





دانشگاه صنعتی شاهرود

دانشکده شیمی

گروه: شیمی تجزیه

مطالعه کمی ساختار – فعالیت IC_{50} برخی از مشتقات گوانین به عنوان

بازدارنده CDK2 در بدن

دانشجو: پیام کلهر

استاد راهنما:

جناب آقای دکتر گودرزی

استاد مشاور:

جناب آقای دکتر نیکوفرد

پایان نامه کارشناسی ارشد جهت اخذ درجه کارشناسی ارشد

اردیبهشت ماه ۱۳۹۰

این مجموعه را به پدر و مادر بزرگوام تقدیم می‌کنم،

در ضمن از زحمات استاد فرهیخته، جناب آقای دکتر گودرزی و سایر اساتید دانشگاه صنعتی شاهرود

کمال تشکر و سپاسگزاری را دارم،

در نهایت از حمایت‌های معنوی سرکار خانم یاری‌وند نیز کمال تشکر را دارم.

چکیده

در بخش اول این تحقیق، ارتباط کمی ساختار - فعالیت (QSAR) ۵۶ ترکیب بازدارنده CDK2 با استفاده از دو روش خطی (رگرسیون خطی چندگانه؛ MLR) و غیر خطی (شبکه عصبی مصنوعی؛ ANN) مدل سازی شد. بازدارنده های CDK2 (کیناز وابسته به حلقه) ترکیباتی هستند که نقش اساسی در تنظیمات سلولی داشته و از تکثیر اضافی سلولها توسط CDK2 که دچار فعالیت بیش از حد شده و منجر به سرطان می شوند جلوگیری می کنند. توانایی پیش بینی مدلها با استفاده از سری ارزیابی، سری تست، روش رد مرحله ای تک تک و Y- تصادفی مورد ارزیابی قرار گرفت و خطای مجذور میانگین (MSE) برای سری تست به روش غیرخطی و خطی به ترتیب ۰/۰۶۳ و ۰/۰۶۹ می باشد. نتایج حاصله نشان دهنده ی قدرت پیش بینی بهتر مدل غیر خطی می باشد.

در بخش دوم، ارتباط کمی ساختار - ویژگی (QSPR) مربوط به شاخص بازداری ۱۹۶ ترکیب مونومتیل آلکان با استفاده از دو روش غیر خطی (شبکه عصبی مصنوعی؛ ANN) و خطی (رگرسیون خطی چندگانه؛ MLR) مدل سازی شد. توانایی پیش بینی مدلها با استفاده از سری ارزیابی، سری تست، روش رد مرحله ای تک تک و Y- تصادفی مورد ارزیابی قرار گرفت و خطای مجذور میانگین (MSE) برای سری تست به روش غیرخطی و خطی به ترتیب ۴/۲۱ و ۲۸/۹۶ می باشد.

کلید واژه ها: ارتباط کمی ساختار - فعالیت، ارتباط کمی ساختار - ویژگی، رگرسیون خطی چندگانه، شبکه عصبی مصنوعی، فعالیت بیولوژیکی، شاخص بازداری

نتایج به دست آمده از این پژوهش طی دو پوستر تحت عناوین

1. Linear and nonlinear QSAR study of N2 and O6 substituted Guanine Derivatives as Cyclin-Dependent Kinase 2 (CDK2) Inhibitors
2. Linear and nonlinear QSPR study of temperature-programmed gas chromatography linear retention indices of all C4 – C30 monomethylalkanes on methylsilicone OV-1 stationary phase

در هفدهمین سمینار شیمی تجزیه ایران در دانشگاه کاشان ارائه گردید.

