع متفاوتی باشد. دو تابع انتقال رایج در شبکه‌های عصبی که برای بهینه‌سازی شبکه به کار گرفته می‌شوند عبارتند از:

- تابع انتقال لگاریتم سیگموئید^۱ (log sig)

از این تابع انتقال در شبکه‌های پس انتشار استفاده می‌شود این تابع مقادیر ورودی را در محدوده منفی بی‌نهایت تا مثبت بی‌نهایت دریافت کرده و خروجی بین ۰ و ۱ تولید می‌نماید.



$$a = \text{logsig}(n)$$

Log-Sigmoid Transfer Function

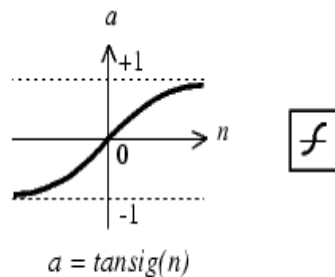
شکل (۶-۲) تابع انتقال لگاریتم سیگموئید

- تابع انتقال تانژانت سیگموئید^۲ (tansig)

این تابع انتقال مقادیر ورودی را در محدوده منفی بی‌نهایت تا مثبت بی‌نهایت دریافت کرده و خروجی بین ۱+ و ۱- تولید می‌کند.

¹ Logarithm sigmoid transfer function

² Hyperbolic tangent transfer function



$a = \text{tansig}(n)$
Tan-Sigmoid Transfer Function

شکل (۷-۲) تابع انتقال تانژانت سیگموئید

۲-۶-۵ - ساختارهای شبکه

بر اساس ساختار اتصال نرون‌ها، شبکه‌های عصبی مصنوعی را می‌توان در دو دسته‌ی کلی طبقه‌بندی کرد:

- شبکه‌های جلوسو^۱ که در آن‌ها گراف حلقه ندارد.
- شبکه‌های برگشتی^۲ که در گراف آن‌ها به دلیل وجود بازخورد^۳ در ساختار شبکه، حلقه به وجود می‌آید.

در معمول‌ترین خانواده‌ی شبکه‌های جلوسو که پرسپترون چندلایه^۴ نامیده می‌شوند، نرون‌ها در لایه‌هایی قرار می‌گیرند و اتصال بین آن‌ها یک طرفه است. اتصالات مختلف سبب رفتارهای متفاوت شبکه‌ها می‌شود. به‌طور کلی می‌توان گفت شبکه‌های جلوسو، استاتیک هستند، به این معنی که از ورودی داده شده تنها یک دسته مقدار خروجی تولید می‌کنند نه یک دنباله از مقادیر خروجی. شبکه‌های جلوسو بی‌حافظه هستند و پاسخ آن‌ها به یک ورودی مستقل از وضعیت قبلی شبکه است. از طرف دیگر شبکه‌های برگشتی،

¹ Feed-Forward Networks

² Recurrent Networks

³ Feedback

⁴ Multi Layer Perceptron

سیستم‌های دینامیک هستند و زمانی که یک دنباله‌ی ورودی جدید به آن‌ها داده شود، خروجی نرون‌ها محاسبه می‌شود. در اثر وجود مسیرهای بازخورد، ورودی هر نرون اصلاح می‌شود و این باعث می‌شود که شبکه وارد یک وضعیت جدید شود. ساختارهای مختلف شبکه، به الگوریتم یادگیری مناسب خود نیاز دارند.

۲-۶-۶- آموزش شبکه‌های جلو سو با تکنیک پس‌انتشار^۱

پس‌انتشار خطا یک روش متداول آموزش با ناظر برای شبکه‌های جلو سو است یعنی برای به دست آوردن ارتباط بین متغیرهای ورودی و خروجی در یادگیری به الگوی آموزشی نیاز است. به‌طور کلی آموزش به کمک تکنیک پس‌انتشار طبق مراحل زیر انجام می‌شود [۵۲].

۱- انتشار ورودی‌ها از نرون‌های ورودی به سمت نرون‌های خروجی

۲- اختصاص ماتریس وزن‌های تصادفی به هر یک از اتصالات

۳- مقایسه خروجی‌های شبکه با مقادیر واقعی (مقادیر هدف) و محاسبه‌ی خطای شبکه

۴- پس‌انتشار خطا از نرون‌های خروجی به سمت نرون‌های ورودی و اصلاح وزن‌ها

۵- ارزیابی عملکرد شبکه با توجه به تابع کارایی تعیین شده

مراحل فوق تا رسیدن به حداکثر تکرار^۲ مجاز تکرار می‌شود و یا این‌که مقدار تابع کارایی از مقداری

که تعیین شده کم‌تر باشد. شبکه عصبی انتخاب شده در این پروژه یک شبکه جلو سو با الگوریتم آموزشی پس‌انتشار می‌باشد.

¹ Back propagation

² Epoch

۲-۷ - ارزیابی مدل

۲-۷-۱ - استفاده از پارامترهای آماری

برای اطمینان از این که مدل نهایی، مدل مناسبی است که توانایی پیش‌بینی نمونه‌های مختلفی از یک جمعیت را داراست، باید مدل را ارزیابی کرد. این ارزیابی از طریق شاخص‌های کمی انجام می‌شود که به وسیله آن‌ها صحت نتایج ارائه شده توسط مدل مورد سنجش قرار می‌گیرند. فرض کنید X_1, X_2, \dots, X_n نشان‌دهنده متغیرهای مستقل (پیش‌گو) و Y_1, Y_2, \dots, Y_n نشان‌دهنده متغیرهای وابسته در مدل مورد بررسی می‌باشند. برخی از این شاخص‌های کمی عبارتند از:

ضریب همبستگی^۱: ساده‌ترین راه برای بررسی میزان همبستگی دو یا چند متغیر، محاسبه آماره‌ی ضریب همبستگی آن‌هاست. ضریب همبستگی دو متغیر X و Y با تعریف می‌شود. مقدار این آماره بین ۱ تا -۱ متغیر است. مقدار بزرگتر آن نشان‌دهنده‌ی این است که ارتباط خطی بیش‌تری میان متغیر وابسته و متغیرهای مستقل وجود دارد.

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad \text{رابطه (۲-۲۴)}$$

ضریب تعیین^۲: به‌عنوان یک شاخص برای بیان دقت خط رگرسیون برآورد شده، به‌کار می‌رود و نشان‌دهنده‌ی نسبت تغییرات متغیر وابسته توضیح داده شده توسط متغیر مستقل است. به‌عنوان مثال R^2 برابر با 0.6376 نشان می‌دهد که $63/76$ درصد تغییرات در متغیر وابسته می‌تواند توسط متغیر مستقل توضیح داده شود. رابطه ریاضی مربوط به ضریب تعیین به‌صورت زیر است:

¹ Correlation coefficient

² Determination coefficient

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad \text{رابطه (۲-۲۵)}$$

مجموع توان‌های دوم رگرسیون^۱ (SSR) طبق رابطه (۲-۲۶) بیانگر مجموع توان‌های دوم انحراف مقادیر پیش‌بینی شده‌ی متغیر وابسته از میانگین مقادیر آن است.

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{رابطه (۲-۲۶)}$$

مجموع توان‌های دوم کامل^۲ (SST) طبق رابطه (۲-۲۷) نشان‌گر مجموع توان‌های دوم انحراف مقادیر واقعی متغیر وابسته از میانگین مقادیر آن است.

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{رابطه (۲-۲۷)}$$

مجموع توان‌های دوم خطا^۳ (SSE) نیز مبین مجموع توان‌های دوم انحراف مقادیر واقعی متغیر وابسته از مقادیر پیش‌بینی شده برای آن است:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{رابطه (۲-۲۸)}$$

بنابراین با توجه به روابط فوق می‌توان نوشت:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{رابطه (۲-۲۹)}$$

طبق رابطه (۲-۲۹) اگر تمام مشاهدات بر روی خط برازش شده قرار گرفته باشند، یعنی به ازای تمام

نقاط $y_i = \hat{y}_i$ باشد، مقدار R^2 برابر یک می‌شود و هرگونه انحرافی از این حالت باعث می‌شود که مقدار R^2 از یک کوچکتر شود.

¹ Sum Square Regression

² Sum Square Total

³ Sum Square Error

ضریب تعیین تصحیح شده^۱: ضریب تعیین تصحیح شده یکی از پارامترهایی است که در ارزیابی و مقایسه مدل‌های مختلف با تعداد متفاوت متغیرهای پیش‌گو استفاده می‌شود و از آنجایی که R^2 (ضریب تعیین)، تعداد پارامترهای موجود در مدل را به حساب نمی‌آورد، به این منظور از ضریب تعیین تصحیح شده طبق رابطه‌ی زیر برای انتخاب مدل برتر استفاده می‌شود:

$$R_{adj}^2 = 1 - \frac{n-1}{n-p-1} \cdot \frac{SSE}{SST} = 1 - (1 - R^2) \frac{n-1}{n-p-1} \quad \text{رابطه (۲-۳۰)}$$

که در این رابطه p تعداد متغیرهای مستقل و n تعداد ترکیبات مورد بررسی می‌باشد.

آماره F: آزمون F یا آزمون فیشر در واقع آزمون معنی‌دار بودن آماری در تحلیل رگرسیون ساده و چند متغیره است و برابر با نسبت میانگین توان‌های دوم رگرسیون (MSR) به میانگین توان‌های دوم باقی‌مانده‌ها (MSE) است. بیان ریاضی آن به صورت زیر می‌باشد:

$$F = \frac{MSR}{MSE} = \frac{SSR/df_m}{SSE/df_{res}} \quad , \quad df_m = p \quad , \quad df_{res} = n - p - 1 \quad \text{رابطه (۲-۳۱)}$$

df_m درجه آزادی مدل، p تعداد متغیرهای مستقل مدل، n تعداد کل ترکیبات مربوط به مدل، df_{res} درجه آزادی باقی‌مانده‌ها است. و SSR ، SSE در روابط قبل توضیح داده شد.

مجموع توان دوم باقی‌مانده‌ها^۲ (PRESS): برابر مجموع توان دوم تفاوت بین مقدار کمیت مشاهده شده (y_i) و مقدار برآورد شده (\hat{y}_i) است.

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{رابطه (۲-۳۲)}$$

¹ Adjusted determination coefficient

² Predictive Residual Sum of Squares

خطای استاندارد پیش‌بینی^۱ (SEP)

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

رابطه (۲-۳۳)

خطای مطلق میانگین^۲ (MAE)

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}|}{n}$$

رابطه (۲-۳۴)

خطای نسبی پیش‌بینی^۳ (REP)

$$REP(\%) = \frac{100}{\bar{y}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n}}$$

رابطه (۲-۳۵)

میانگین توان‌های دوم خطا^۴ (MSE)

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

رابطه (۲-۳۶)

میانگین خطای نسبی^۵ (MRE)

$$MRE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right|}{n} \times 100$$

رابطه (۲-۳۷)

تولرانس^۶: همان‌طور که می‌دانیم وجود یا عدم وجود هم‌خطی^۷ در بین متغیرهای پیش‌بینی کننده مسئله

مهمی است که باید بررسی شود. پدیده هم‌خطی می‌تواند به یکی از دو دلیل زیر باشد:

¹ Standard Error of Prediction

² Mean Absolute Error

³ Relative Error of Prediction

⁴ Mean Square Error

⁵ Mean Relative Error

⁶ Tolerance

⁷ Multicollinearity

❖ حداقل دو تا از متغیرهای مستقل از لحاظ خطی خیلی به هم وابسته باشند. یعنی قدر مطلق ضریب همبستگی بین آن دو خیلی به یک نزدیک باشد.

❖ حداقل یکی از متغیرهای مستقل وابستگی شدیدی به مجموعه سایر متغیرهای مستقل داشته باشد که معمولاً این مورد مهم‌تر است.

لذا برای کشف پدیده هم‌خطی برای هر متغیر مستقل مقداری محاسبه می‌شود که به آن تلرانس گفته می‌شود. تلرانس برای هر متغیر X_i از رابطه زیر به دست می‌آید [۵۳]:

$$T_i = 1 - R_i^2 \quad i = 1.2.3 \dots p \quad \text{رابطه (۲-۳۸)}$$

که در آن p تعداد متغیرهای پیش‌گو و R_i^2 مجذور ضریب همبستگی چندگانه است که از رگرسیون X_i بر تمام متغیرهای پیش‌بینی کننده‌ی دیگر به دست می‌آید. پیشنهاد شده است که اگر برای هر یک از متغیرهای مستقل، T_i از 0.1 بیش‌تر باشد، پدیده هم‌خطی وجود ندارد.

عامل تورم واریانس^۱: در برخی مواقع از عامل افزایش واریانس به جای تلرانس استفاده می‌شود که رابطه آن با تلرانس به صورت زیر است:

$$VIF = \frac{1}{T_i} = \frac{1}{1-R_i^2} \quad i = 1.2.3 \dots p \quad \text{رابطه (۲-۳۹)}$$

واضح است که اگر X_i رابطه خطی با سایر متغیرها داشته باشد، آن‌گاه R_i^2 نزدیک به یک است و عامل افزایش واریانس بزرگ می‌شود. مقادیر VIF_i بین محدوده ۱ تا ۱۰ قابل قبول بوده و مقادیر بیشتر از ۱۰ اغلب به عنوان علامتی از این‌که داده‌ها مشکل هم‌خطی دارند، تلقی می‌شود. اگر هیچ‌گونه رابطه‌ی خطی بین متغیرهای پیش‌بینی کننده نباشد، آن‌گاه R_i^2 صفر و VIF_i برابر یک است. [۵۴, ۵۵].

¹ Variance Inflation Factor (VIF)

برای بررسی بیشتر مدل QSPR می‌توان از پارامترهای دیگری علاوه بر پارامترهای فوق استفاده نمود که در ذیل به آن‌ها اشاره شده است.

مقادیر ضرایب تعیین (R^2_0 و R'^2_0)

حداقل یکی از ضرایب تعیین که از عرض از مبدا صفر گذشته (پیش‌بینی برحسب واقعی و یا واقعی برحسب پیش‌بینی) یعنی R^2_0 یا R'^2_0 باید به R^2 نزدیک باشند. مقادیر به‌دست آمده با استفاده از رابطه (۲-۴۰) و رابطه (۲-۴۱) نیز باید از ۰/۱ کمتر باشد [۵۶].

$$R^2_0 = \frac{(R^2 - R'^2_0)}{R^2} \quad \text{رابطه (۲-۴۰)}$$

$$R'^2_0 = \frac{(R^2 - R^2_0)}{R^2} \quad \text{رابطه (۲-۴۱)}$$

شیب منحنی پیش‌بینی بر حسب مقدار واقعی و شیب منحنی مقدار واقعی بر حسب پیش‌بینی (k و k')

حداقل یکی از شیب‌های رگرسیون که از مبدا صفر گذشته است (مربوط به R^2_0 و R'^2_0) به یک نزدیک باشد و یا در این محدوده قرار بگیرد ($0/۱۵ \leq k \leq ۱/۱۵$ یا $0/۱۵ \leq k' \leq ۱/۱۵$). k و k' به ترتیب شیب منحنی پیش‌بینی بر حسب مقدار واقعی و شیب منحنی واقعی بر حسب مقدار پیش‌بینی می‌باشد که از عرض از مبدا صفر گذشته است [۵۶].

همچنین تفاوت بین مقادیر R^2_0 و R'^2_0 ($|R^2_0 - R'^2_0|$) بهتر است از مقدار ۰/۳ کمتر باشد [۵۷].

۲-۷-۲ - استفاده از نمودار برگشتی

در نمودار برگشتی مقادیر پیش‌بینی شده‌ی کمیت مورد نظر برحسب مقادیر تجربی رسم می‌شود و به کمک ضریب تعیین (R^2) به‌دست آمده از نمودار، پراکندگی نقاط اطراف خط برگشت تعیین می‌شود. محدوده‌ی تغییرات ضریب تعیین به‌دست آمده بین ۰ و ۱ است. هرچه قدر به ۱ نزدیک‌تر باشد، نتایج پیش‌بینی شده به مقادیر واقعی نزدیک‌ترند. اما اگر این مقدار برابر صفر باشد، بین داده‌ها هیچ گونه همبستگی وجود ندارد.

۲-۷-۳ - استفاده از نمودار خطای باقی‌مانده‌ها

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

۲-۷-۴ - استفاده از سری آزمون

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

فصل سوم

بخش تجربی

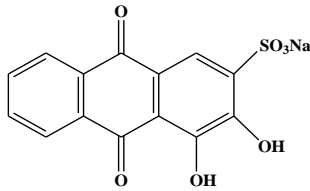
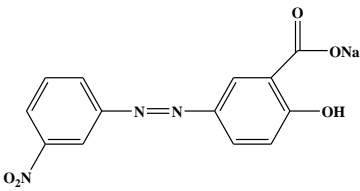
در این فصل، به منظور مطالعه ارتباط کمی ساختار-خاصیت رنگ‌های مختلف، از شبکه عصبی مصنوعی به‌عنوان روش مدل‌سازی برای پیش‌بینی درصد رنگبری استفاده شد و توانایی مدل به‌دست آمده مورد ارزیابی قرار گرفت. به‌طور کلی این بخش از تحقیق شامل به‌دست آوردن سری داده‌ها با رنگبری ۳۱ رنگ از گروه‌های مختلف، بهینه‌سازی ساختار مولکول‌ها و محاسبه توصیف‌کننده‌های مولکولی، انتخاب توصیف‌کننده‌های مولکولی مناسب و هم‌چنین مدل‌سازی و ارزیابی مدل است.

