



## مطالعه ارتباط کمی ساختار - فعالیت جهت پیش بینی فعالیت دارویی برخی از ترکیبات پیریدین در درمان سرطان پوست با استفاده از الگوریتم کرم شب تاب

فاطمه شمس، مهدی نکویی\*

دانشگاه آزاد اسلامی، واحد شاهرود، دانشکده علوم پایه، گروه شیمی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۳۹۷/۰۲/۱۰، تاریخ دریافت نسخه اصلاح شده: ۱۳۹۷/۰۳/۲۲، تاریخ پذیرش قطعی: ۱۳۹۷/۰۴/۱۲

### چکیده

مطالعه ارتباط کمی-ساختار فعالیت (QSAR) جهت پیش بینی فعالیت دارویی برخی از ترکیبات پیریدین و مشتقات آن در درمان سرطان پوست انجام شد. در ابتدا ساختار ترکیبات مورد نظر توسط نرم افزار هایپرکم رسم و بهینه گردید. سپس تعداد ۱۴۸۱ توصیف کننده توسط نرم افزار دراگون محاسبه شد. برای انتخاب مناسب ترین توصیف کننده ها از روش رگرسیون مرحله ای و الگوریتم کرم شب تاب استفاده شد. پس از انتخاب مناسبترین توصیف کننده ها توسط این دو روش، جهت مدلسازی و پیش بینی فعالیت دارویی ترکیبات مورد نظر از روش رگرسیون خطی چندگانه (MLR) استفاده شد. عملکرد هر مدل توسط سری تست مورد بررسی قرار گرفت. با استفاده از روش های آماری، ریشه میانگین مربعات خطا سری پیش بینی (RMSEP)، ضریب تعیین ( $R^2$ ) و خطای استاندارد پیش بینی (SEP) در روشهای SW-MLR و FF-MLR بدست آمد. نتایج برتری نسبی روش انتخاب متغیر الگوریتم کرم شب تاب را نسبت به رگرسیون مرحله ای جهت پیش بینی فعالیت دارویی ترکیبات ضد سرطان پوست نشان می دهد.

واژه های کلیدی: ارتباط کمی ساختار- فعالیت، ملانوما (سرطان پوست)، الگوریتم کرم شب تاب، رگرسیون خطی چندگانه

### ۱. مقدمه

سرطان پوست یکی از مشکلات پوستی است که متأسفانه در سالهای اخیر روند رو به گسترشی داشته است. در سرطان پوست سلولهای بدخیم که از بافت پوست نشأت می گیرند به طور نامنظم و فزاینده ای تکثیر و به طریقی از سیستم ایمنی و دفاعی بدن عبور می کنند، بدون این که موجب عکس العمل تدافعی و تهاجمی در سیستم ایمنی بدن شوند. مهم ترین عامل ایجاد سرطان پوست، جهش

\*عهده دار مکاتبات: مهدی نکویی

نشانی: گروه شیمی، دانشکده علوم پایه، دانشگاه آزاد اسلامی، شاهرود، ایران

پست الکترونیک: E-mail: m\_nekoei1356@yahoo.com

تلفن: ۰۲۳۳۲۳۹۴۳۲۰۰

در سلول بر اثر اشعه ماورای بنفش خورشید است با توجه به نوع سلول پوستی که دچار جهش ژنی شده است اشکال مختلفی از سرطان پوست به وجود می‌آید که هر کدام نشانه‌ها و عوارض خاص خود را دارند.



شکل ۱. تصاویری از سرطان پوست

ملانوما اجدی‌ترین نوع سرطان پوست می‌باشد که از سلول‌هایی در پوست بنام ملانوسیت منشاء می‌گیرد و به سایر نواحی بدن و عمدتاً گره‌های لنفاوی، کبد، ریه‌ها و دستگاه عصبی مرکزی گسترش می‌یابد. این سرطان از بدخیم‌ترین سرطان‌های پوست محسوب می‌شود. شیوع ملانوما در دهه‌های اخیر رو به افزایش بوده، به طوری که امروزه به عنوان یک سرطان شایع مطرح است. درصد ابتلا به ملانوما در طی سی سال گذشته حدوداً دو برابر شده است. علت اصلی ملانوم ناشناخته است اما عوامل متعددی از جمله عوامل ژنتیکی، تابش اشعه ماورای بنفش و تماس‌های محیطی در ایجاد این بیماری دخیل است. اهمیت این سرطان بیش از یک دهه ذهن دانشمندان را در جهت یافتن داروهای موثر و کارآمد جهت مقابله با این بیماری به خود معطوف نموده است [۱]. در گذشته روند کشف و توسعه داروهای جدید مبتنی بر روش آزمون و خطا بوده که این امر بسیار وقت گیر، طاقت فرسا و هزینه‌بر می‌باشد. مشکل دیگری که در این راه دانشمندان را آزار می‌دهد، عدم اطلاع آن‌ها از فعالیت داروئی ترکیبات، قبل از سنتز و بررسی تجربی آن‌ها بوده و به همین دلیل یکی از مهم‌ترین اهداف شیمیدان‌ها و محققان دارویی ارزیابی مقدماتی فعالیت داروها، قبل از ساخت آن‌ها می‌باشد. از این رو نیاز به استفاده از روش‌های تئوری و محاسباتی که بدون انجام آزمایش بتوانند ویژگی و یا فعالیت ترکیبات دارویی را پیش بینی کنند اجتناب‌ناپذیر به نظر می‌رسد. یکی از مهمترین این روش‌ها، مطالعه ارتباط کمی ساختار-فعالیت<sup>۲</sup> (QSAR) می‌باشد که به بررسی نحوه ارتباط بین فعالیت ترکیبات دارویی با مشخصات ساختاری و ذاتی آن‌ها می‌پردازد. بررسی ساختار شیمیایی و فعالیت ترکیبات، پیش بینی فعالیت ترکیبات جدید را بر اساس اطلاعات مرتبط به ساختار شیمیایی آن‌ها امکان‌پذیر می‌سازد [۲-۵]. در سالیان اخیر، روند رو به افزایش مقالات منتشره در منابع علمی بر اساس QSAR، دلالت بر جایگاه منحصر به فرد این دیدگاه در شیمی نظری و تعمیق بیش دانشمندان در فهم و توجیه سازوکارهای دخیل در فعالیت گسترده‌ی وسیعی از ترکیبات دارد [۶-۱۶].