فهرست مطالب

فصل اول : مقدمه

۲	-----	۱-۱- تعریف تحقیق
۳	-----	۱-۱-۱- کاربرد QSAR در طراحی دارو
۴	-----	۱-۱-۲- شاخص بازداري
۴	-----	۲-۱- ضرورت تحقیق
۵	-----	۳-۱- پیشینه‌ی تحقیق

فصل دوم : کمومتریکس

۸	-----	۱-۲- ارتباط کمی ساختار - ویژگی / فعالیت
۱۰	-----	۱-۱-۲- سری داده‌ها
۱۰	-----	۲-۱-۲- بهینه‌سازی ساختار هندسی
۱۱	-----	۳-۱-۲- محاسبه‌ی توصیف‌کننده‌ها
۱۱	-----	۴-۱-۲- انتخاب بهترین توصیف‌کننده‌ها
۱۲	-----	۵-۱-۲- مدل سازی
۱۳	-----	۶-۱-۲- رگرسیون خطی چندگانه (MLR)
۱۶	-----	۷-۱-۲- شبکه عصبی مصنوعی
۱۸	-----	۱-۷-۱-۲- اساس بیولوژیکی شبکه‌های عصبی مصنوعی
۱۹	-----	۲-۷-۱-۲- ساختار و عملکرد شبکه‌های عصبی مصنوعی (مدل ریاضی)
۲۳	-----	۳-۷-۱-۲- توابع انتقال
۲۵	-----	۴-۷-۱-۲- آموزش شبکه‌های عصبی مصنوعی
۲۵	-----	۵-۷-۱-۲- الگوریتم آموزشی پس انتشار خطا
۲۹	-----	۸-۱-۲- ارزیابی مدل

فصل سوم : مدل سازی QSAR مشتقات گوانین به عنوان بازدارنده های CDK2 با

استفاده از روش رگرسیون خطی و شبکه عصبی مصنوعی

۳۳	-----	۱-۳- معرفی بازدارنده‌های CDK2
۳۴	-----	۲-۳- نرم افزار های استفاده شده
۳۴	-----	۱-۲-۳- نرم افزار HyperChem

۳۵	-----	Dragon نرم افزار ۲-۲-۳
۳۵	-----	SPSS نرم افزار ۳-۲-۳
۳۵	-----	MATLAB نرم افزار ۴-۲-۳
		-
۳۶	-----	انتخاب سری داده‌های استفاده شده در مدل‌سازی ۳-۳
۴۳	-----	محاسبه ی توصیف کننده‌ها ۴-۳
۴۵	-----	انتخاب توصیف کننده‌های مناسب ۵-۳
۴۶	-----	انتخاب مدل مناسب ۱-۵-۳
		-
۵۱	-----	مدل سازی به روش شبکه عصبی ۲-۵-۳
۵۲	-----	انتخاب اولیه ی مقدار وزن‌ها و بایاس ۱-۲-۵-۳
		-
۵۲	-----	بهینه سازی پارامترهای شبکه ۲-۲-۵-۳
		-
۵۳	-----	انتخاب تعداد متغیرهای ورودی شبکه، الگوریتم آموزشی، نوع تابع انتقال، تعداد گره لایه ی پنهان و تعداد دوره‌های آموزشی ۳-۲-۵-۳
۵۹	-----	بهینه سازی پارامتر μ ۴-۲-۵-۳
۶۰	-----	معماری شبکه‌ی عصبی بهینه شده ۳-۵-۳
		-
		فصل چهارم: بحث و نتیجه گیری
۶۲	-----	ارزیابی مدل‌های خطی و غیر خطی ۱-۴
۶۲	-----	ارزیابی مدل‌های خطی و غیر خطی با استفاده از سری‌های ارزیابی و تست ۱-۱-۴
۶۵	-----	ارزیابی مدل‌های خطی و غیر خطی با استفاده از روش رد مرحله‌ای تک‌تک ۲-۱-۴
۷۰	-----	ارزیابی مدل‌های ارائه شده با استفاده از پارامترهای آماری ۳-۱-۴
۷۱	-----	ارزیابی شبکه با استفاده از Y - تصادفی ۴-۱-۴
		-
۷۲	-----	بررسی ارتباط توصیف کننده‌ها ساختاری وارد شده در مدل با فعالیت بیولوژیکی ترکیبات مورد بررسی ۲-۴
		-
۷۷	-----	بررسی میزان مشارکت توصیف کننده‌ها در مدل شبکه‌ی عصبی ۳-۴