۱-۳ - مواد شیمیایی به کار رفته در پروژه

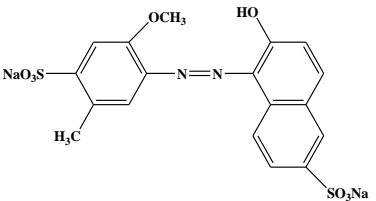
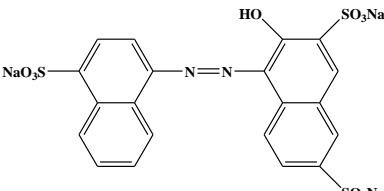
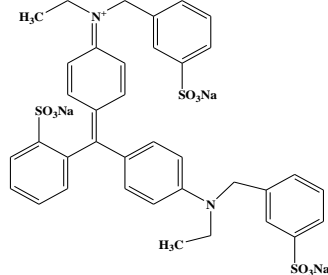
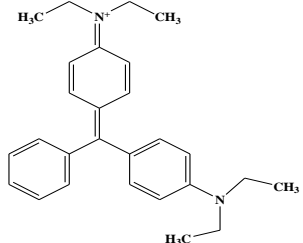
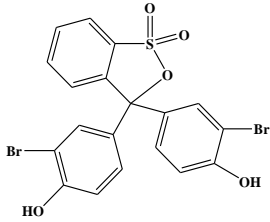
تمامی مواد شیمیایی استفاده شده در این کار با خلوص بالا از شرکت مرک آلمان خریداری شد.

لیست مواد شیمیایی مورد استفاده در جدول (۱-۳) آورده شده است.

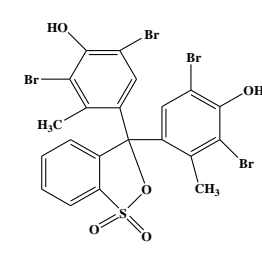
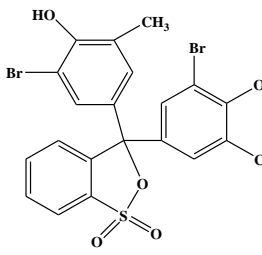
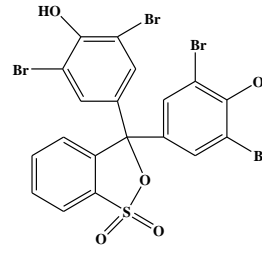
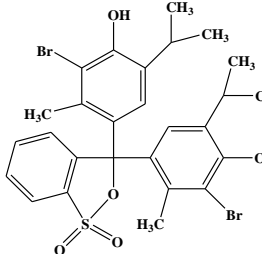
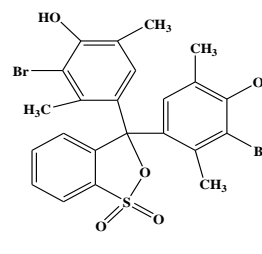
جدول (۱-۳) مواد شیمیایی مورد استفاده در این پروژه

شماره	گروه رنگ	نام ماده	ساختار ماده
۱	آنتراکینون	Alizarin red S	
۲	آزو	Alizarin yellow GG	

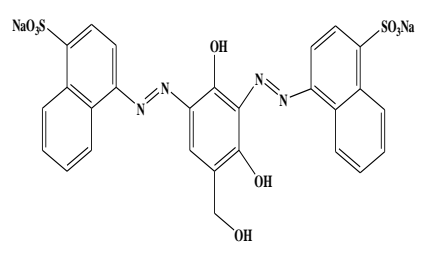
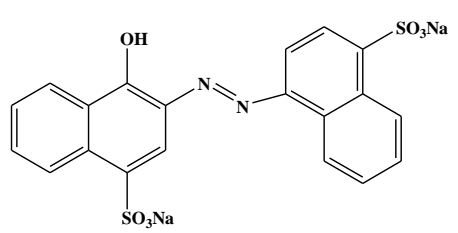
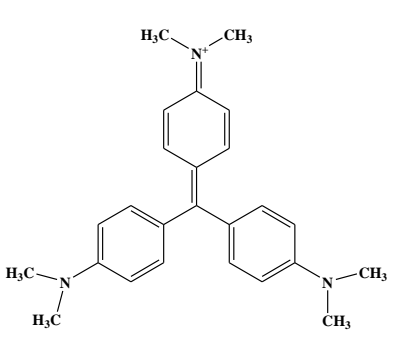
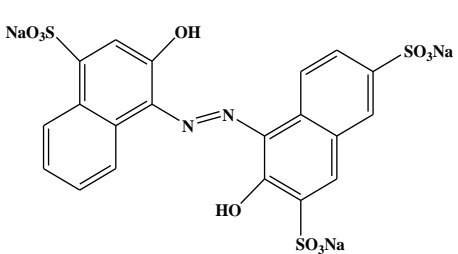
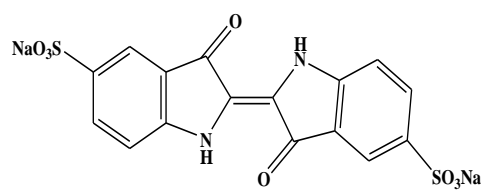
ادامه جدول (۳-۱) مواد شیمیایی مورد استفاده در این پروژه

شماره	گروه رنگ	نام ماده	ساختار ماده
۳	آزو	Allura Red	
۴	آزو	Amaranth	
۵	تری آریل متان	Brilliant blue e 313	
۶	تری آریل متان	Brilliant green	
۷	تری آریل متان	Bromophenol Red	

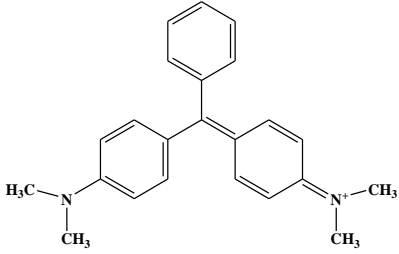
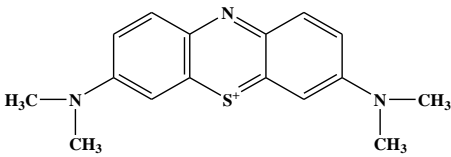
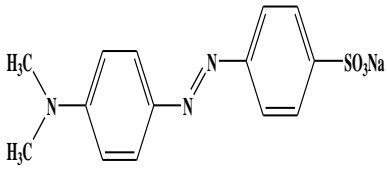
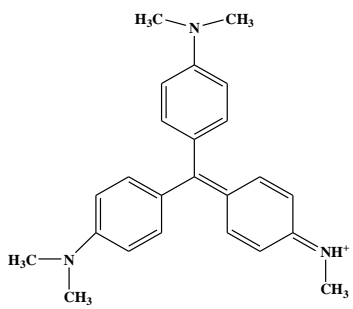
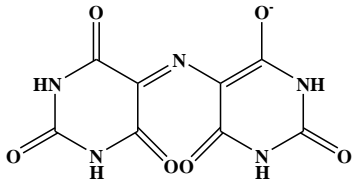
ادامه جدول (۳-۱) مواد شیمیایی مورد استفاده در این پروژه

شماره	گروه رنگ	نام ماده	ساختار ماده
۸	تری آریل متان	Bromocresol green	
۹	تری آریل متان	Bromocresol purple	
۱۰	تری آریل متان	Bromophenol blue	
۱۱	تری آریل متان	Bromothymol Blue	
۱۲	تری آریل متان	Bromoxyleneol Blue	

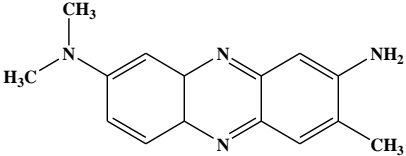
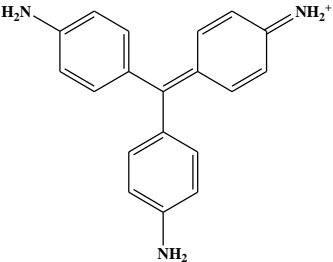
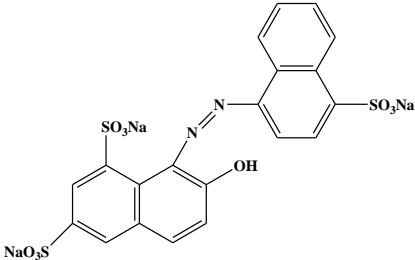
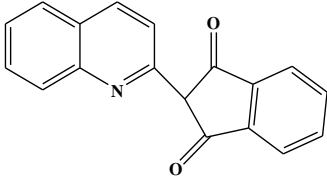
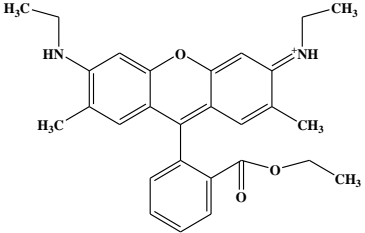
ادامه جدول (۳-۱) مواد شیمیایی مورد استفاده در این پروژه

شماره	گروه رنگ	نام ماده	ساختار ماده
۱۳	آزو	Brown HT	
۱۴	آزو	carmoisine	
۱۵	تری آریل متان	crystal violet	
۱۶	آزو	Hydroxynaphthol blue	
۱۷	ایندیگو	Indigo carmine	

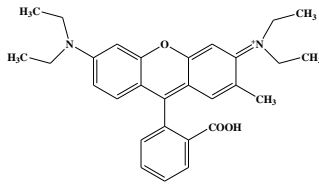
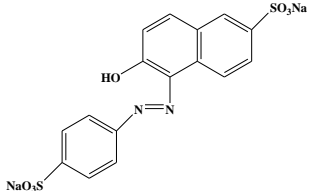
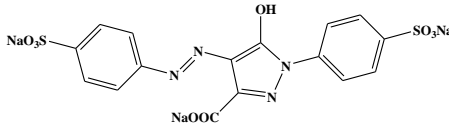
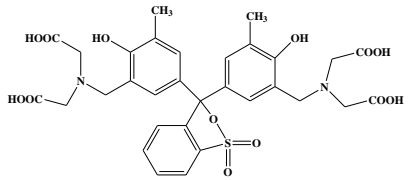
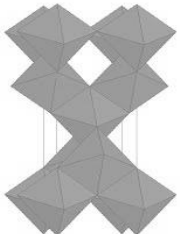
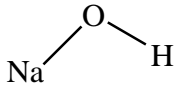
ادامه جدول (۳-۱) مواد شیمیایی مورد استفاده در این پروژه

شماره	گروه رنگ	نام ماده	ساختار ماده
۱۸	تری آریل متان	Malachite green	
۱۹	تiazین	Methylene Blue	
۲۰	آزو	Methyl orange	
۲۱	تری آریل متان	Methyl violet	
۲۲	باربیتورات	Murexide	

ادامه جدول (۳-۱) مواد شیمیایی مورد استفاده در این پروژه

شماره	گروه رنگ	نام ماده	ساختار ماده
۲۳	آزین	Neutral red	
۲۴	تری آریل متان	Pararosaniline	
۲۵	آزو	Ponceau 4R	
۲۶	کینولین	Quinoline Yellow	
۲۷	زانتن	Rhodamine 6G	

ادامه جدول (۳-۱) مواد شیمیایی مورد استفاده در این پروژه

شماره	گروه رنگ	نام ماده	ساختار ماده
۲۸	زانتن	Rhodamine B	
۲۹	آزو	Sunset Yellow	
۳۰	آزو	Tartrazine	
۳۱	تری آریل متان	Xylenol Orange	
۳۲	-	دی اکسید تیتانیوم	
۳۳	-	سدیم هیدروکسید	
۳۴	-	هیدروکلریک اسید	H-Cl

۳-۲ - تهیه محلول‌ها

جهت تهیه‌ی محلول‌های ۰/۰۱ مولار از رنگ‌ها، گرم نمک رنگ مورد نظر به یک بالن حجمی ۲۵/۰ میلی‌لیتری منتقل شد، و با اضافه کردن اندکی آب مقطر آنرا حل کرده و سپس بالن تا خط نشانه به حجم رسانده شد. محلول‌های مادر پس از تهیه در تاریکی نگهداری شدند. محلول‌های رنگ مورد نیاز با رقیق سازی از محلول مادر تهیه گردید. لازم به ذکر است که برای تمامی رنگ‌ها از آب مقطر به عنوان حلال استفاده شد. به منظور تبعیت از قانون بیر محلول‌های مورد استفاده رقیق شدند تا جذب آن‌ها به زیر ۲ برسد. برای ساخت محلول ۱/۰ مولار سدیم هیدروکسید، ۴/۰ گرم سدیم هیدروکسید در یک بالن حجمی ۱۰۰/۰ میلی‌لیتری حل و با آب مقطر به حجم رسانده شد. محلول‌های ۰/۱۰ مولار سدیم هیدروکسید با رقیق سازی از محلول مادر تهیه گردید.

جهت تهیه محلول ۱/۰ مولار هیدروکلریک اسید، ابتدا محلول مادر ۱/۰ مولار اسید تهیه گردید. برای ساخت محلول ۱/۰ مولار، ۸/۳ میلی‌لیتر از اسید غلیظ در یک بالن حجمی ۱۰۰/۰ میلی‌لیتری با آب مقطر به حجم رسانده شد و سپس با استفاده از محلول سدیم کربنات استاندارد گردید. محلول ۰/۱۰ مولار اسید، با رقیق سازی از محلول مادر تهیه گردید.

۳-۳ - دستگاهوری

- جهت ثبت طیف‌های ماوراء بنفش - مرئی از دستگاه اسپکتروفوتومتر تک‌پرتویی ریلی^۱ مدل UV-۱۸۰۰ با یک جفت سل کوارتز ۱/۰ سانتی‌متری استفاده شد.
- آزمایش‌های تخریب و رنگبری فتوکاتالیستی رنگدانه‌ها در یک فتوراکتور که در آزمایشگاه

^۱ Rayleigh

شیمی تجزیه طراحی و ساخته شد انجام گرفت.

➤ اندازه‌گیری‌های pH به وسیله‌ی دستگاه pH- متر متر اهم^۱ مدل ۷۴۴ مجهز به یک

الکتروود مرکب شیشه - کالومل (۳/۰۰ مولار) انجام گرفت.

➤ جهت توزین مواد از ترازوی سارتریوس^۲ مدل S 200 A با دقت ۰/۱ میلی‌گرم استفاده

شد.

➤ همزن مغناطیسی فرزانه آرمان مدل HPMA-700 برای همزدن مخلوط واکنش استفاده

شد.

➤ اندازه‌گیری زمان با یک زمان‌سنج صورت گرفت.

➤ جهت جداسازی نمونه‌ی رسوب از محلول آن، از دستگاه سانتریفیوژ بنیامین طب استفاده

شد.

۳-۴ - نرم‌افزارهای مورد استفاده

دانش کمومتری کس بسته‌های نرم‌افزاری متنوعی را برای انجام تمام مراحل مدل‌سازی به کار گرفته

است که در ادامه به اختصار بسته‌های نرم‌افزاری استفاده شده در این پژوهش معرفی می‌شوند.

۳-۴-۱ - بسته نرم‌افزاری Hyperchem 8.0.5

از بسته نرم‌افزاری Hyperchem [۳۹] برای رسم شکل مولکول‌ها و بهینه‌سازی ساختار با استفاده

از روش‌های کوانتومی و مکانیکی، استفاده شد.

¹ Metrohm

² Sartorius

۳-۴-۲ - بسته نرم‌افزاری Dragon 5.5

نرم‌افزار Dragon که توسط گروه تحقیقاتی کمومتریکس میلانو ارائه شده است [۵۸]، برای بررسی محاسبه ۳۲۲۴ توصیف‌کننده ساختاری مختلف مورد استفاده قرار گرفت. جهت محاسبه توصیف‌کننده به کمک این نرم‌افزار لازم است ساختار هندسی بهینه مولکول مورد استفاده قرار گیرد. برای این منظور می‌توان ساختار بهینه مولکول‌ها را به صورت فایل‌هایی با فرمت hin, mol و ... به‌عنوان اطلاعات ورودی به کار برد. که در این پروژه ساختار بهینه ترکیبات مورد مطالعه با پسوند hin به عنوان ورودی نرم‌افزار دراگون استفاده گردید.

۳-۴-۳ - بسته نرم‌افزاری SPSS 21

SPSS [۵۹] که نخستین نسخه آن در سال ۱۹۷۰ توسط جمعی از فارغ‌التحصیلان دانشگاه استانفورد آمریکا ارائه شده، برای بررسی تجزیه و تحلیل آماری داده‌ها و انجام رگرسیون تک متغیره به کار گرفته شد.