۴-۴- نتیجه‌گیری نهایی ----- ۷۸

فصل پنجم : مدل سازی QSPR شاخصهای بازداري خطی کروماتوگرافی مربوط به مونومتیل آلکان های C30 – C4 بر روی فاز ثابت متیل سلیکون 1 – OV با استفاده از روش رگرسیون خطی و شبکه عصبی مصنوعی

۱-۵- معرفی مونومتیل آلکانها ----- ۸۰

۲-۵- نرم افزارهای استفاده شده ----- ۸۱

۳-۵- انتخاب سری داده‌ها ----- ۸۱

۴-۵- ویژگیهای سیستم اندازه گیری اندیس بازداري سری داده‌ها ----- ۸۱

۵-۵- محاسبه‌ی توصیف کننده‌ها ----- ۸۲

۶-۵- انتخاب توصیف کننده های مناسب ----- ۸۳

۱-۶-۵- مدل سازی به روش شبکه عصبی ----- ۹۳

۱-۱-۶-۵- انتخاب اولیه‌ی مقدار وزن‌ها و بایاس‌ها ----- ۹۴

۲-۱-۶-۵- بهینه سازی پارامترهای شبکه ----- ۹۴

۳-۱-۶-۵- انتخاب تعداد ورودی‌های شبکه، الگوریتم آموزشی، نوع تابع انتقال، تعداد گره‌های لایه‌ی پنهان و
تعداد دوره‌های آموزشی ----- ۹۴

۴-۱-۶-۵- بهینه سازی پارامتر μ ----- ۱۰۰

۳-۶-۵- معماری شبکه‌ی عصبی بهینه شده ----- ۱۰۱

فصل ششم : بحث و نتیجه گیری

۱-۶- ارزیابی مدل‌های خطی و غیرخطی ----- ۱۰۳

۱-۱-۶- ارزیابی مدل‌های خطی و غیر خطی با استفاده از سری‌های ارزیابی و تست ----- ۱۰۳

۲-۱-۶- ارزیابی مدل‌های خطی و غیر خطی با استفاده از روش رد مرحله‌ای تک تک ----- ۱۰۸

۳-۱-۶- ارزیابی مدل‌های ارائه شده با استفاده از پارامترهای آماری ----- ۱۱۷

۴-۱-۶- ارزیابی شبکه به روش Y- تصادفی ----- ۱۱۸

۲-۶- بررسی ارتباط توصیف کننده های ساختاری وارد شده در مدل با اندیس بازداري ترکیبات مورد بررسی ----- ۱۱۹

۱۲۲	-----	۳-۶- بررسی میزان مشارکت توصیف‌کننده‌ها در مدل شبکه عصبی
۱۲۳	-----	۴-۶- نتیجه‌گیری نهایی
۱۲۳	-----	آینده‌نگری
۱۲۶	-----	فهرست منابع و مراجع

-

فهرست جداول

۲۴	-----	جدول ۱-۲- نام و تعریف انواع مختلف توابع در شبکه عصبی
۳۷	-----	جدول ۱-۳- ساختار، نام و فعالیت بیولوژیکی ترکیبات شرکت کننده در مدل سازی
۴۴	-----	جدول ۲-۳- هجده سری از توصیف‌کننده‌های مورد استفاده در کمومتریکس
۴۶	-----	جدول ۳-۳- کل مدل‌های منتخب از روش رگرسیون مرحله‌ای به همراه پارامترهای مربوطه
۴۸	-----	جدول ۴-۳- توصیف‌کننده‌های مدل منتخب به همراه کلاس مربوطه
۴۸	-----	جدول ۵-۳- ماتریس همبستگی توصیف‌کننده‌های انتخاب شده
۴۹	-----	جدول ۶-۳- ضریب رگرسیون و اثر متوسط توصیف‌کننده‌ها
۵۰	-----	جدول ۷-۳- مقادیر عددی توصیف‌کننده‌های وارد شده در مدل سازی
۵۴	-----	جدول ۸-۳- پارامترهای شبکه‌های بهینه بدست آمده براساس مقادیر میانگین مربعات خطا
		جدول ۹-۳- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی
۵۵	-----	لونبرگ - مارکوارت و تابع انتقال لگاریتم سیگموئید
		جدول ۱۰-۳- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی بایزین
۵۶	-----	و تابع تبدیل لگاریتم سیگموئید
		جدول ۱۱-۳- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونبرگ - مارکوارت
۵۷	-----	و تابع تبدیل تانژانت سیگموئید
		جدول ۱۲-۳- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین
۵۸	-----	و تابع تبدیل تانژانت سیگموئید
۵۹	-----	جدول ۱۳-۳- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ
		جدول ۱-۴- مقادیر تجربی و پیش بینی شده‌ی pIC_{50} برای ترکیبات سری ارزیابی با استفاده از روش‌های
۶۳	-----	خطی و غیرخطی
		جدول ۲-۴- مقادیر تجربی و پیش بینی شده‌ی pIC_{50} برای ترکیبات سری تست با استفاده از روش‌های
۶۴	-----	خطی و غیرخطی
		جدول ۳-۴- مقادیر تجربی، پیش بینی شده و خطای حاصله به روش رد مرحله‌ای تک تک با
۶۶	-----	مدل ANN و MLR
۷۰	-----	جدول ۴-۴- مقایسه‌ی آماره‌های بدست آمده از دو مدل ANN و MLR

- جدول ۴-۵- مقادیر مجذور ضریب همبستگی سری تست بعد از چندین Y - تصادفی ----- ۷۱
- جدول ۵-۱- هفت مدل اول حاصل از رگرسیون مرحله‌ای به همراه پارامترهای آماری مربوطه ----- ۸۴
- جدول ۵-۲- توصیف‌کننده‌های انتخاب شده توسط رگرسیون خطی مرحله‌ای ----- ۸۴
- جدول ۵-۳- ماتریس همبستگی توصیف‌کننده‌های انتخاب شده توسط رگرسیون خطی چندگانه
مرحله‌ای ----- ۸۵
- جدول ۵-۴- مقادیر R^2 و MSE مربوط به سری ارزیابی ----- ۸۶
- جدول ۵-۵- نام و معنای توصیف‌کننده به همراه اثر متوسط ----- ۸۶
- جدول ۵-۶- مقادیر عددی توصیف‌کننده‌ها و مقادیر تجربی اندیس بازداري برای ترکیبات مورد بررسی - ۸۷

- جدول ۵-۷- پارامترهای شبکه‌های بهینه بدست آمده بر اساس مقادیر میانگین مربعات خطا (MSE) ۹۵
- جدول ۵-۸- مقادیر میانگین مربعات خطای مربوط به سری ارزیابی برای شبکه با الگوریتم آموزشی لونیبرگ- مارکوارت و تابع انتقال لگاریتم سیگموئیدی ----- ۹۶
- جدول ۵-۹- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل لگاریتم سیگموئید ----- ۹۷
- جدول ۵-۱۰- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی لونیبرگ- مارکوارت و تابع تبدیل تانژانت سیگموئید ----- ۹۸
- جدول ۵-۱۱- مقادیر میانگین مربع خطاها برای شبکه‌هایی با الگوریتم آموزشی تنظیم بایزین و تابع تبدیل تانژانت سیگموئید ----- ۹۹
- جدول ۵-۱۲- میانگین مربع خطای شبکه‌ی بهینه در مقادیر متفاوت μ ----- ۱۰۰
- جدول ۶-۱- مقادیر تجربی و پیش بینی شده‌ی RI برای ترکیبات سری ارزیابی با استفاده از روش‌های خطی و غیر خطی ----- ۱۰۴
- جدول ۶-۲- مقادیر تجربی و پیش بینی شده‌ی اندیس بازداری برای ترکیبات سری تست با استفاده از روش‌های خطی و غیر خطی ----- ۱۰۶
- جدول ۶-۳- مقادیر تجربی، پیش بینی شده و خطای حاصله به روش رد مرحله‌ای تک تک با مدل ANN و MLR ----- ۱۰۹
- جدول ۶-۴- مقایسه‌ی آماره‌های بدست آمده از دو مدل ANN و MLR ----- ۱۱۷
- جدول ۶-۵- مقادیر مجذور ضریب همبستگی سری تست بعد از چندین Y - تصادفی ----- ۱۱۸
- جدول ۶-۶- اثر توصیف‌کننده‌ی MW بر اندیس بازداری ----- ۱۲۰