۳-۴-۴ - نرم‌افزار MATLAB

نرم‌افزار MATLAB [۶۰] یکی از جامع‌ترین و کارآمدترین نرم‌افزارهای علمی و محاسباتی است که طی چند سال گذشته تهیه و به بازار عرضه شد و در سال‌های اخیر با تدوین نسخه‌های جدیدتر و کامل‌تر روز به روز بر غنای آن افزوده می‌شود. MATLAB به معنای آزمایشگاه ماتریس است. از جمله کاربردهای جالب آن که مورد توجه شیمی‌دانان قرار گرفته است شبکه عصبی مصنوعی است که در این پژوهش از آن استفاده شده است. اطلاعات شیمیایی به عنوان ورودی و به صورت یک ماتریس به نرم‌افزار داده شد. در محیط این برنامه با استفاده از آرایه‌ها و فرامین موجود، امکان مدل‌سازی غیرخطی خاصیت با ساختار ترکیبات فراهم می‌گردد.

۳-۵ - روش کار در آزمایش رنگبری فتوکاتالیستی

برای رنگبری واکنش فتوکاتالیستی، تغییرات طیف جذبی مخلوط رنگ‌ها در زمان‌های مختلف مورد بررسی قرار گرفت. برای به دست آوردن طیف جذبی رنگ‌ها به صورت زیر عمل شد:

پس از تهیه محلول‌های مادر در بالن‌های ۲۵/۰ میلی‌لیتری و سپس رقیق‌سازی مرحله به مرحله، ابتدا ۵۰/۰ میلی‌لیتر محلول رنگ مورد نظر با غلظت مشخص که اسیدیته آن با افزایش سود یا هیدروکلریدریک اسید ۰/۱ مولار در $\text{pH}=7/0$ تنظیم شده بود به داخل یک بشر ۵۰/۰ میلی‌لیتری منتقل گردید. سپس مقدار ۰/۱۰۰ گرم از فتوکاتالیست TiO_2 به دقت توزین و به محلول افزوده شد. محلول برای انجام جذب سطحی، به مدت ۲۰ دقیقه به وسیله همزن مغناطیسی در تاریکی هم‌زده شد. پس از تعادل جذب سطحی، ۲/۰ میلی‌لیتر از محلول به داخل یک لوله آزمایش منتقل شده و به مدت ۲ دقیقه با سرعت ۴۰۰۰ دور بر دقیقه سانتریفیوژ گردید تا ذرات معلق فتوکاتالیست ته نشین شود. سپس محلول به منظور ثبت طیف جذبی، به سل دستگاه اسپکتروفوتومتر سر ریز شد و طیف جذبی محلول در محدوده طول موجی ۲۰۰-۷۰۰ نانومتر ثبت گردید. پس از ثبت طیف جذبی، محلول به مخلوط اصلی برگردانده شد و مخلوط واکنش در داخل فتوراکتور قرار گرفت و به منظور شروع واکنش رنگبری لامپ UV (۴۰۰ وات) روشن گردید. فاصله لامپ UV تا سطح محلول در ۸ سانتی متر تنظیم گردید. در فواصل زمانی ۵ دقیقه‌ای پس از روشن شدن لامپ UV و شروع واکنش فتوشیمیایی، حجم‌های ۲/۰ میلی‌لیتری از مخلوط واکنش به داخل لوله سانتریفیوژ منتقل شده و مطابق قبل، پس از جداسازی فتوکاتالیست، طیف جذبی محلول در ناحیه ۲۰۰-۷۰۰ نانومتر ثبت گردید. بعد از ثبت طیف‌های جذبی مقدار جذب در طول موج‌های ماکزیمم هر رنگ در زمان‌های مختلف اندازه‌گیری شد و از آن‌ها برای محاسبه پارامتر میزان رنگبری استفاده گردید. برای کلیه رنگ‌ها شرایط آزمایش مانند $\text{pH}=7/0$ ، حجم محلول (۵۰/۰ میلی‌لیتر)، مقدار کاتالیست

(۰/۰۱۰۰ گرم) و شدت تابش برای رنگبری ثابت بود. لازم به ذکر است که تخریب و رنگبری کلیه رنگ‌ها سه مرتبه تکرار شد. پارامتر کسر تبدیل (X) به عنوان معیاری از میزان رنگبری و تخریب رنگ در نظر گرفته شد. مقدار X از رابطه (۳-۱) محاسبه گردید. در این رابطه C_0 بیانگر غلظت اولیه رنگ و C غلظت رنگ در هر لحظه می‌باشد و X نیز کسر تبدیل در آن لحظه است.

$$\% X = \frac{C_0 - C}{C_0} \times 100 \quad (3-1)$$

جذب محلول‌ها در طول موج‌های ماکزیمم در زمان‌های مختلف اندازه‌گیری شد و با استفاده از منحنی‌های کالیبراسیون هر رنگ، مقدار C_0 و C محاسبه شد و مقدار کسر تبدیل X در سه زمان مختلف به دست آمد. نتایج میانگین حاصل از درصد رنگبری در جدول (۲-۳) نشان داده شده است.

جدول (۲-۳) میانگین درصد رنگبری در زمان‌های مختلف

شماره	نام رنگ	غلظت رنگ (مولار)	λ_{max}	زمان رنگبری	درصد رنگبری
۱	Alizarin red S	$5/0 \times 10^{-5}$	۵۱۷	a	۴۲
				b	۵۳
				c	۶۹
۲	Alizarin yellow GG	$5/0 \times 10^{-5}$	۳۵۳	a	۴۰
				b	۶۲
				c	۷۷
۳	Allura Red	$5/0 \times 10^{-5}$	۵۰۱	a	۵
				b	۲۳
				c	۷۰
۴	Amaranth	$1/0 \times 10^{-4}$	۵۲۷	a	۱۷
				b	۴۱
				c	۸۴

c: رنگبری در ۱۵ دقیقه

b: رنگبری در ۱۰ دقیقه

a: رنگبری در ۵ دقیقه

ادامه جدول (۲-۳) میانگین درصد رنگبری در زمان‌های مختلف

شماره	نام رنگ	غلظت رنگ (مولار)	λ_{max}	زمان رنگبری	درصد رنگبری
۵	Brilliant blue e 313	$1/0 \times 10^{-5}$	۶۳۰	a	۵
				b	۳۰
				c	۷۲
۶	Brilliant green	$5/0 \times 10^{-5}$	۶۲۵	a	۹
				b	۲۵
				c	۶۲
۷	Bromophenol Red *	$2/0 \times 10^{-5}$	۵۷۵	a	۱۵
				b	۶۰
				c	۸۹
۸	Bromocresol green	$1/0 \times 10^{-4}$	۶۶۶	a	۱۴
				b	۵۴
				c	۷۶
۹	Bromocresol purple	$3/0 \times 10^{-5}$	۵۸۹	a	۱۹
				b	۶۴
				c	۸۹
۱۰	Bromophenol Blue *	$2/0 \times 10^{-5}$	۵۹۱	a	۵
				b	۳۶
				c	۵۹
۱۱	Bromothymol Blue	$1/0 \times 10^{-4}$	۴۲۸	a	۹
				b	۲۹
				c	۸۳
۱۲	Bromoxyleneol Blue *	$1/0 \times 10^{-4}$	۶۱۴	a	۲۱
				b	۴۴
				c	۷۴

ادامه جدول (۳-۲) میانگین درصد رنگبری در زمان‌های مختلف

شماره	نام رنگ	غلظت رنگ (مولار)	λ_{max}	زمان رنگبری	درصد رنگبری
۱۳	Brown HT	$۱/۰ \times ۱۰^{-۴}$	۴۶۴	a	۴
				b	۱۶
				c	۴۰
۱۴	Carmoisine	$۱/۰ \times ۱۰^{-۴}$	۵۱۵	a	۳
				b	۱۳
				c	۳۶
۱۵	Crystal Violet	$۲/۰ \times ۱۰^{-۵}$	۵۸۶	a	۲۲
				b	۷۲
				c	۹۶
۱۶	Hydroxynaphthol Blue	$۱/۰ \times ۱۰^{-۴}$	۵۳۵	a	۱۵
				b	۵۹
				c	۹۵
۱۷	Indigo Carmine	$۵/۰ \times ۱۰^{-۵}$	۶۱۰	a	۳۷
				b	۹۲
				c	۹۹
۱۸	Malachite green	$۵/۰ \times ۱۰^{-۵}$	۶۱۷	a	۷
				b	۵۶
				c	۸۸
۱۹	Methylene Blue	$۲/۰ \times ۱۰^{-۵}$	۶۶۶	a	۱۹
				b	۶۷
				c	۹۳

ادامه جدول (۳-۲) میانگین درصد رنگبری در زمان‌های مختلف

شماره	نام رنگ	غلظت رنگ (مولار)	λ_{max}	زمان رنگبری	درصد رنگبری
۲۰	Methyl Orange	$5/0 \times 10^{-5}$	۴۶۴	a	۱
				b	۱۰
				c	۳۰
۲۱	Methyl Violet	$2/0 \times 10^{-5}$	۵۸۶	a	۱۳
				b	۵۵
				c	۹۲
۲۲	Murexide	$1/0 \times 10^{-4}$	۵۱۹	a	۳۱
				b	۷۵
				c	۹۵
۲۳	Neutral Red	$5/0 \times 10^{-5}$	۵۲۶	a	۴۹
				b	۸۳
				c	۸۸
۲۴	Pararosaniline	$2/0 \times 10^{-5}$	۵۴۳	a	۱۱
				b	۴۷
				c	۷۰
۲۵	Ponceau 4R	$5/0 \times 10^{-5}$	۵۰۸	a	۵
				b	۲۴
				c	۵۱
۲۶	Quinoline Yellow	$5/0 \times 10^{-5}$	۴۱۳	a	۴
				b	۱۴
				c	۴۴

ادامه جدول (۲-۳) میانگین درصد رنگبری در زمان‌های مختلف

شماره	نام رنگ	غلظت رنگ (مولار)	λ_{max}	زمان رنگبری	درصد رنگبری
۲۷	Rhodamine 6G	$1/0 \times 10^{-5}$	۵۲۷	a	۲۷
				b	۶۰
				c	۸۵
۲۸	Rhodamine B	$1/0 \times 10^{-5}$	۵۵۴	a	۱۸
				b	۵۰
				c	۸۷
۲۹	Sunset Yellow	$5/0 \times 10^{-5}$	۴۸۲	a	۳
				b	۱۳
				c	۳۷
۳۰	Tartrazine	$5/0 \times 10^{-5}$	۴۲۹	a	۵
				b	۲۰
				c	۴۷
۳۱	Xylenol Orange	$1/0 \times 10^{-4}$	۴۳۵	a	۱۴
				b	۳۱
				c	۵۸

* در محلول بازی حل شدند.

۳-۶ - محاسبات و مدل سازی QSPR

هدف از این مطالعه برقراری ارتباط بین ساختار رنگ‌های مورد استفاده و درصد رنگبری آنها است به منظور ارائه مدل QSPR جهت پیش بینی میزان رنگبری مراحل زیر انجام شد و در ادامه بحث به تفصیل به آنها پرداخته شده است.

۳-۶-۱ - سری داده‌ها

سری داده‌ها شامل درصد رنگبری ۳۱ رنگ از گروه‌های مختلف است که توسط تکنیک فتوکاتالیستی با استفاده از نیمه‌رسانای TiO_2 در زمان‌های ۵، ۱۰ و ۱۵ دقیقه اندازه‌گیری شد. نام و ساختار این ترکیبات به همراه درصد رنگبری در زمان‌های متفاوت در جدول (۳-۱) نشان داده شده است. کل ۹۳ داده به طور تصادفی به دو دسته‌ی آموزش (۸۰٪)، و آزمون (۲۰٪) تقسیم شد. داده‌های سری آموزش به منظور ساخت مدل و داده‌های سری آزمون برای ارزیابی و مقایسه‌ی توانایی مدل در پیش‌بینی رنگبری به کار برده شدند.

میانگین درصد رنگبری حاصل از سه مرتبه تکرار آزمایش به عنوان پاسخ (متغیر وابسته) در شبکه عصبی مصنوعی در نظر گرفته شد.

۳-۶-۲ - رسم و بهینه‌سازی ساختار مولکول‌ها

در این مرحله از مطالعه، ساختار مولکولی هر ترکیب ابتدا در نرم افزار Hyperchem 8.0 ترسیم شد. سپس با استفاده از روش AM1 بهینه شد و بهینه‌سازی تا زمانی ادامه یافت که جذر میانگین مربعات

گرایان^۱ انرژی به ۰/۰۰۱ کیلو کالری بر مول برسد. همه ترکیبات پس از بهینه‌سازی با پسوند hin ذخیره شدند.

۳-۶-۳ - محاسبه توصیف‌کننده‌ها

برای ایجاد ارتباط ریاضی بین ساختار- خاصیت، به مقادیر عددی نیاز است پس باید ساختار مولکول‌ها به مقادیر کمی تبدیل شود. بدین منظور ساختار بهینه شده ترکیبات به نرم‌افزار Dragon منتقل شد و برای هر یک از مولکول‌ها ۳۲۲۴ توصیف‌کننده در ۲۲ طبقه مختلف محاسبه گردید. جدول (۳-۳) فهرست این ۲۲ طبقه را نشان می‌دهد.

جدول (۳-۳) فهرست توصیف‌کننده‌های محاسبه شده توسط نرم‌افزار Dragon

ردیف	نام توصیف‌کننده	ردیف	نام توصیف‌کننده
۱	توصیف‌کننده‌های زیر ساختاری	۱۲	توصیف‌کننده‌های هندسی
۲	توصیف‌کننده‌های توپولوژیکی	۱۳	توصیف‌کننده‌های RDF
۳	شمارنده های گام مولکولی	۱۴	توصیف‌کننده‌های سه بعدی مورس
۴	شاخص‌های اتصال	۱۵	توصیف‌کننده‌های WHIM
۵	شاخص‌های اطلاعات	۱۶	توصیف‌کننده‌های GETAWAY
۶	خود ارتباطی های دوبعدی	۱۷	گروه های عاملی
۷	شاخص‌های مجاورت edge	۱۸	اجزای میان اتمی
۸	توصیف‌کننده BCUT	۱۹	توصیف‌کننده‌های بار
۹	شاخص بار توپولوژی Galvez	۲۰	خصوصیات مولکولی
۱۰	شاخص‌های مقادیر خاص	۲۱	اثر انگشت باینری دو بعدی
۱۱	پروفایل های مولکولی راندیک	۲۲	اثر انگشت فرکانس دو بعدی

¹ Root Mean Square Gradient

۳-۶-۴ - انتخاب بهترین توصیف‌کننده‌ها با توجه درصد مشارکت در مدل‌های

تصادفی حاصل از رد گروهی

ایراد وارد بر رگرسیون مرحله به مرحله این است که دارای مشکل عدم استواری می‌باشد. به عنوان مثال با تغییر کوچک در داده‌ها مدل خیلی متفاوتی را به وجود می‌آورد، که این امر درستی پیش‌بینی را کاهش می‌دهد. به منظور کاهش نوسان در انتخاب متغیر از روش فراوانی توصیف‌کننده‌ها استفاده شد تا توصیف‌کننده‌هایی با بیشترین فراوانی در مدل‌سازی مورد استفاده قرار گیرد. به منظور ساخت مدل‌های خطی به روش رگرسیون مرحله به مرحله، به صورت تصادفی ۲۰٪ از مولکول‌ها (شش مولکول) کنار گذاشته شد و با مولکول‌های باقیمانده، مدل خطی با استفاده از رگرسیون مرحله به مرحله به کمک نرم افزار SPSS ساخته شد. این کار ۲۰ مرتبه متوالی انجام شد به گونه‌ای که هر بار به صورت تصادفی شش مولکول (با جایگزینی) از سری داده‌ها خارج می‌شد. در نهایت ۲۰ مدل متفاوت برای سری داده‌ها ایجاد شد. توصیف‌کننده‌های مهم ۲۰ مدل به دست آمده در جدول (۳-۴) خلاصه شده‌اند.