فهرست اشکال

- شکل ۲-۱- ساختار یک نرون بیولوژیکی ----- ۱۸
- شکل ۲-۲- مدل ریاضی یک نرون ----- ۱۹
- شکل ۲-۳- یک ساختار سه لایه ای از نرونها ----- ۲۱
- شکل ۲-۴- توابع تبدیل (الف) خطی، (ب) سیگموئیدی ----- ۲۴
- شکل ۳-۱- تغییرات R^2 بر حسب شماره‌ی مدل ----- ۴۷
- شکل ۳-۲- تغییرات SE بر حسب شماره‌ی مدل ----- ۴۷
- شکل ۳-۳- تغییرات F بر حسب شماره‌ی مدل ----- ۴۷
- شکل ۳-۴- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکوارت و تابع تبدیل لگاریتم سیگموئید ----- ۵۵
- شکل ۳-۵- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل لگاریتم سیگموئید ----- ۵۶
- شکل ۳-۶- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکوارت و تابع تبدیل تانژانت سیگموئید ----- ۵۷
- شکل ۳-۷- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل تانژانت سیگموئید ----- ۵۸
- شکل ۳-۸- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی ----- ۵۸

- ۵۹ ----- بر حسب پارامتر μ -----
- ۶۰ ----- شکل ۳-۹- طرح شماتیک ساختار شبکه عصبی بدست آمده پس از بهینه سازی -----
- ۶۳ ----- شکل ۴-۱- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل ANN -----
- ۶۳ ----- شکل ۴-۲- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل MLR -----
- ۶۴ ----- شکل ۴-۳- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل ANN -----
- ۶۴ ----- شکل ۴-۴- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل MLR -----
- شکل ۴-۵- نمودار مقادیر پیش‌بینی شده pIC_{50} با استفاده از شبکه‌ی عصبی مصنوعی بهینه شده
- ۶۸ ----- بر حسب مقادیر واقعی به روش رد مرحله ای تک تک -----
- شکل ۴-۶- نمودار مقادیر پیش‌بینی شده pIC_{50} با استفاده از روش خطی بر حسب مقادیر واقعی
- ۶۸ ----- به روش رد مرحله‌ای تک تک -----
- شکل ۴-۷- نمودار مقادیر خطای مطلق در ارزیابی مدل شبکه‌ی عصبی مصنوعی به روش
- ۶۹ ----- رد مرحله ای تک‌تک در مقابل مقادیر تجربی -----
- شکل ۴-۸- نمودار مقادیر خطای مطلق در ارزیابی مدل خطی به روش رد مرحله ای تک‌تک
- ۶۹ ----- در مقابل مقادیر تجربی -----
- شکل ۴-۹- نمودار درصد مشارکت توصیف‌کننده‌های بکار گرفته شده در مدل شبکه عصبی
- ۷۸ ----- برای ترکیبات CDK2 -----
- ۸۶ ----- شکل ۵-۱- تغییرات R^2 بر حسب تعداد توصیف‌کننده‌ها برای سری ارزیابی -----
- شکل ۵-۲- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکوارت و تابع تبدیل لگاریتم
- ۹۶ ----- سیگموئید -----
- ۹۷ ----- شکل ۵-۳- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل لگاریتم سیگموئید --
- شکل ۵-۴- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی لونیبرگ - مارکوارت و
- ۹۸ ----- تابع تبدیل تانژانت سیگموئید -----
- ۹۹ ----- شکل ۵-۵- نمایش سه بعدی شبکه‌هایی با الگوریتم آموزشی بایزین و تابع تبدیل تانژانت سیگموئید--
- شکل ۵-۶- نمودار میانگین مربع خطای حاصل از شبکه‌ی عصبی بهینه مربوط به سری ارزیابی
- ۱۰۰ ----- بر حسب پارامتر μ -----
- ۱۰۱ ----- شکل ۵-۷- تصویر شماتیک ساختار شبکه عصبی بدست آمده پس از بهینه سازی -----
- ۱۰۵ ----- شکل ۶-۱- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل ANN -----
- ۱۰۵ ----- شکل ۶-۲- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری ارزیابی با مدل MLR -----
- ۱۰۷ ----- شکل ۶-۳- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل ANN -----
- ۱۰۷ ----- شکل ۶-۴- نمودار مقادیر پیش‌بینی شده بر حسب مقادیر واقعی برای سری تست با مدل MLR -----
- شکل ۶-۵- نمودار مقادیر پیش‌بینی شده اندیس بازداري با استفاده از شبکه‌ی عصبی مصنوعی
- ۱۱۵ ----- بهینه شده بر حسب مقادیر واقعی به روش رد مرحله ای تک تک -----
- شکل ۶-۶- نمودار مقادیر پیش‌بینی شده اندیس بازداري به روش خطی بر حسب مقادیر واقعی
- ۱۱۵ ----- به روش رد مرحله ای تک تک -----
- شکل ۶-۷- نمودار مقادیر خطای مطلق در ارزیابی مدل شبکه‌ی عصبی مصنوعی به روش رد مرحله ای
- ۱۱۶ ----- تک‌تک در مقابل مقادیر تجربی -----

۱۱۶	شکل ۶-۸- نمودار مقادیر خطای مطلق در ارزیابی مدل به روش رد مرحله ای تک-تک در مقابل مقادیر تجربی -----
۱۲۲	شکل ۶-۹- نمودار درصد مشارکت توصیف کننده‌های بکار گرفته شده در مدل شبکه عصبی برای مونومتیل آلکان‌ها -----

فصل اول

مقدمه

۱-۱- تعریف تحقیق

همان‌طور که در فصل دوم بیشتر توضیح داده خواهد شد، کمومتریکس شاخه‌ای از شیمی است که ابزارهای ریاضی و آمار را برای بهبود بخشیدن به داده‌های شیمیایی و بدست آوردن اطلاعات بیشتر از آنها به کار می‌گیرد. نگرش چند متغیره به مسئله، ساخت و ارزیابی مدل‌هایی با قابلیت پیش‌بینی، مقایسه‌ی نتایج بدست آمده از روش‌های مختلف، قابلیت تعریف و استفاده از شاخص-هایی که قادرند کیفیت اطلاعات استخراج شده و مدل‌های بدست آمده را بسنجند از کاربردهای کمومتریکس می‌باشد.

از شاخه‌های مهم کمومتریکس، ارتباط کمی ساختار - فعالیت / ساختار - ویژگی^۱ بوده که در این پایان‌نامه هر دو بررسی شده است. در قسمت نخست پروژه، تکنیک ارتباط کمی ساختار - فعالیت برای پیش‌بینی فعالیت بیولوژیک تعدادی ترکیب بازدارنده و در قسمت دوم پیش‌بینی شاخص‌های بازدارندگی از مونو متیل آلکان‌ها بررسی شده است. جهت بررسی ارتباط کمی ساختار - فعالیت / ویژگی، روش‌های مورد استفاده به دو دسته‌ی روش‌های خطی و روش‌های غیرخطی تقسیم می‌شوند. از جمله روش‌های خطی که بدین منظور مورد استفاده قرار می‌گیرد می‌توان به روش‌های رگرسیون خطی چندگانه^۲، رگرسیون اجزای اصلی^۳ و روش حداقل مربعات جزئی^۴ اشاره کرد. از مهمترین روش‌های غیرخطی نیز می‌توان به شبکه‌های عصبی مصنوعی^۵ اشاره کرد. در تحقیق حاضر مطالعه‌ی ارتباط کمی ساختار - فعالیت / ویژگی با استفاده از روش‌های خطی و غیرخطی انجام گرفته و نتایج باهم مقایسه شده‌اند. در فصل‌های سوم و چهارم، رگرسیون خطی چندگانه به منظور ایجاد مدل مناسبی برای ایجاد ارتباط خطی بین فعالیت بیولوژیکی و ساختار ۵۶ ترکیب از نوع بازدارنده‌های CDK2 و سپس، شبکه عصبی مصنوعی به منظور بررسی روابط غیرخطی بین ساختار این ترکیبات و