جدول (۳-۴) توصیف‌کننده‌های به دست آمده از تقسیم‌بندی مختلف داده‌ها

شماره مدل	توصیف‌کننده‌ها
۱	MATS1m, E2s, BEHm4, C005, IC2, R1p
۲	MATS1m, E2s, C005, BIC2, MATS5e, Mor11p
۳	JGI7, B08NN, G2u, EEig10d, Mor15u, Du
۴	JGI7, nArNHR, MATS6e, MATS7m, R2v, HATS6e
۵	JGI7, B08NN, G2u, IDDE, RDF135e, nR06
۶	MATS1m, B03OS, IC1, C005, E2m, H8m
۷	JGI7, MATS7m, G2e, Mor08m, Mor26p, F10CO
۸	MATS1m, GATS2e, IC2, MATS4p, N078, Mor11p
۹	JGI7, F08OS, GATS6e, R8u, CIC1, E3m
۱۰	MATS1m, B05OO, EEig03d, nArCOOH, nNN, MATS4m
۱۱	MATS1m, N070, B04OS, RDF040v, ATS8p, Mor29e
۱۲	JGI7, B08NN, G2u, Hypertens80, Mor12e, De
۱۳	B09OO, MATS1m, B04OS, Mor31p
۱۴	nNN, B08OS, GATS6e, BIC2, Mor30m, Mor20e
۱۵	MATS1m, G2e, F04OS, GATS2v, N068, HATS4m
۱۶	JGI7, VRZ2, HATS7u, G3m, RBF, P2m
۱۷	JGI6, MATS7m, F10OS, G1p, JGI7, B10CN
۱۸	JGI7, MATS7m, G2u, B07CS, Mor29p, MATS4v
۱۹	MATS1m, nHBonds, GATS2m, R8u, HyDp, E2m
۲۰	DISPe, F03OS, Mor20m, GATS2e, HATS2m, Mor30m

با توجه به جدول (۳-۴) در این ۲۰ مدل، ۸ توصیف‌کننده بیش‌ترین درصد مشارکت (بیش‌ترین فراوانی) را داشته‌اند که نشان‌دهنده‌ی اهمیت این توصیف‌کننده‌ها است. این ۸ توصیف‌کننده به همراه طبقه مربوط به توصیف‌کننده در جدول (۳-۵) نشان داده شده است. برای بررسی میزان رنگیری در زمان‌های متفاوت، توصیف‌کننده زمان نیز به مجموع توصیف‌کننده‌های منتخب افزوده شد. جدول (۳-۶) ماتریس هم‌بستگی توصیف‌کننده‌ها را نشان می‌دهد که حاکی از عدم وجود هم‌بستگی معنادار بین توصیف‌کننده‌ها می‌باشد.

جدول (۳-۵) توصیف کننده‌های انتخاب شده بر اساس فراوانی توصیف کننده‌ها

No	Symbol	Class	Meaning	درصد مشارکت
۱	JGI7	topological charge indices	topological charge index of order 7	۴۵
۲	MATS1m	2D autocorrelation	Moaran autocorrelation –lag1	۴۵
۳	G2u	WHIM descriptors	2st component symmetry directional WHIM index	۲۰
۴	MATS7m	2D autocorrelation	Moaran autocorrelation –lag7	۲۰
۵	C-005	atom-centred fragments	CH3X	۱۵
۶	B08NN	2D binary fingerprints	Presence/absence of N_N at topological distance 08	۱۵
۷	E2s	WHIM descriptors	2nd component accessibility directional WHIM index	۱۰
۸	IC2	Information indices	information content index	۱۰

جدول (۳-۶) ماتریس همبستگی کل توصیف کننده‌های انتخاب شده بر اساس فراوانی توصیف کننده‌ها

	Time	JGI7	MATS1m	G2u	MATS7m	C-005	B08NN	E2s	IC2
Time	۱								
JGI7	۰	۱							
MATS1m	۰/۰۰۰	-۰/۶۳۳	۱						
G2u	۰/۰۰۰	-۰/۰۰۷	۰/۱۵۴	۱					
MATS7m	۰/۰۰۰	-۰/۷۵۹	۰/۴۳۹	۰/۰۱۳	۱				
C-005	۰/۰۰۰	۰/۲۷۸	-۰/۴۸۶	-۰/۱۰۰	-۰/۰۵۰	۱			
B08NN	۰/۰۰۰	-۰/۱۴۲	-۰/۴۱۲	-۰/۰۹۲	۰/۳۲۸	۰/۱۷۰	۱		
E2s	۰/۰۰۰	۰/۲۹۹	۰/۰۹۳	-۰/۱۶۸	-۰/۱۶۳	-۰/۴۸۸	-۰/۳۵۷	۱	
IC2	۰/۰۰۰	-۰/۴۱۵	۰/۵۰۰	-۰/۰۲۶	۰/۴۴۷	-۰/۵۹۹	۰/۱۴۵	۰/۳۲۹	۱

۳-۶-۵ - بررسی وجود هم‌خطی بین بهترین توصیف کننده‌ها

به منظور بررسی وجود پدیده‌ی هم‌خطی در بین توصیف کننده‌های انتخاب شده، عامل افزایش واریانس (VIF) مربوط به هر توصیف کننده محاسبه شد. برای محاسبه VIF مربوط به متغیر اول، ابتدا مقدار ضریب تعیین برای مدلی که متغیر اول به عنوان متغیر وابسته و ۸ متغیر دیگر به عنوان متغیرهای مستقل حضور دارند بدست آمد و مقدار VIF با استفاده از رابطه (۳۹-۲) محاسبه شد و برای ۸ متغیر دیگر نیز VIF به همین ترتیب محاسبه گردید که نتایج آن در جدول (۳-۷) نشان داده شده است. با توجه به اینکه برای کلیه متغیرهای مستقل، عامل افزایش واریانس از ۱۰ کمتر است، عدم وجود هم‌خطی معنادار در بین توصیف کننده‌ها اثبات می‌گردد.

جدول (۷-۳) مقادیر عامل افزایش واریانس توصیف کننده های انتخاب شده توسط روش فراوانی توصیف کننده ها

توصیف کننده	Time	JGI7	MATS1m	G2u	MATS7m	C-005	B08NN	E2s	IC2
VIF	۱/۰۰	۵/۲۶	۴/۲۸	۱/۱۵	۳/۸۷	۳/۰۰	۲/۶۱	۲/۴۷	۲/۶۷

۳-۶-۶ - مدل سازی شبکه عصبی مصنوعی با توصیف کننده های انتخاب شده توسط روش

فراوانی توصیف کننده ها

یکی از راه های یافتن رابطه غیرخطی بین متغیرهای مستقل و متغیر وابسته، استفاده از شبکه عصبی مصنوعی برای مدل سازی می باشد. شبکه عصبی انتخاب شده در این پروژه یک شبکه پیشخور با الگوریتم آموزشی پس انتشار می باشد که الگوریتم آن با برنامه متلب ۲۰۱۵ نوشته شد.

توصیف کننده زمان و توصیف کننده های انتخاب شده به روش فراوانی توصیف کننده ها، به عنوان ورودی به شبکه عصبی مصنوعی داده شدند. توصیف کننده های به دست آمده، بر اساس بیشترین فراوانی به ترتیب اهمیت وارد شبکه شده است که اهمیت نیز بر اساس درصد مشارکت بیشتر بود و اگر درصد مشارکت یکسان بود در این صورت متغیری بیشترین اهمیت را داشت که R^2 بزرگتری با پاسخ داشته است. خروجی شبکه نیز میزان رنگبری می باشد. کل ۹۳ داده به طور تصادفی به دو دسته ی آموزش (۸۰٪)، و آزمون (۲۰٪) تقسیم شد. داده های سری آموزش به منظور ساخت مدل و داده های سری آزمون برای ارزیابی و مقایسه ی توانایی مدل در پیش بینی رنگبری به کار برده شدند.

برای ایجاد یک مدل شبکه عصبی مصنوعی پیشگو با درصد خطای پیش بینی پایین، باید ابتدا

پارامترهای مؤثر بر شبکه بهینه شوند که این پارامترها عبارتند از:

✓ تعداد متغیرهای ورودی (تعداد توصیف‌کننده‌ها)

✓ تعداد گره‌های لایه پنهان

✓ نوع تابع انتقال

✓ تعداد دوره‌های آموزش

۳-۶-۱ - انتخاب تعداد لایه‌های پنهان

هر شبکه دارای حداقل یک لایه ورودی، یک لایه خروجی و تعدادی لایه پنهان می‌باشد. متأسفانه هیچ راهنمای مناسبی برای انتخاب تعداد لایه‌های پنهان وجود ندارد و ساختار شبکه‌های عصبی، اغلب به روش سعی و خطا ایجاد می‌شود. با این وجود در بیشتر موارد به نظر می‌رسد که یک لایه پنهان مناسب باشد [۶۲، ۶۱]. بنابراین در این تحقیق شبکه عصبی پیشخور با الگوریتم آموزشی پس‌انتشار با یک لایه پنهان مورد استفاده قرار گرفت.

۳-۶-۲ - انتخاب تعداد گره‌های لایه پنهان، تعداد متغیر ورودی، نوع تابع آموزش و

نوع تابع انتقال

پس از طراحی اولیه شبکه، برای اینکه شبکه قدرت پیشگویی مناسبی داشته باشد باید به‌طور مناسب آموزش دیده باشد و با توجه به اینکه فرایند آموزش وقت‌گیر است، انتخاب الگوریتم آموزشی سریعت‌ر امری مهم است. از میان الگوریتم‌های آموزشی متفاوت که در جعبه ابزار متلب وجود دارد، الگوریتم لونبرگ - مارکوارت و تنظیم بایزین دارای سرعت مناسب، قابلیت پیشگویی و تعمیم پذیری بهتری نسبت به سایر روش‌های معمول می‌باشند. شبکه عصبی مصنوعی دارای مشکلاتی نظیر بیش‌برازش^۱، آموزش بیش از حد^۲،

¹ overfitting

² overtraining

پیش بینی و ارزیابی نامناسب مدل و هزینه بر بودن می‌باشد. برای کاهش مشکلات نامبرده استفاده از شبکه عصبی مصنوعی با تابع آموزش تنظیم بایزین پیشنهاد شده است [۶۸-۶۳].

این تابع آموزش بسیاری از مشکلاتی موجود در مدل سازی، چون انتخاب مدل، استحکام مدل، انتخاب سری ارزیابی و بهینه‌سازی ساختار شبکه را رفع کرده است [۶۹]. همچنین از دیگر مزایای این تابع می‌توان به موارد زیر اشاره نمود:

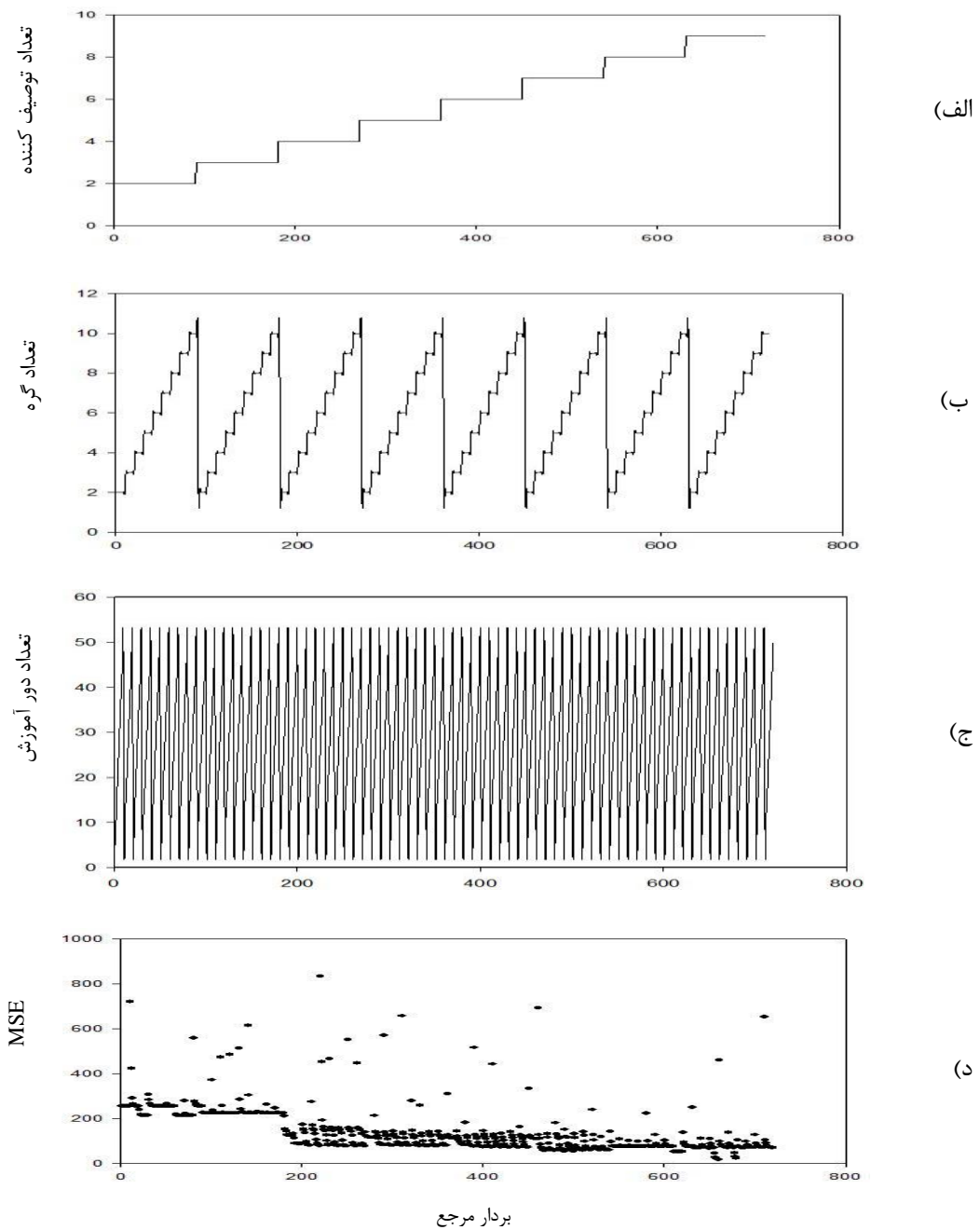
✓ به سختی دچار آموزش بیش از حد می‌شوند و اغلب به سری ارزیابی نیازی ندارند.

✓ به سختی دچار بیش برآش می‌شوند.

✓ یک مدل مطلوب بهینه را ارائه می‌کند به طوری که در داده‌های کم حتی به داده‌های سری

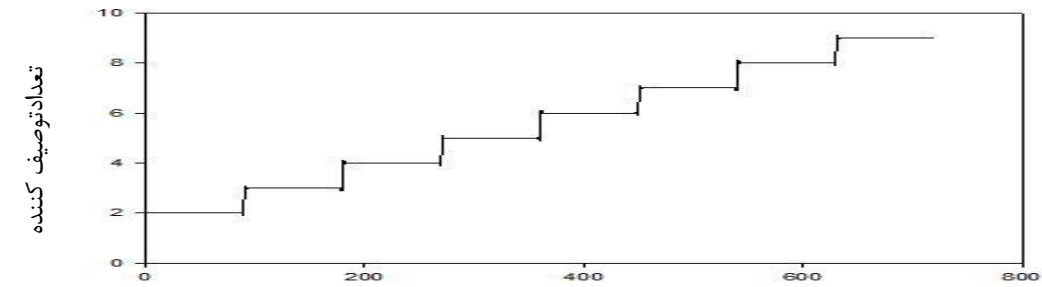
آزمون نیز نیازی نیست و این امر به طور ریاضی اثبات شده است [۷۰].

بر این اساس شبکه‌های عصبی توسط الگوریتم آموزشی تنظیم بایزین، و دو تابع انتقال لگاریتم سیگموئیدی (logsig) و تانژانت سیگموئیدی (tansig) ایجاد شد. ورودی‌های هر دو شبکه عصبی مصنوعی از ۲ تا ۹ و تعداد گره از ۲ تا ۱۰ و تعداد دور آموزش از ۵ تا ۵۰ با گام ۵ تنظیم شد و به طور همزمان آموزش داده شد. در روند بهینه‌سازی فوق، به حداقل رساندن MSE برای سری آموزش به عنوان معیار انتخاب شد، این کار موجب می‌شود بتوان نقش تمام پارامترها را به طور همزمان و یکجا بررسی کرد. بخشی از روند تغییرات پارامترهای شبکه در حین بهینه‌سازی همزمان پارامترها به همراه مقادیر MSE به دست آمده، به صورت نموداری بر حسب یک بردار مرجع فرضی در شکل (۳-۱) و شکل (۳-۲) آمده است. نتایج بهینه‌سازی این پارامترها در جدول (۳-۸) گردآوری شده است.