^۱Quantitative Structure – Activity / Property Relationship (QSAR/QSPR)

^۲Multiple Linear Regression (MLR)

^۳Principal Component Regression

^۴Partial Least Square(PLS)

^۵Artificial Neural Network (ANN)

فعالیت بیولوژیکی به کار گرفته شده است. در فصل‌های پنجم و ششم نیز به کمک این روش‌ها ارتباط خطی و غیر خطی بین ساختار ۱۹۶ ترکیب از مونومتیل آلکان‌ها و شاخص بازدارنده آنها بررسی شده است.

۱-۱-۱- کاربرد QSAR در طراحی دارو

QSAR به راحتی می‌تواند برای کمک به بهینه‌سازی سنتز داروها مورد استفاده قرار گیرد. تحقیقات برای پیشنهاد داروهای جدید در روش‌های داروسازی سنتی تا حد زیادی بر پایه‌ی آزمون و خطا و استخراج موادی با خاصیت دارویی از طبیعت می‌باشد. پس از شناسایی اصول مولکولی حیات، دانشمندان به اطلاعاتی ارزشمند در مورد نحوه‌ی اثر مولکول‌های دارویی بر بدن دست یافتند. در این تلاش‌ها مشخص شد که اثر دارویی هر ترکیب به ویژگی‌های ساختاری آن ترکیب وابسته است. هر یک از این ویژگی‌های ساختاری باعث بروز یکی از آثار دارویی مربوط به آن می‌شود. در علم داروسازی مدرن با تکیه بر روش‌های QSAR، تولید دارو به یک روند تحقیقاتی کاملاً جهت‌دار تبدیل شده است. برای انجام یک مطالعه‌ی QSAR ابتدا باید مقادیر فعالیت دارویی ترکیبات از طریق تجربی بدست آیند و همچنین توصیف‌کننده‌های مناسب از طریق تجربی یا محاسبه توسط کامپیوتر تولید شوند، مرحله‌ی بعد دستیابی به ارتباط ریاضی بین ویژگی‌های ساختاری و مقادیر فعالیت است که معمولاً -ترین تکنیک چندمتغیری نیز رگرسیون خطی چندگانه (MLR) می‌باشد. با این حال این روش تنها زمانی به کار برده می‌شوند که تعداد مولکول‌های مورد بررسی بسیار بیشتر از تعداد توصیف‌کننده‌ها بوده و همچنین متغیرها با هم همبستگی زیادی نداشته باشند. استفاده از شبکه‌های عصبی مصنوعی به جای MLR، خصوصاً هنگامی که ارتباط بین توصیف‌کننده‌ها و فعالیت مورد نظر خطی نبوده و یا اینکه بین آنها برهمکنش‌هایی وجود داشته باشد، باعث بهبود مدل حاصله خواهد شد.

۱-۱-۲- شاخص بازداری

شاخص بازداری پارامتری کمی است که میزان بازداری نسبی هر جزء در نمونه را نسبت به آلکان‌های نرمال بر روی یک فاز ساکن معین و در یک دمای مشخص نشان می‌دهد. شاخص بازداری، متغیرهای مختلف سیستم‌های کروماتوگرافی را نرمالیزه کرده، به گونه‌ای که امکان مقایسه‌ی نتایج سیستم‌های مختلف فراهم می‌شود. مقدار شاخص بازداری از معادله‌ی (۱-۱) محاسبه می‌شود:

$$I = 100Y(Z - Y) \left[\frac{\log t_r' X - \log t_r' Y}{\log t_r' Z - \log t_r' Y} \right] \quad \text{معادله‌ی ۱-۱}$$

در معادله‌ی بالا $t_r' Y$ زمان بازداری اصلاح شده برای آلکان نرمال که قبل از ترکیب X از ستون خارج می‌شود، $t_r' X$ زمان بازداری اصلاح شده برای ترکیب مورد نظر، $t_r' Z$ زمان بازداری اصلاح شده برای آلکان نرمال که بعد از ترکیب X از ستون خارج می‌شود، Y تعداد کربن‌های آلکان نرمال که قبل از ترکیب مورد نظر از ستون خارج می‌شود. Z نیز تعداد کربن‌های آلکان نرمال که بعد از ترکیب مورد نظر از ستون خارج می‌شود، می‌باشد.