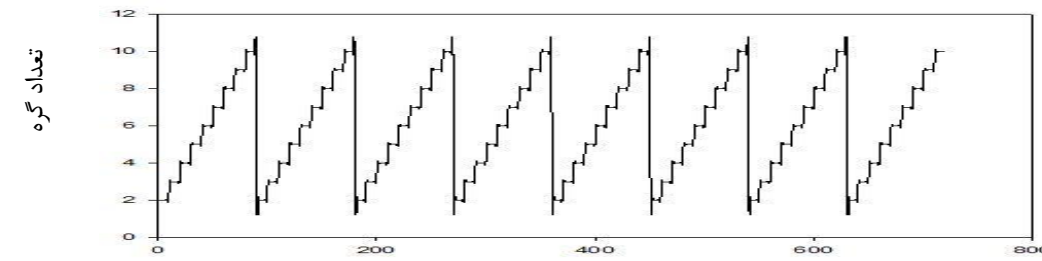


شکل (۳-۱) نمودارهای الف) تعداد توصیف کننده، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE

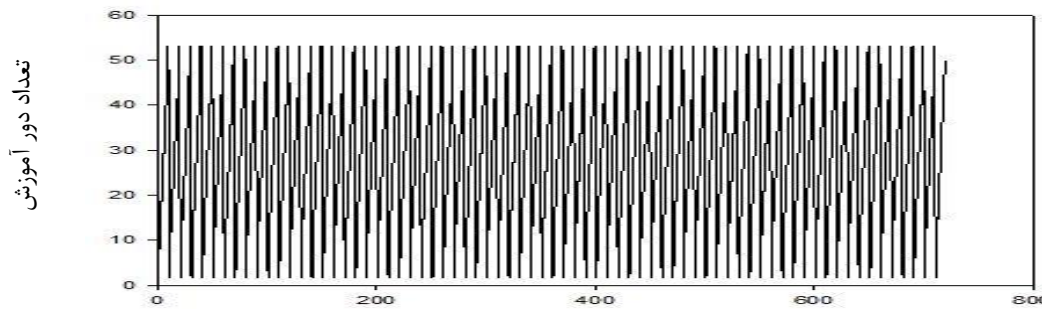
برای تابع آموزش تنظیم بایزین به همراه تابع انتقال لگاریتم سیگموئید



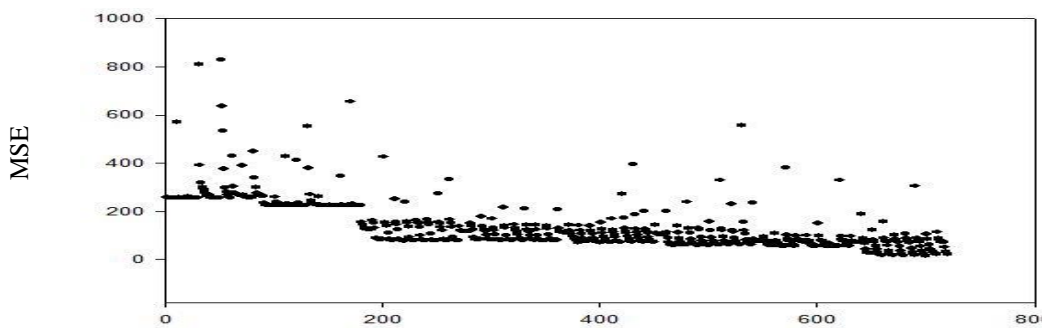
(الف)



(ب)



(ج)



(د)

نمودار مرجع

شکل (۲-۳) نمودارهای الف) تعداد توصیف کننده، ب) تعداد گره‌ها، ج) تعداد دورهای آموزش و د) مقادیر MSE

برای تابع آموزش تنظیم بایزین به همراه تابع انتقال تانژانت سیگموئید

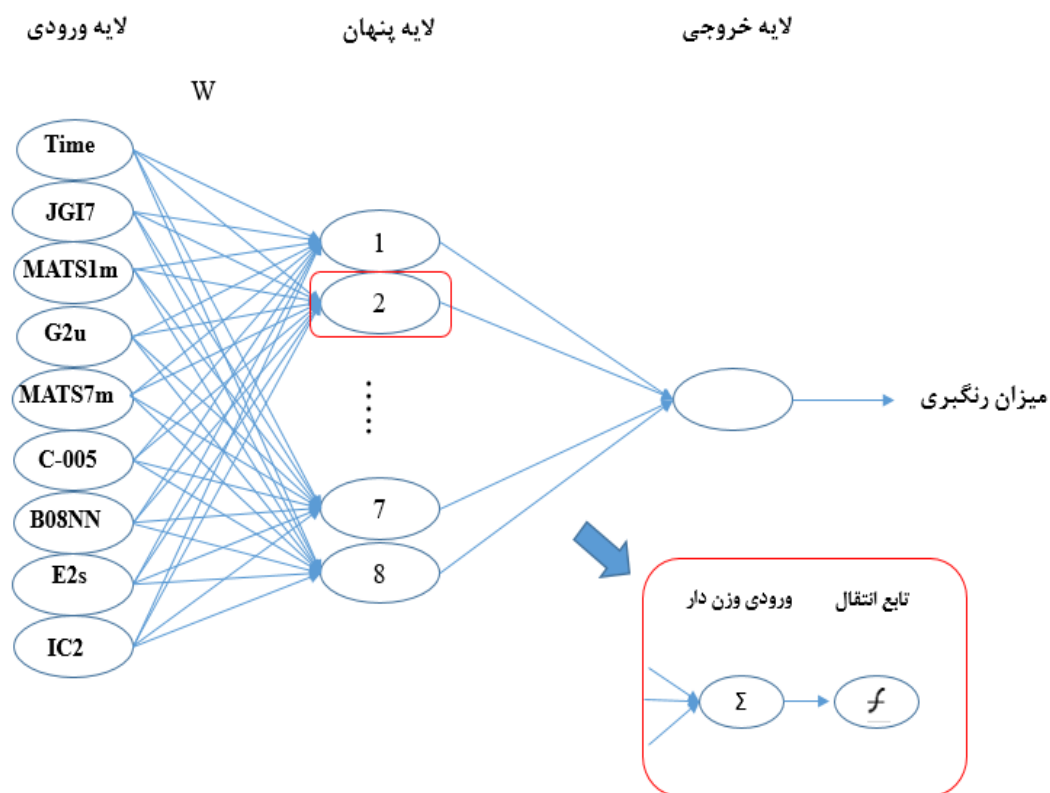
جدول (۳-۸) تابع و پارامترهای شبکه بهینه توصیف‌کننده‌های انتخاب شده بر اساس فراوانی توصیف‌کننده‌ها

تعداد توصیف‌کننده	تابع آموزش	تابع انتقال	تعداد گره	تعداد دور آموزش	MSE	MAE
۹	تنظیم بایزین	لگاریتم-سیگموئید	۴	۵۰	۱۸/۱	۳/۲۸
۹	تنظیم بایزین	تانژانت-سیگموئید	۸	۵۰	۱۵/۶	۳/۰۸

با توجه به نتایج به دست آمده، الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی برای سری داده‌های دارای میانگین مربع خطا (MSE) کمتری است.

۳-۶-۳ - ساختار شبکه عصبی مصنوعی بهینه شده

با توجه به نتایج به دست آمده در بهینه‌سازی شبکه، شبکه عصبی سه لایه پس انتشار با الگوریتم آموزشی تنظیم بایزین و تابع انتقال تانژانت سیگموئیدی، به عنوان شبکه بهینه به دست آمد که ساختار آن در شکل (۳-۳) نشان داده شده است.



شکل (۳-۳) ساختار شبکه عصبی بهینه شده

توابع و مقادیر پارامترهای بهینه شده شبکه عصبی برای ترکیبات مورد بررسی در جدول (۳-۹)

ارائه شده است.

جدول (۳-۹) توابع و مقادیر پارامترهای بهینه شده شبکه عصبی

train br	تابع آموزش
tansig	تابع انتقال لایه پنهان
۸	تعداد نرون لایه پنهان
۹	تعداد متغیرهای ورودی
۵۰	تعداد دورهای آموزش

۳-۷ - ارزیابی مدل

با توجه به روش مدل سازی مورد استفاده، ارزیابی مدل با استفاده از داده های سری آزمون و به وسیله رسم نمودار تغییرات مقادیر پیش بینی شده در مقابل مقادیر تجربی و با استفاده از کل داده ها به وسیله رسم نمودار مقادیر باقی مانده بر حسب مقادیر تجربی مورد بررسی قرار گرفت.

۳-۷-۱ - ارزیابی مدل با استفاده از داده های سری آزمون

اعتبار و اهمیت مدل با پیش بینی رنگبری، رنگ هایی که در مدل سازی حضور نداشته اند مشخص می شود. بدین منظور مدل انتخاب شده جهت پیش بینی درصد رنگبری ۱۸ ترکیب که در فرایند مدل سازی استفاده نشده اند، به کار گرفته شد و با استفاده از شبکه عصبی بهینه شده، مقادیر رنگبری این ترکیبات پیش بینی شد. جدول (۳-۱۰) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده های سری آزمون را نشان می دهد. شکل (۳-۴) نمودار تغییرات مقادیر پیش بینی شده در مقابل مقادیر تجربی را برای داده های سری آزمون نشان می دهد. با توجه به تکرارهای انجام شده در طی اندازه گیری رنگبری محدوده اطمینان برای هر رنگ محاسبه شد که در جدول (۳-۱۰) نشان داده شده است.

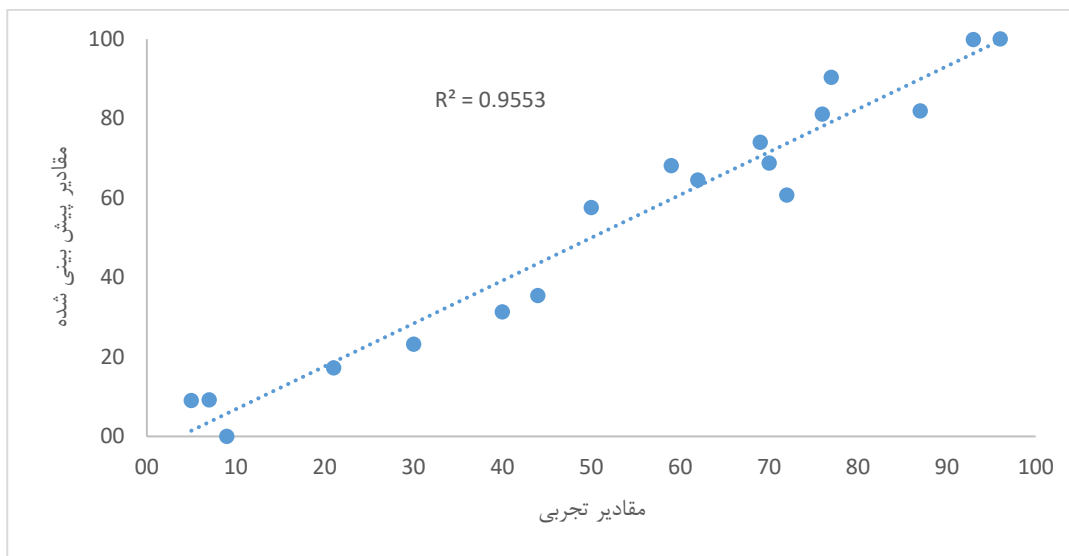
جدول (۳-۱۰) نتایج حاصل از ارزیابی مدل شبکه عصبی مصنوعی با استفاده از داده‌های سری آزمون

شماره ترکیب	میزان رنگبری تجربی	میزان رنگبری پیش‌بینی شده	محدوده اطمینان		پذیرش فرضیه صفر
			پایین	بالا	
۱c	۶۹	۷۴/۰	۶۲/۵	۷۵/۵	
۲c	۷۷	۹۰/۴	۶۵/۰	۸۹/۰	x
۵c	۷۲	۶۰/۷	۵۹/۸	۸۴/۲	
۶c	۶۲	۶۴/۵	۵۱/۷	۷۲/۳	
۸c	۷۶	۸۱/۱	۷۱/۶	۸۰/۴	x
۱۰c	۵۹	۶۸/۱	۴۸/۳	۶۹/۷	
۱۱a	۹	۰/۰	۰/۰	۱۹/۵	
۱۲a	۲۱	۱۷/۲	۶/۵	۳۵/۵	
۱۳c	۴۰	۳۱/۳	۲۹/۰	۵۱/۰	
۱۵c	۹۶	۱۰۰/۰	۸۷/۷	۱۰۰/۰	
۱۸a	۷	۹/۲	۱/۲	۱۲/۸	
۱۹c	۹۳	۹۹/۹	۸۴/۴	۱۰۰/۰	
۲۰c	۳۰	۲۳/۲	۲۲/۵	۳۷/۵	
۲۴c	۷۰	۶۸/۸	۶۰/۶	۷۹/۴	
۲۵a	۵	۹/۰	۰/۰	۱۱/۳	
۲۶c	۴۴	۳۵/۴	۳۳/۸	۵۴/۲	
۲۸b	۵۰	۵۷/۶	۴۱/۱	۵۸/۹	
۲۸c	۸۷	۸۱/۹	۷۷/۱	۹۶/۹	

a: رنگبری در ۵ دقیقه

b: رنگبری در ۱۰ دقیقه

c: رنگبری در ۱۵ دقیقه



شکل (۳-۴) نمودار مقادیر پیش‌بینی شده بر حسب مقادیر تجربی برای سری آزمون

۳-۷-۲ - ارزیابی مدل غیر خطی به روش رد مرحله‌ای تک‌تک

برای ارزیابی بیشتر مدل ارائه شده، تکنیک رد مرحله‌ای تک‌تک (برای کل داده‌ها) استفاده گردید. در این روش هر بار یک ترکیب از سری داده‌ها کنار گذاشته شد و در شرایط بهینه به‌دست آمده با ترکیبات باقی‌مانده مدل‌سازی صورت گرفت، سپس مدل به‌دست آمده برای پیش‌بینی خاصیت ترکیب کنار گذاشته شده به کار گرفته شد و این فرآیند برای تمام ترکیبات سری داده‌ها تکرار گردید.

نتایج حاصل از تکنیک رد مرحله‌ای تک‌تک در جدول (۳-۱۱) ارائه شده است. ضریب تعیین مشاهده شده بین مقادیر پیش‌بینی شده و مقادیر تجربی در شکل (۳-۵) نشان‌دهنده میزان نزدیکی رنگبری پیش‌بینی شده توسط مدل نسبت به مقدار تجربی آن است.

جدول (۱۱-۳) نتایج حاصل از ارزیابی مدل با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها

شماره ترکیب	میزان رنگبری تجربی	میزان رنگبری پیش‌بینی شده	محدوده اطمینان		پذیرش فرضیه صفر
۱a	۴۲	۳۰/۷	۵۲/۰	۳۲/۰	x
۱b	۵۳	۴۰/۱	۶۰/۲	۴۵/۸	x
۱c	۶۹	۷۵/۴	۷۵/۵	۶۲/۵	
۲a	۴۰	۳۳/۹	۵۲/۱	۲۷/۹	
۲b	۶۲	۵۰/۱	۶۸/۳	۵۵/۷	x
۲c	۷۷	۸۶/۳	۸۹/۰	۶۵/۰	
۳a	۵	۰/۰	۱۳/۷	۰/۰	
۳b	۲۳	۳۱/۴	۳۴/۱	۱۱/۹	
۳c	۷۰	۶۵/۳	۷۸/۶	۶۱/۴	
۴a	۱۷	۹/۹	۲۷/۳	۶/۷	
۴b	۴۱	۴۳/۵	۵۱/۶	۳۰/۴	
۴c	۸۴	۸۵/۷	۹۱/۸	۷۶/۲	
۵a	۵	۱۰/۶	۱۲/۲	۰/۰	
۵b	۳۰	۳۵/۷	۳۹/۲	۲۰/۸	
۵c	۷۲	۶۲/۳	۸۴/۲	۵۹/۸	
۶a	۹	۸/۰	۱۹/۸	۰/۰	
۶b	۳۵	۳۹/۷	۴۳/۴	۲۶/۶	
۶c	۶۲	۶۵/۷	۷۲/۳	۵۱/۷	
۷a	۱۵	۲۳/۷	۲۵/۲	۴/۸	
۷b	۶۰	۴۷/۳	۶۸/۶	۵۱/۴	x
۷c	۸۹	۸۸/۸	۹۴/۷	۸۳/۳	
۸a	۱۴	۱۷/۸	۲۱/۶	۶/۴	
۸b	۵۴	۴۷/۲	۶۳/۱	۴۴/۹	
۸c	۷۶	۸۰/۱	۸۰/۴	۷۱/۶	
۹a	۱۹	۱۰/۷	۳۳/۱	۴/۹	
۹b	۶۴	۴۶/۰	۷۲/۰	۵۶/۰	x
۹c	۸۹	۷۴/۰	۹۵/۶	۸۲/۴	x
۱۰a	۵	۹/۵	۱۰/۲	۰/۰	
۱۰b	۳۶	۴۲/۳	۴۷/۳	۲۴/۷	
۱۰c	۵۹	۶۸/۹	۶۹/۷	۴۸/۳	
۱۱a	۹	۰/۰	۱۹/۵	۰/۰	
۱۱b	۲۹	۴۱/۰	۴۱/۴	۱۶/۶	
۱۱c	۸۳	۸۳/۵	۹۱/۶	۷۴/۴	
۱۲a	۲۱	۱۶/۲	۳۵/۵	۶/۵	
۱۲b	۴۴	۴۸/۴	۵۲/۴	۳۵/۶	

ادامه جدول (۱۱-۳) نتایج حاصل از ارزیابی مدل با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها

شماره ترکیب	میزان رنگبری تجربی	میزان رنگبری پیش‌بینی شده	محدوده اطمینان		پذیرش فرضیه صفر
۱۲c	۷۴	۷۸/۶	۸۵/۰	۶۳/۰	
۱۳a	۴	۰/۶	۸/۳	۰/۰	
۱۳b	۱۶	۱۱/۰	۲۲/۳	۹/۷	
۱۳c	۴۰	۳۳/۵	۴۸/۳	۳۱/۷	
۱۴a	۳	۶/۵	۷/۵	۰/۰	
۱۴b	۱۳	۱۹/۵	۲۰/۹	۵/۱	
۱۴c	۳۶	۴۰/۹	۴۱/۰	۳۱/۰	
۱۵a	۲۲	۲۱/۲	۲۸/۶	۱۵/۴	
۱۵b	۷۲	۶۳/۶	۸۱/۴	۶۲/۶	
۱۵c	۹۶	۱۰۰/۰	۱۰۰/۰	۸۷/۷	
۱۶a	۱۵	۲۳/۰	۲۵/۱	۴/۹	
۱۶b	۵۹	۵۲/۳	۶۷/۱	۵۰/۹	
۱۶c	۹۵	۸۹/۹	۱۰۰/۰	۸۹/۲	
۱۷a	۳۷	۶۲/۳	۴۶/۹	۲۷/۱	x
۱۷b	۹۲	۶۱/۸	۹۹/۴	۸۴/۶	x
۱۷c	۹۹	۱۰۰/۰	۱۰۰/۰	۹۴/۱	
۱۸a	۷	۱۶/۴	۱۲/۸	۱/۲	x
۱۸b	۵۶	۵۵/۸	۶۴/۵	۴۷/۵	
۱۸c	۸۸	۸۰/۷	۱۰۰/۰	۷۴/۹	
۱۹a	۱۹	۲۵/۲	۲۶/۴	۱۱/۶	
۱۹b	۶۷	۶۳/۱	۷۵/۳	۵۸/۷	
۱۹c	۹۳	۱۰۰/۰	۱۰۰/۰	۸۴/۴	
۲۰a	۱	۱/۵	۵/۱	۰/۰	
۲۰b	۱۰	۸/۶	۱۴/۳	۵/۷	
۲۰c	۳۰	۲۷/۸	۳۷/۵	۲۲/۵	
۲۱a	۱۳	۱۴/۵	۲۳/۰	۳/۰	
۲۱b	۵۵	۶۰/۸	۶۲/۵	۴۷/۵	
۲۱c	۹۲	۸۲/۳	۱۰۰/۰	۸۱/۵	
۲۲a	۳۱	۳۱/۵	۳۶/۸	۲۵/۲	
۲۲b	۷۵	۷۱/۴	۸۴/۳	۶۵/۷	
۲۲c	۹۵	۱۰۰/۰	۱۰۰/۰	۸۷/۰	
۲۳a	۴۹	۵۰/۴	۵۳/۴	۴۴/۶	
۲۳b	۸۳	۷۳/۹	۹۲/۹	۷۳/۱	
۲۳c	۸۸	۹۵/۰	۹۵/۳	۸۰/۷	
۲۴a	۱۱	۱۷/۳	۲۰/۲	۱/۸	

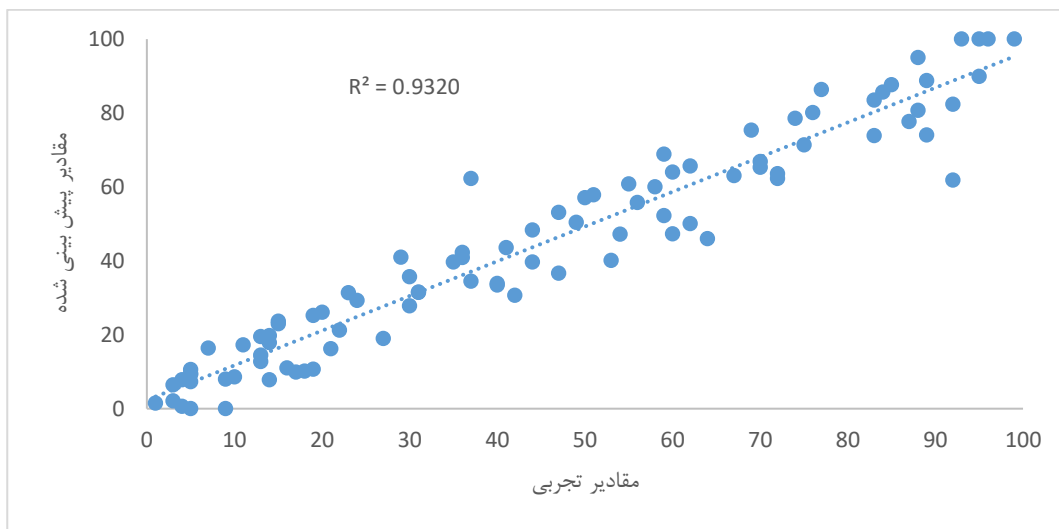
ادامه جدول (۳-۱۱) نتایج حاصل از ارزیابی مدل با استفاده از رد مرحله‌ای تک‌تک برای کل داده‌ها

شماره ترکیب	میزان رنگبری تجربی	میزان رنگبری پیش‌بینی شده	محدوده اطمینان		پذیرش فرضیه صفر
۲۴b	۴۷	۳۶/۷	۵۷/۸	۳۶/۲	
۲۴c	۷۰	۶۶/۸	۷۹/۴	۶۰/۶	
۲۵a	۵	۹/۳	۱۱/۳	۰/۰	
۲۵b	۲۴	۲۹/۳	۳۲/۲	۱۵/۸	
۲۵c	۵۱	۵۷/۸	۵۸/۶	۴۳/۴	
۲۶a	۴	۷/۸	۱۰/۳	۰/۰	
۲۶b	۱۴	۱۹/۷	۲۵/۴	۲/۶	
۲۶c	۴۴	۳۹/۶	۵۴/۲	۳۳/۸	
۲۷a	۲۷	۱۹/۰	۳۶/۱	۱۷/۹	
۲۷b	۶۰	۶۴/۰	۶۷/۵	۵۲/۵	
۲۷c	۸۵	۸۷/۷	۸۸/۴	۸۱/۶	
۲۸a	۱۸	۱۰/۲	۲۸/۳	۷/۷	
۲۸b	۵۰	۵۷/۱	۵۸/۹	۴۱/۱	
۲۸c	۸۷	۷۷/۷	۹۶/۹	۷۷/۱	
۲۹a	۳	۲/۱	۵/۹	۰/۱	
۲۹b	۱۳	۱۲/۸	۱۴/۴	۱۱/۶	
۲۹c	۳۷	۳۴/۵	۴۴/۶	۲۹/۴	
۳۰a	۵	۷/۳	۷/۹	۲/۱	
۳۰b	۲۰	۲۶/۱	۳۱/۲	۸/۸	
۳۰c	۴۷	۵۳/۱	۵۶/۰	۳۸/۰	
۳۱a	۱۴	۷/۸	۲۰/۴	۷/۶	
۳۱b	۳۱	۳۱/۶	۳۹/۱	۲۲/۹	
۳۱c	۵۸	۶۰/۰	۶۵/۰	۵۱/۰	

c: رنگبری در ۱۵ دقیقه

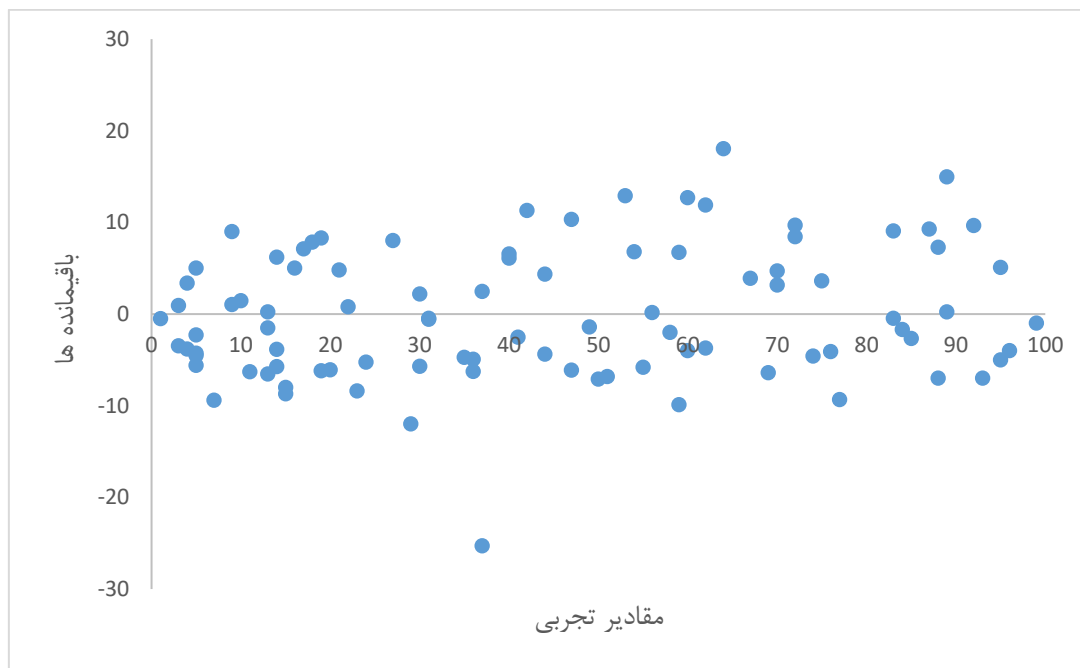
b: رنگبری در ۱۰ دقیقه

a: رنگبری در ۵ دقیقه



شکل (۳-۵) نمودار مقادیر پیش بینی شده بر حسب مقادیر تجربی به روش رد مرحله‌ای تک تک برای داده‌ها

نمودار باقیمانده‌ها معیاری برای شایستگی مدل به دست آمده، می‌باشد و اگر باقیمانده‌ها به‌طور یکنواخت حول محور افقی پراکنده باشند، نشان می‌دهد که مدل مناسبی به دست آمده و خطای سیستماتیکی وجود ندارد. شکل (۳-۶) نمودار باقیمانده‌ها را که از تفاضل مقدار محاسبه شده توسط مدل شبکه عصبی و مقدار تجربی درصد رنگبری به دست می‌آید، بر حسب مقدار تجربی نشان می‌دهد. تقارن پراکندگی نقاط در دو طرف محور افقی حاکی از عدم وجود خطای سیستماتیک می‌باشد.



شکل (۳-۶) نمودار باقیمانده‌ها بر حسب مقادیر تجربی برای کل داده‌ها

۳-۸ - ارزیابی مدل با استفاده از پارامترهای آماری

کارایی مدل طراحی شده توسط شبکه عصبی مصنوعی با استفاده از چندین پارامتر آماری که در بخش (۷-۲-۱) به آن‌ها اشاره شد، مورد بررسی قرار گرفت که نتایج آن در جدول (۳-۱۲) ارائه شده است. مقادیر این پارامترها نشان می‌دهد که مدل ارائه شده از قدرت پیشگویی مناسبی برخوردار است.

جدول (۳-۱۲) پارامترهای آماری برای مدل برتر طراحی شده توسط شبکه عصبی مصنوعی

شماره	پارامترهای آماری	سری آموزش	سری آزمون
۱	PRESS	۱۱۷۲	۹۰۸/۳۱۰
۲	SEP	۳/۹۵۴۱	۷/۱۰۳۶
۳	MAE	۳/۰۸۴۰	۶/۳۵۰۰
۴	REP(%)	۹/۴۲۹۵	۱۳/۲۲۲۹
۵	MSE	۱۵/۶۳۴۸	۵۰/۴۶۱۷
۶	MRE	۱۷/۵۰۵۱	۲۱/۹۰۸۸
۷	R^2	۰/۹۸۲۰	۰/۹۵۵۳
۸	R_0^2	۰/۹۸۱۹	۰/۹۴۴۴
۹	$\frac{(R^2 - R_0^2)}{R^2}$	۰/۰۰۰۱	۰/۰۱۱۴۱
۱۰	$R_0'^2$	۰/۹۸۱۳	۰/۹۵۱۸
۱۱	$\frac{(R^2 - R_0'^2)}{R^2}$	۰/۰۰۰۷	۰/۰۰۳۶
۱۲	$ R_0^2 - R_0'^2 $	۰/۰۰۰۶	۰/۰۰۷۴
۱۳	k	۱/۰۰۲۳	۰/۹۶۶۴
۱۴	k'	۰/۹۹۱۸	۱/۰۲۲

۳-۹ - ارزیابی مدل ارائه شده با استفاده از آزمون Y-تصادفی

این تکنیک ارزیابی مدل با هدف بررسی هر گونه ارتباط تصادفی بین داده‌ها انجام شد. در این

آزمون مقادیر تصادفی از متغیر وابسته با استفاده از دستور Rand بسته نرم افزاری متلب بین بازه ۱-۹۹

تولید گردید. مدل QSPR جدید با استفاده از ماتریس متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر

وابسته توسعه یافت. اگر مدل اصلی هیچ گونه ارتباط تصادفی نداشته باشد، تفاوت قابل توجهی بین مقدار

ضریب تعیین مدل اصلی و مدل QSPR که با پاسخ‌های تصادفی توسعه یافته، وجود دارد. نتایج حاصل از چندین بار اجرای آزمون Y- تصادفی در جدول (۳-۱۳) نشان داده شده است. مقادیر کوچک ضریب تعیین (R^2) بیانگر عدم وجود ارتباط تصادفی یا وابستگی ساختاری به‌سری آموزش در مدل توسعه یافته توسط شبکه عصبی مصنوعی می‌باشد.

جدول (۳-۱۳) مقادیر R^2 برای سری آزمون با استفاده از آزمون Y-تصادفی

تکرار	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
R^2	۰/۰۵۴۳	۰/۰۰۷۹	۰/۰۹۴۱	۰/۰۰۸۴	۰/۰۰۰۴	۰/۰۲۲۱	۰/۰۱۳۹	۰/۰۰۴۹	۰/۰۱۶۴	۰/۰۰۰۵

۳-۱۰ - بررسی ارتباط توصیف‌کننده‌های وارد شده در مدل با درصد رنگبری

الف- توصیف‌کننده‌های topological charge indices

این توصیف‌کننده‌ها جهت ارزیابی و بررسی انتقال بار بین جفت اتم‌ها در یک مولکول و بنابراین

انتقال بار در تمام مولکول پیشنهاد شدند. برای توضیح این توصیف‌کننده‌ها ابتدا ماتریس M را به‌صورت

زیر تعریف می‌کنیم:

$$M=A*D^{-2}$$

(۳-۲)

که در آن A ماتریس مجاورت^۱ و D^{-2} ماتریس عکس مربع فاصله^۲ می‌باشد. M ماتریس گالوز^۳ می‌باشد و یک ماتریس مربعی نامتقارن $A \times A$ می‌باشد که در آن A تعداد اتم‌های مولکول می‌باشد. M باعث یک ماتریس ترم بار^۴ نامتقارن، CT می‌شود که به صورت زیر تعریف می‌گردد:

$$CT = \partial_i \quad \text{if } i=j \quad (3-3)$$

$$CT = m_{ij} - m_{ji} \quad \text{if } i \neq j$$

$$CT_{ij} = D(M) + [M - M^T] \quad (4-3)$$

که در آن m_{ij} عناصر M و ∂_i درجه‌ی بردار اتم i می‌باشد. ورودی‌های قطری CT ظرفیت توپولوژیکی اتم‌ها و ورودی‌های قطری CT_{ij} مقدار بار منتقل شده از اتم j به اتم i می‌باشد.

برای هر فاصله طولی K می‌توان یک اندیس بار توپولوژیکی، G_k به صورت زیر تعریف نمود:

$$G_k = \frac{1}{2} \sum_{i=1}^A \sum_{j=1}^A |CT_{ij}| \delta(k; d_{ij}) \quad (5-3)$$

که در آن $\delta(k; d_{ij})$ دلتای کرونکر^۵ می‌باشد.

$$\delta(k; d_{ij}) = 1 \quad \text{if } d_{ij} = k$$

$$= 0 \quad \text{if } d_{ij} \neq k$$

¹ Adjacency

² Reciprocal square distance

³ Galvez matrix

⁴ Charge term matrix

⁵ Kronecker

که در آن d_{ij} عناصر ماتریس فاصله هستند. بنابراین G_k بار کلی انتقال بین اتم‌ها در یک فاصله‌ی توپولوژیکی k می‌باشد. حداکثر تعداد ترم‌های G_k برابر با قطر توپولوژیکی D است.

این شاخص‌ها یک مقدار دقیقاً توپولوژیکی را نشان می‌دهد، که به طور قابل ملاحظه‌ای با توزیع شارژ داخل مولکول ارتباط دارد. این توزیع یک ویژگی مهم است که بسیاری از خواص فیزیکی و شیمیایی مولکول‌ها را تحت تأثیر قرار می‌دهد. توصیف‌کننده JGI7 از این گروه توصیف‌کننده‌ها وارد مدل شده است.