۱-۲- ضرورت تحقیق

پیشرفت علوم در زمینه‌های مختلف روش‌های آزمایشگاهی کمک شایانی به بهبود نتایج حاصل از این روش‌ها کرده است. با وجود این، مشکلات و محدودیت‌هایی از جمله وجود عدم قطعیت در نتایج آزمایشگاهی، هزینه‌ی بالا، در دسترس نبودن امکانات و عدم توانایی در اندازه‌گیری همه‌ی ترکیبات باعث کشف و توسعه‌ی روش‌های نظری شده است.

در این تحقیق سعی شده است با به‌کارگیری روش‌های QSAR/QSPR مدلی برای پیش‌بینی فعالیت بیولوژیکی در بخش اول تحقیق و پیش‌بینی شاخص بازداری در بخش دوم تحقیق ارائه شود.

۳-۱- پیشینه‌ی تحقیق

در سال ۲۰۰۵ فرناندز و کامبا^۱ بازدارنده‌های CDK2 را با استفاده از شبکه‌ی عصبی مصنوعی مدل‌سازی نمودند که در آن دقت مدل برای سری آموزش ۸۷٪ بود [۱].

سوما سامانتا^۲ در سال ۲۰۰۶، در مطالعه‌ی QSAR ترکیبات ۳-آمینوپیرازول‌ها به عنوان ترکیبات ضد تومور برای فعالیت بازدارندگی CDK2 بر مبنای پارامترهای شاخص^۳ و بارهای اتمی برای بررسی تاثیر کمی تفاوت‌های ساختاری روی فعالیت ضد تومور به این نتیجه رسیدند که اتم شماره‌ی ۵ برای فعالیت این گروه از ترکیبات نقش مهمی دارد. همچنین مطالعات نشان داد که با افزایش خاصیت الکترونگاتیویته‌ی اکسیژن در موقعیت ۸، فعالیت افزایش می‌یابد [۲].

در سال ۲۰۰۷ لی جی و لیو^۴ ارتباط ساختار- فعالیت بازدارنده‌های اکسیندول CDK را با استفاده از روش حداقل مربعات برداری (LSV)^۵ مطالعه کردند [۳].

کاتریتزکی^۶ و همکارانش در سال ۲۰۰۰، مدل‌های QSPR شامل ۱۷۸ ترکیب مونومتیل، دی-متیل، تری‌متیل و تترا متیل آلکان‌ها را توسط توصیف‌کننده‌های کوانتومی و توپولوژیکی برای پیش-بینی شاخص بازدارندگی آنها ارائه کردند که انحراف استاندارد میانگین مدل ۴/۶ بود [۴].

دوو و لیانگ^۷ در سال ۲۰۰۳ با بررسی ارتباط کمی بین ساختار مولکولی و شاخص‌های بازدارندگی ایزومتریکی ۱۳۴ آلکان متفاوت، خطای استاندارد $\pm 3im$ را بدست آوردند [۵].

در سال ۲۰۰۵ ژیانگ، لیانگ و هوکیو^۸ با مدل‌سازی QSPR مربوط به ۲۰۷ ترکیب از متیل آلکان‌ها برای پیش‌بینی اندیس بازدارندگی با بکارگیری روش‌های ژنتیک الگوریتم (GA)^۱ و حداقل

^۱Fernandez and Camba^۲Soma Samanta^۳Indicatore parameters^۴Li. J. and Liu^۵Least Square Vortex^۶Katritzky^۷DU and Liang^۸Index Unit^۹Ziang, Liang and Hu Q