ب- توصیف‌کننده‌های 2D autocorrelation

توصیف‌کننده‌های گروه Autocorrelation Moran توصیف‌کننده‌های دوبعدی هستند که از

طریق رابطه زیر محاسبه می‌شوند:

$$I(d) = \frac{\frac{1}{2} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - \bar{w})(w_j - \bar{w})}{\frac{1}{A} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad (6-3)$$

که w یک ویژگی اتم، \bar{w} میانگین مقدار آن روی مولکول، A تعداد اتم‌ها و d فاصله‌ی توپولوژیکی است و δ_{ij} نیز که به تابع کرونکر معروف است. در حالتی که $d_{ij} = d$ باشد یک است، در غیر این صورت صفر است. مقدار این توصیف‌کننده که از جنس فاصله است، از $+1$ تا -1 متغیر است [۷۲، ۷۱]. MATS1m^۱ و MATS7m^۲ از این گروه توصیف‌کننده‌هاست که در مدل برتر ظاهر شده است که با جرم اتمی وزن‌دار شده در مدل انتخاب شده و نشان می‌دهد که وزن کلی استخلاف، فاکتوری تعیین‌کننده برای رنگبری آن

¹ - Moran autocorrelation-lag1 weighted by atomic masses

² Moran autocorrelation-lag7 weighted by atomic masses

ترکیب می‌باشد که با کاهش مقادیر این توصیف‌کننده‌ها رنگبری نیز افزایش می‌یابد. جدول (۳-۱۴) چگونگی ارتباط این توصیف‌کننده با رنگبری را برای تعدادی از مولکول‌های سری داده‌ها نشان می‌دهد.

جدول (۳-۱۴) نمایش ارتباط خاصیت با مقدار توصیف‌کننده MATS1m برای بعضی از ترکیبات مورد مطالعه

شماره ترکیب	مقدار MATS1m	درصد رنگبری
۲۰ c	۰/۳۰۱	۳۰/۰۰
۲۵ c	۰/۱۷۵	۵۱/۰۰
۱۸ c	-۰/۱۵۵	۸۸/۰۰
۱۵ c	-۰/۲۱۶	۹۶/۰۰

a: رنگبری در ۵ دقیقه

b: رنگبری در ۱۰ دقیقه

c: رنگبری در ۱۵ دقیقه

ج- توصیف‌کننده‌های WHIM^۱

این شاخص از مختصات کارتیزین ساختار سه‌بعدی مولکول، با استفاده از صورت‌بندی با حداقل انرژی محاسبه می‌شود و شامل اطلاعاتی درباره‌ی اندازه، شکل، تقارن و توزیع اتمی ساختار سه‌بعدی مولکول می‌باشد. این توصیف‌کننده از رابطه (۷-۳) به دست می‌آید:

$$S_{jk} = \frac{\sum_{i=1}^A w_i (q_{ij} - \bar{q}_j) (q_{ik} - \bar{q}_k)}{\sum_{i=1}^A w_i} \quad (7-3)$$

¹ -Weighted Holistic Invariant Molecular descriptors

که S_{ijk} کوواریانس وزن دار شده بین کئوردینه j ام و k ام، A تعداد اتم‌ها، w_i وزن i امین اتم، q_{ij} و q_{ik} به ترتیب j امین و k امین کئوردینه‌های اتم i ام، و \bar{q} مقدار میانگین متناظر است. شش طرح وزن دار شدن پیشنهاد شده است که عبارتند از:

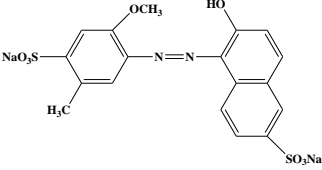
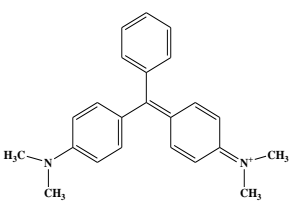
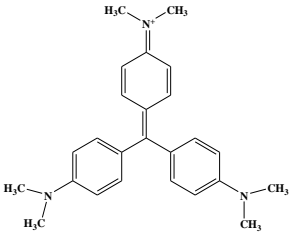
- ۱ - حالت بدون وزن (u) ۲- وزن دار شده با جرم اتمی ۳- وزن دار شده با حجم و اندروالس ۴-
- وزن دار شده با الکترونگاتیویته ساندرسون ۵- وزن دار شده با حالت الکتروتوپولوژیکی کییر و هال ۶- وزن دار شده با قطبش‌پذیری [۷۱، ۷۳]. از این نوع توصیف‌کننده‌ها $E2s$ و $G2u$ توسط مدل انتخاب گردید که به تأثیر صورتبندی رنگ‌ها در رنگبری آنها اشاره دارد.

د- توصیف‌کننده‌های atom-centred fragments

این توصیف‌کننده تأثیر گروه‌های CH_3-X را بر روی رنگبری نشان می‌دهد. که X نشان دهنده‌ی اتم الکترونگاتیو (O, N, S, P, Se و هالوژن‌ها) است. وجود این گروه‌ها بر روی ساختار مولکول باعث افزایش رنگبری می‌گردد. جدول (۳-۱۵) چگونگی ارتباط این توصیف‌کننده با میزان رنگبری را برای تعدادی از مولکول‌های سری داده‌ها نشان می‌دهد.

¹ - Kier-hall

جدول (۳-۱۵) نمایش ارتباط خاصیت با مقدار توصیف کننده $\text{CH}_3\text{-X}$ برای بعضی از ترکیبات مورد مطالعه

ساختار ترکیب	تعداد $\text{CH}_3\text{-X}$	زمان رنگبری	میزان رنگبری
	۱	a	۵
		b	۲۳
		c	۷۰
	۴	a	۷
		b	۵۶
		c	۸۸
	۶	a	۲۲
		b	۷۲
		c	۹۶

c: رنگبری در ۱۵ دقیقه

b: رنگبری در ۱۰ دقیقه

a: رنگبری در ۵ دقیقه

۵- توصیف کننده‌های 2D binary fingerprints

این توصیف کننده حضور یا عدم حضور N-N در فاصله توپولوژیکی ۸ را نشان می‌دهد. اثر انگشت باینری اندازه مولکول‌ها را به‌طور غیر مستقیم بررسی می‌کند. در حالت تجربی مولکول‌های بزرگتر، دارای ویژگی‌های متنوع‌تری هستند. در حالت تئوری فاصله توپولوژیکی مولکول‌های بزرگ ثابت است و اثر انگشت باینری حضور یا عدم حضور ویژگی‌های موثر در عملکرد مولکول را نسبت به حالتی که ویژگی‌های واقعی

بررسی شده‌اند مورد بررسی قرار می‌دهد. توصیف‌کننده‌های اثر انگشتی اندازه مولکول را بررسی می‌کنند و همه این ویژگی‌های بررسی شده برای مولکول باید با اندازه مولکول در سطح اولیه همبسته باشد. که از میان توصیف‌کننده‌های اثر انگشت B08NN در مدل برتر دیده می‌شود [۷۴] که نشان می‌دهد در حضور N-N در فاصله توپولوژیکی ۸، رنگبری افزایش پیدا می‌کند.

ز- توصیف‌کننده‌های Information indices

این توصیف‌کننده‌ها جزء شاخص‌های اطلاعاتی هستند و در بردارنده اطلاعاتی راجع به ماهیت شیمیایی، پیوند اتم‌ها در فضا، توپولوژی مولکولی و تقارن می‌باشند. این شاخص‌های اطلاعاتی از سطح صفر تا سطح پنج محاسبه می‌شوند. توصیف‌کننده شاخص حاوی اطلاعات (تقارن همسایگی از مرتبه ۲) بر پایه محاسبه‌ی دسته‌های تعادلی از گراف مولکول است [۷۱]. از میان توصیف‌کننده‌های شاخص اطلاعاتی، توصیف‌کننده IC2 در مدل برتر دیده می‌شود.

۱۱-۳- بررسی میزان مشارکت توصیف‌کننده‌های منتخب در شبکه عصبی

میزان مشارکت توصیف‌کننده‌های منتخب به صورت زیر تعیین شد [۷۵-۷۷]:

۱- توصیف‌کننده مورد نظر به همراه اوزان مربوطه‌اش از شبکه بهینه شده حذف گردید.

۲- با استفاده از بقیه توصیف‌کننده‌ها مقدار متغیر وابسته (میزان رنگبری) برای هر ترکیب سری

آموزش به روش BR-ANN پیش‌بینی شد.

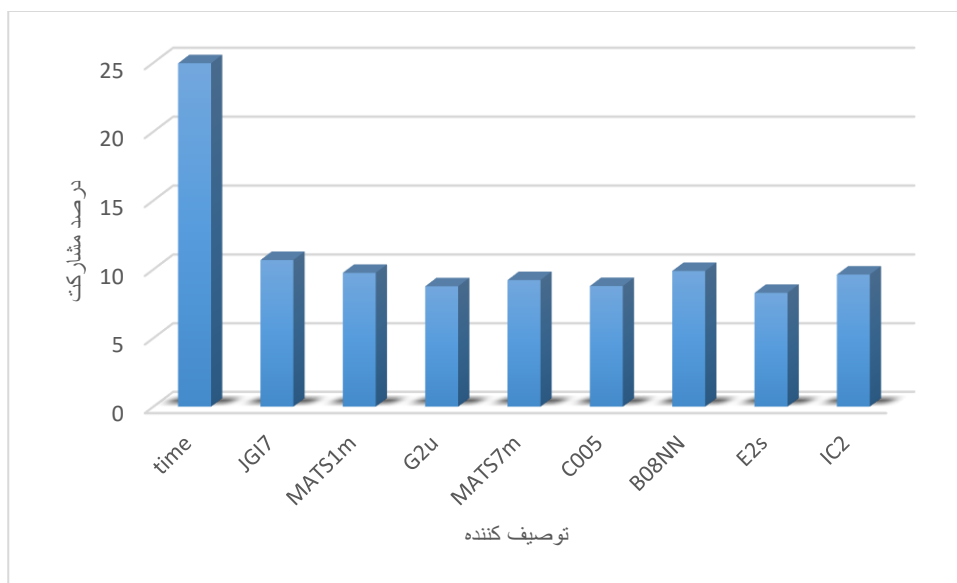
۳- میانگین خطای مطلق حاصل از BR-ANN برای ترکیبات سری آموزش محاسبه گردید.

۴- مراحل ۱ تا ۳ برای دیگر توصیف‌کننده‌های منتخب نیز تکرار شد.

۵- سرانجام درصد مشارکت هر توصیف‌کننده توسط رابطه (۸-۳) برآورد شد.

$$c_i = 100 \frac{\Delta m_i}{\sum_{i=1}^N \Delta m_i} \quad \text{رابطه (۳-۸)}$$

در رابطه (۳-۸) درصد مشارکت توصیف‌کننده حذف شده Δm ، N تعداد توصیف‌کننده‌های مدل و Δm_i میانگین خطای مطلق حاصل از روش BR-ANN برای سری آموزش در غیاب توصیف‌کننده Δm را نشان می‌دهد. میزان مشارکت توصیف‌کننده‌های منتخب در پیش‌بینی رنگبری مورد نظر برآورد شد که نتایج آن در شکل (۳-۷) ارائه شده است.



شکل (۳-۷) مشارکت توصیف‌کننده‌ها در شبکه عصبی بهینه

بر اساس نمودار فوق توصیف‌کننده زمان دارای بیشترین اثر مشارکت می‌باشد. همان‌گونه که مشخص است با افزایش زمان رنگبری نیز افزایش می‌یابد. توصیف‌کننده JGI7 یک مقدار دقیقاً توپولوژیکی را نشان می‌دهد، که به طور قابل ملاحظه‌ای با توزیع بار داخل مولکول ارتباط دارد. این توصیف‌کننده بعد از توصیف‌کننده زمان دارای بیشترین درصد مشارکت می‌باشد.

فصل چهارم

نتیجه‌گیری و آینده نگری

۴-۱ - نتیجه‌گیری

امروزه با توجه به این موضوع که تحقیقات آزمایشگاهی نیازمند صرف هزینه و وقت بسیار بوده و همواره با خطرات زیاد و عوارض زیست محیطی همراه است، استفاده از مدل‌های پیش‌بینی‌کننده جهت انجام تحقیقات با هدف و نتیجه‌بخش، بسیار گسترش یافته است. تاکنون با استفاده از ابزارهای مختلف QSPR/QSAR مدل‌های بسیاری توسعه یافته‌اند که قابلیت اطمینان و قدرت پیش‌بینی‌کنندگی بسیار بالایی دارند. سهولت، سادگی، دقت بالا، قابلیت اطمینان و قدرت پیش‌بینی‌کنندگی بالا از مزایای اغلب مدل‌های ارائه شده با این روش می‌باشد. در این تحقیق از روش فوتوکاتالیستی برای حذف رنگ‌های مختلف استفاده شد. نتایج حاصل از این تحقیق در زیر آورده شده است.

در این تحقیق از روش مدل‌سازی شبکه عصبی مصنوعی با الگوریتم آموزشی تنظیم بایزین به عنوان روشی قدرتمند و کارآمد برای پیش‌بینی رنگبری استفاده شد که بر اساس اطلاعات موجود این تحقیق به عنوان اولین گزارش در مورد مدل‌سازی QSPR برای پیش‌بینی میزان رنگبری فوتوکاتالیستی رنگ‌ها، با ساختارهای متنوع و گروه‌های مختلف می‌باشد.

همان‌طور که گفته شد هدف اصلی از کار حاضر، بررسی ارتباط کمی ساختار-خاصیت برای پیش‌بینی میزان رنگبری از رنگ‌های مختلف می‌باشد. نتایج نشان می‌دهند که مدل غیرخطی توسعه یافته می‌تواند یک رابطه‌ی رضایت‌بخش بین ساختار مولکولی و رنگبری برقرار کند. در انتخاب توصیف‌کننده‌ها، زمان به عنوان توصیف‌کننده با اهمیت بالا انتخاب شد. برای انتخاب سایر توصیف‌کننده‌ها در این پایان‌نامه از روش فراوانی توصیف‌کننده‌ها استفاده شد تا ارتباط مناسبی بین ساختار هر رنگ و رنگبری آن برقرار شود. ۸ توصیف‌کننده ساختاری به علاوه توصیف‌کننده زمان که با استفاده از روش فراوانی توصیف‌کننده‌ها انتخاب شد به عنوان توصیف‌کننده‌های ساختاری موثر در رنگبری، مورد بررسی قرار گرفت. نتایج نشان

دهنده این است که سه توصیف کننده زمان، zGI7 و B08NN بیشترین تاثیر را در رنگبری دارند. که توصیف کننده zGI7 با توزیع بار داخل مولکول ارتباط دارد و توصیف کننده B08NN که مربوط به اثر انگشت باینری دو بعدی است، حضور یا عدم حضور N-N در فاصله توپولوژیکی ۸ را نشان می دهد.

نتایج حاصل از ارزیابی مدل برتر با کمک آزمون Y-تصادفی، روش حذف تک تک، نمودار باقی مانده و سایر پارامترهای ارزیابی، تاییدی بر قدرت مدل انتخاب شده است.

۴-۲ - آینده نگری

- ✓ پژوهشگران می‌توانند در شرایط متغیر آزمایشگاهی به پیش بینی رنگیری بپردازند.
- ✓ می‌توان از روش‌هایی همچون الگوریتم ژنتیک و طرح‌ریزی متوالی^۱ برای انتخاب توصیف کننده‌های مهم استفاده کرد همچنین می‌توان روش‌هایی مانند SVM^۲ و LS-SVM^۳ را به جای شبکه عصبی به کار برد.
- ✓ می‌توان روش‌های خطی مانند حداقل مربعات جزئی و آنالیز اجزای اصلی را جایگزین روش MLP کرد و نتایج به دست آمده را با هم مقایسه کرد.

^۱ - Successive Projection Algorithm

^۲ - Supported Vector Machine

^۳ - Least Square Supported Vector Machine

مراجع:

- [۱] م. پیکری، ا. مهربانی، مبانی تصفیه آب، انتشارات اشکان، ۱۳۸۳.
- [۲] ح. گنجی، ب. آیتی، کاربرد راکتورهای بیوفیلمی با بستر متحرک (MBBR)، انتشارات دانشگاه تربیت مدرس، تهران، ۱۳۸۹.
- [3] J. Khayatzadeh, E. Abbasi, The effects of heavy metals on aquatic animals, in: The 1st International Applied Geological Congress, Department of Geology, Islamic Azad University–Mashad Branch, Iran, **2010**, pp. 26-28.
- [4] S. Wang, A comparative study of Fenton and Fenton-like reaction kinetics in decolourisation of wastewater, *Dyes and Pigments*, 76 (**2008**) 714-720.
- [۵] م.ص. نیاسری، ا. سبحانی، نانو رنگدانه، تهران، انتشارات علم ودانش، ۱۳۸۸.
- [6] W. Herbst, K. Hunger, Industrial organic pigments: production, properties, applications, John Wiley & Sons, **2006**.
- [۷] ب. فهیمی، سنتز مورفولوژی‌های مختلف نانو ساختارهای ZnO و ZnO/Au و مطالعه فعالیت فتوکاتالیزوری آنها در تخریب نوری رنگدانه، دانشکده شیمی، دانشگاه صنعتی شاهرود، ۱۳۹۰.
- [8] A. Greaves, J. Churchley, M. Hutchings, D. Phillips, J. Taylor, A chemometric approach to understanding the bioelimination of anionic, water-soluble dyes by a biomass using empirical and semi-empirical molecular descriptors, *Water research*, 35 (**2001**) 1225-1239.
- [9] T. Suzuki, S. Timofei, L. Kurunczi, U. Dietze, G. Schüürmann, Correlation of aerobic biodegradability of sulfonated azo dyes with the chemical structure, *Chemosphere*, 45 (**2001**) 1-9.

[10] H. Kusic, N. Koprivanac, A.L. Bozic, Environmental aspects on the photodegradation of reactive triazine dyes in aqueous media, *Journal of Photochemistry and Photobiology A: Chemistry*, 252 (2013) 131-144.

[11] B. Li, Y. Dong, Z. Ding, Heterogeneous Fenton degradation of azo dyes catalyzed by modified polyacrylonitrile fiber Fe complexes: QSPR (quantitative structure property relationship) study, *Journal of Environmental Sciences*, 25 (2013) 1469-1476.

[12] U.G. Akpan, B.H. Hameed, Parameters affecting the photocatalytic degradation of dyes using TiO₂-based photocatalysts: a review, *Journal of hazardous materials*, 170 (2009) 520-529.

[13] F. Erol, Advanced Oxidation Techniques for the Removal of Refractory Organics From Textile Wastewaters, Master Of Science, The Graduate School of Natural And Applied Sciences Of Middle East Technical University, Ankara, (2008).

[14] F. Harrelkas, A. Paulo, M. Alves, L. El Khadir, O. Zahraa, M.-N. Pons, F. Van Der Zee, Photocatalytic and combined anaerobic-photocatalytic treatment of textile dyes, *Chemosphere*, 72 (2008) 1816-1822.

[15] D. Sannino, P. Ciambelli, M. Ricciardi, A.I. Lyubov, Wastewater treatment by high efficiency heterogeneous photo-Fenton process, in, *Google Patents*, 2010.

[16] N.M. Mahmoodi, M. Arami, Bulk phase degradation of Acid Red 14 by nanophotocatalysis using immobilized titanium (IV) oxide nanoparticles, *Journal of Photochemistry and Photobiology A: Chemistry*, 182 (2006) 60-66.

[17] س.م. ها شمی, خ. بدیعی, ا. اکبرزاده, رنگبری پساب های رنگی با استفاده از نانو ذرات دی اکسید

تیتانیوم بر روی پوشش پلیمری, علوم پایه پژوهشگران جوان, دانشگاه فردوسی مشهد, ۱۳۹۱.

[18] C.G. da Silva, J.L. Faria, Photochemical and photocatalytic degradation of an azo dye in aqueous solution by UV irradiation, *Journal of Photochemistry and Photobiology A: Chemistry*, 155 (2003) 133-143.

[19] V. Augugliaro, M. Litter, L. Palmisano, J. Soria, The combination of heterogeneous photocatalysis with chemical and physical operations: A tool for improving the photoprocess performance, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 7 (2006) 127-144.

[20] S. Esplugas, J. Gimenez, S. Contreras, E. Pascual, M. Rodríguez, Comparison of different advanced oxidation processes for phenol degradation, *Water research*, 36 (2002) 1034-1042.

[21] J.C. Colmenares, R. Luque, J.M. Campelo, F. Colmenares, Z. Karpiński, A.A. Romero, Nanostructured photocatalysts and their applications in the photocatalytic transformation of lignocellulosic biomass: an overview, *Materials*, 2 (2009) 2228-2258.

[۲۲] م. صلواتی, نانوکاتالیست, انتشارات علم و دانش, تهران, ۱۳۸۸.

[23] A.L. Linsebigler, G. Lu, J.T. Yates Jr, Photocatalysis on TiO₂ surfaces: principles, mechanisms, and selected results, *Chemical reviews*, 95 (1995) 735-758.

[24] S. Li, Structural design, characterization, and property investigation of iron oxide: nanoparticles with visible light photoactivity, Metz, 2009.

[25] U.I. Gaya, A.H. Abdullah, Heterogeneous photocatalytic degradation of organic contaminants over titanium dioxide: a review of fundamentals, progress and problems, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 9 (2008) 1-12.

[26] M. Rauf, S.S. Ashraf, Fundamental principles and application of heterogeneous photocatalytic degradation of dyes in solution, *Chemical engineering journal*, 151 (2009) 10-18.

[27] Q. Wang, J. Lian, Y. Bai, J. Hui, J. Zhong, J. Li, N. An, J. Yu, F. Wang, Photocatalytic activity of hydrogen production from water over TiO₂ with different crystal structures, *Materials Science in Semiconductor Processing*, 40 (2015) 418-423.

- [28] Ü. Ocak, H. Alp, P. Gökçe, M. Ocak, The synthesis of new N₂S₂-macrocyclic schiff base ligands and investigation of their ion extraction capability from aqueous media, *Separation science and technology*, 41 (2006) 391-401.
- [29] M. Nasirian, C.F. Bustillo-Lecompte, M. Mehrvar, Photocatalytic efficiency of Fe₂O₃/TiO₂ for the degradation of typical dyes in textile industries: Effects of calcination temperature and UV-assisted thermal synthesis, *Journal of environmental management*, 196 (2017) 487-498.
- [30] S. Sakthivel, B. Neppolian, M. Shankar, B. Arabindoo, M. Palanichamy, V. Murugesan, Solar photocatalytic degradation of azo dye: comparison of photocatalytic efficiency of ZnO and TiO₂, *Solar Energy Materials and Solar Cells*, 77 (2003) 65-82.
- [31] M.A. Behnajady, H. Eskandarloo, F. Eskandarloo, Artificial neural network modeling of the influence of sol-gel synthesis variables on the photocatalytic activity of TiO₂ nanoparticles in the removal of Acid Red 27, *Research on Chemical Intermediates*, 41 (2015) 6463-6476.
- [32] R.G. Brereton, *Chemometrics: data analysis for the laboratory and chemical plant*, John Wiley & Sons, 2003.
- [33] I.E. Frank, B.R. Kowalski, *Chemometrics, Analytical Chemistry*, 54 (1982) 232-243.
- [34] B.R. Kowalski, *Chemometrics, Journal of Analytical Chemistry*, (1980) 112-122.
- [35] S. Deming, Y. Michotte, D.L. Massart, L. Kaufman, B. Vandeginste, *Chemometrics: a textbook*, Elsevier, 1988.
- [36] D.B. Hibbert, *Genetic algorithms in chemistry, Chemometrics and Intelligent Laboratory Systems*, 19 (1993) 277-293.
- [37] G.G. Andersson, P. Kaufmann, *Development of a generalized neural network, Chemometrics and Intelligent Laboratory Systems*, 50 (2000) 101-105.
- [38] M. Arab Chamjangali, *Modelling of Cytotoxicity Data (CC50) of Anti-HIV 1-[5-Chlorophenyl] Sulfonyl]-1H-Pyrrole Derivatives Using Calculated Molecular Descriptors*

and Levenberg–Marquardt Artificial Neural Network, Chemical biology & drug design, 73 (2009) 456-465.

[39] H. Inc, HyperChem8.0.5, Toronto, Canada: .

[۴۰] ا. لواين, شیمی کوانتومی, موسسه انتشارات فاطمی, ۱۳۸۷.

[41] P.W. Atkins, R.S. Friedman, Molecular quantum mechanics, Oxford university press, 2011.

[۴۲] م. عرب.چم جنگلی, پیش‌بینی فعالیت دارویی ضد ایدز (سیتوتوکسیستی) مشتقات ۵-فنیل-۱-فنیل آمینو ۱-H-ایمیدازول به وسیله شبکه عصبی مصنوعی "، دانشگاه صنعتی شاهرود, "، دانشگاه صنعتی شاهرود, ۱۳۸۶.

[43] www.iasbs.ac.ir/chemistry/chemometrics/.../8th/qsar_introduction.ppt

[44] A. Habibi-Yangjeh, M. Danandeh-Jenagharad, Prediction of acidity constant for substituted acetic acids in water using artificial neural networks, (2007) 478-487.

[۴۵] ع. فرشادفر, اصول و روشهای پیشرفته آماری (تجزیه رگرسیون), انتشارات طاق بستان, ۱۳۸۰.

[۴۶] م. اشرفی, مطالعه ارتباط کمی ساختار- فعالیت مشتقات تیوکرباماتها به عنوان دسته ی جدیدی از بازدارنده های غیر نوکلئوزیدی, HIV دانشکده شیمی دانشگاه صنعتی شاهرود, ۱۳۸۹.

[47] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics, 5 (1943) 115-133.

[48] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychological review, 65 (1958) 386-408.

[49] D.E. Rumelhart, J.L. McClelland, Parallel distributed processing: explorations in the microstructure of cognition. volume 1. foundations, (1986).

[۵۰] م. منهاج، مبانی شبکه های عصبی (هوش محاسباتی)، مرکز نشر دانشگاه صنعتی امیرکبیر تهران، ۱۳۸۷.

[۵۱] ت. هاگان، ه. دیموث، م. بیل، طراحی شبکه های مصنوعی، انتشارات کیان رایانه سبز، ۲۰۰۱.

[۵۲] م.ع. فردوسی، پیش بینی ثابت های هنری بعضی از ترکیبات آلی با استفاده از روش های خطی و غیرخطی، QSPR، دانشکده شیمی، دانشگاه صنعتی شاهرود، ۱۳۸۹.

[۵۳] س. چترچی، ع. هادی، ب. پرایس، تحلیل رگرسیون با مثال، مشهد، انتشارات دانشگاه فردوسی مشهد، ۱۳۸۴.

[54] E.B. de Melo, M.M.C. Ferreira, Nonequivalent effects of diverse LogP algorithms in three QSAR studies, *QSAR & Combinatorial Science*, 28 (2009) 1156-1165.

[55] R. Bansal, C. Karthikeyan, N. Moorthy, P. Trivedi, QSAR analysis of some phthalimide analogues based inhibitors of HIV-1 integrase, *Arkivoc*, 15 (2007) 66-81.

[56] A. Golbraikh, A. Tropsha, Beware of $q^2!$, *Journal of molecular graphics and modelling*, 20 (2002) 269-276.

[57] S. Funar-Timofei, L. Kurunczi, V. Vlaia, T. Olariu, D. Ciubotariu, Quantitative dye structure-toxicity relationships study by PLS,: Proceedings of the 2nd conference on European computing conference, World Scientific and Engineering Academy and Society (WSEAS), 2008, pp. 108-113.

[58] R. Todeschini, DRAGON,, Milano Chemometrics and QSPR Group, 2006.

[59] M.J. Norušis, SPSS, SPSS advanced statistics: Student guide. SPSS, Incorporated, 1990.

[60] I. MathWorks, , MATLAB: Application program interface guide. Vol. 5. MathWorks, 1996.

- [61] M. Zahouily, J. Rakik, M. Lazar, M. Bahlaoui, A. Rayadh, N. Komiha, Exploring QSAR of non-nucleoside reverse transcriptase inhibitors by artificial neural networks: HEPT derivatives, *ARKIVOC*, 14 (2007) 245-256.
- [62] J. Zupan, J. Gasteiger, *Neural networks for chemists: an introduction*, John Wiley & Sons, Inc., 1993.
- [63] M.A. Chamjangali, M. Ashrafi, QSAR study of necroptosis inhibitory activities (EC 50) of [1, 2, 3] thiadiazole and thiophene derivatives using Bayesian regularized artificial neural network and calculated descriptors, *Medicinal Chemistry Research*, 22 (2013) 392-400.
- [64] B.M. Bababdani, M. Mousavi, Gravitational search algorithm: a new feature selection method for QSAR study of anticancer potency of imidazo [4, 5-b] pyridine derivatives, *Chemometrics and Intelligent Laboratory Systems*, 122 (2013) 1-11.
- [65] M.A. Chamjangali, M. Mohammadrezaei, Z. Kalantar, A.H. Amin, Bayesian Regularized Artificial Neural Network Modeling of the Anti-protozoal Activities of 1-Methylbenzimidazole Derivatives Against *T. Vaginalis* Infection, *Journal of the Chinese Chemical Society*, 59 (2012) 743-752.
- [66] M. Fernández, J. Caballero, Bayesian-regularized genetic neural networks applied to the modeling of non-peptide antagonists for the human luteinizing hormone-releasing hormone receptor, *Journal of Molecular Graphics and Modelling*, 25 (2006) 410-422.
- [67] M. Fernández, J. Caballero, A. Tundidor-Camba, Linear and nonlinear QSAR study of N-hydroxy-2-[(phenylsulfonyl) amino] acetamide derivatives as matrix metalloproteinase inhibitors, *Bioorganic & medicinal chemistry*, 14 (2006) 4137-4150.
- [68] F.W. Bell, A.S. Cantrell, M. Hoegberg, S.R. Jaskunas, N.G. Johansson, C.L. Jordan, M.D. Kinnick, P. Lind, J.M. Morin Jr, Phenethylthiazolethiourea (PETT) compounds, a new class of HIV-1 reverse transcriptase inhibitors. 1. Synthesis and basic structure-activity relationship studies of PETT analogs, *Journal of Medicinal Chemistry*, 38 (1995) 4929-4936.
- [69] F. Burden, D. Winkler, Bayesian regularization of neural networks, *Artificial neural networks*, Springer, (2008), pp. 23-42.

- [70] X. Chen, X. Liu, Q. Meng, D. Wang, H. Liu, E. De Clercq, C. Pannecouque, J. Balzarini, X. Liu, Novel piperidinylamino-diarylpyrimidine derivatives with dual structural conformations as potent HIV-1 non-nucleoside reverse transcriptase inhibitors, *Bioorganic & medicinal chemistry letters*, 23 (2013) 6593-6597.
- [71] R. Todeschini, V. Consonni, *Handbook of molecular descriptors*, John Wiley & Sons, 2008.
- [72] P. Broto, G. Moreau, C. Vanduycke, Molecular structures: perception, autocorrelation descriptor and SAR studies. Autocorrelation descriptor, *European journal of medicinal chemistry*, 19 (1984) 66-70.
- [73] N. Moorthy, C. Karthikeyan, P. Trivedi, QSAR studies of cytotoxic acridine 5, 7-diones: A comparative study using P-VSA descriptors and topological descriptors, (2007).
- [74] T. Ewing, J.C. Baber, M. Feher, Novel 2D fingerprints for ligand-based virtual screening, *Journal of chemical information and modeling*, 46 (2006) 2423-2431.
- [75] L. Douali, D. Villemin, D. Cherqaoui, Neural networks: accurate nonlinear QSAR model for HEPT derivatives, *Journal of chemical information and computer sciences*, 43 (2003) 1200-1207.
- [76] J. Caballero, M. Garriga, M. Fernández, 2D Autocorrelation modeling of the negative inotropic activity of calcium entry blockers using Bayesian-regularized genetic neural networks, *Bioorganic & medicinal chemistry*, 14 (2006) 3330-3340.
- [77] Z. Cheng, Y. Zhang, W. Fu, QSAR study of carboxylic acid derivatives as HIV-1 Integrase inhibitors, *European journal of medicinal chemistry*, 45 (2010) 3970-3980.

ABSTRACT

The industrial textile effluents contain a large variety of dyes that without proper treatment prior to their disposal become toxic to microorganisms and aquatic life. The indiscriminate release of dyes has a considerable negative impact on the environment and human health.

In this study, QSPR model based on Bayesian regularized artificial neural network (BR-ANN) was developed for the modelling and accurate prediction of photo-catalytic dye decolorization efficiency at differentiate based on their chemical structure. For this purpose, the photo-catalytic decolorization of various dyes (31 dyes) have been investigated in TiO₂/UV suspensions. Under constant experimental conditions including amount of photo-catalyst (0.0100g), sample volume (50.0 ml) and pH of 7.0 decolorization efficiency of each dye were measured at 5, 10 and 15 minutes after initiation of UV irradiation. Dyes structures were converted to the molecular descriptors using Dragon software and the most significant descriptors were selected and used as inputs in ANN modelling. After obtaining the BR-ANN model as an optimum model, the prediction ability of the model was evaluated by test set and leave-one-out method. The squared correlation coefficients obtained for the test set and LOO were 0.9553 and 0.9320, respectively. The results obtained, showed the superior prediction ability of the proposed model in the prediction of decolorization efficiency.

Keywords: “Artificial Neural Network, TiO₂/UV, Advanced Oxidation Processes, decolorization efficiency, QSPR”



Shahrood University of Technology

Faculty of Chemistry

M.Sc. Thesis in Analytical Chemistry

Modeling of dyes removal from waste water using the
experimental parameters and calculated structural descriptors

By:

Davood Nadali

Supervisor:

Dr. Mansour Arab Chamjangali

Advisor:

Dr. Nasser Goudarzi

September 2018