<sup>1</sup> -Melanoma

<sup>2</sup> - Quantitative Structure-Activity Relationship (QSAR)

هدف از مطالعه حاضر، ارائه‌ی مدل‌های مناسب جهت پیش‌بینی فعالیت دارویی برخی از مشتقات پیریدین در درمان سلولهای سرطانی پوست با استفاده از روش‌های SW-MLR و FFA-MLR می‌باشد.

## ۲. روش‌های محاسباتی

### ۲-۱. انتخاب سری داده‌ها

در این پژوهش، سری داده‌ها شامل فعالیت دارویی ۳۰ ترکیب از مشتقات پیریدین به عنوان بازدارنده برای درمان سرطان پوست استفاده گردید [۱۷-۱۸]. قدرت بازدارندگی این ترکیبات به صورت  $IC_{50}$  گزارش شده است.  $IC_{50}$  عبارتست از مینیمم غلظتی از ترکیب دارویی که باعث ۵۰٪ اثر بازدارندگی بر روی بیماری می‌شود. این مقادیر به مقیاس لگاریتمی ( $pIC_{50}$ ) تبدیل و مورد استفاده قرار گرفت. در این کار پس از تقسیم تصادفی ترکیبات به دو گروه سری آموزش (۲۴ مولکول) و سری پیش‌بینی (۶ مولکول)، مقادیر  $pIC_{50}$  به عنوان متغیر وابسته و توصیف‌کننده‌ها به عنوان متغیر مستقل انتخاب شد. سری آموزش جهت ایجاد یک مدل مناسب و سری پیش‌بینی جهت ارزیابی مدل مورد استفاده قرار گرفت.

### ۲-۲. محاسبه و کاهش تعداد توصیف‌کننده‌ها

در ابتدا برای محاسبه توصیف‌کننده‌ها، ساختار ترکیبات به کمک نرم افزار هایپرکم<sup>۱</sup> رسم شد. سپس ساختارهای مولکولی ترسیم شده به وسیله الگوریتم AM1 بهینه شدند. ساختارهای بهینه شده به نرم افزار دراگون<sup>۲</sup> منتقل و توصیف‌کننده‌ها به تعداد ۱۴۸۱ مورد به کمک این نرم افزار محاسبه گردیدند. این نرم افزار برای محاسبه هیجده دسته از توصیف‌کننده‌های مولکولی مورد استفاده قرار می‌گیرد. یکی از مشکلاتی که در هنگام ایجاد مدل‌های QSAR با آن مواجه می‌شویم، تعداد زیاد متغیرهای مستقل می‌باشد. در اغلب موارد تعداد توصیف‌کننده‌ها از تعداد مولکول‌ها بسیار بیشتر است. در این صورت استفاده از روش‌های حداقل مربعات باعث ایجاد مشکلاتی نظیر انتخاب شانس و همبستگی تصادفی می‌گردد. با توجه به این که بعضی از متغیرهای مستقل، ثابت بوده و همچنین برخی دیگر با یکدیگر همبستگی<sup>۳</sup> نشان می‌دهند، لذا به روش زیر برخی از متغیرها حذف شدند.

۱- توصیف‌کننده‌هایی که مقادیر ثابت و یا تقریباً ثابت دارند (بیش از ۹۰٪ داده‌های ثابت دارند)، حذف شدند. در این مرحله تعداد ۳۴۴ توصیف‌کننده حذف و بدین ترتیب ۱۱۳۷ توصیف‌کننده باقی ماند.

۲- توصیف‌کننده‌هایی که با یکدیگر همبستگی بالای ۰/۹ دارند مورد بررسی قرار گرفتند و بین آن‌ها، توصیف‌کننده‌ای که همبستگی کمتری با متغیر مستقل داشت حذف گردید. بدین ترتیب تعداد ۷۹۲ توصیف‌کننده، حذف و تعداد ۳۴۵ توصیف‌کننده باقی ماند.

### ۲-۳. انتخاب توصیف‌کننده‌های مؤثر

<sup>۱</sup>- Hyperchem

<sup>۲</sup>- Dragon

<sup>۳</sup>- Correlation

مهم‌ترین بخش در ایجاد یک مدل کارآمد، انتخاب توصیف‌کننده‌های مناسب است. پس از محاسبه توصیف‌کننده‌های مختلف، تعدادی از آنها به عنوان توصیف‌کننده‌های مناسب برای ساخت مدل انتخاب می‌شوند. این مرحله شامل یافتن توصیف‌کننده‌های حاوی اطلاعات مفید است به طوری که قدرت پیش‌بینی مدل در سطح قابل قبولی باشد. در این پژوهش از دو روش انتخاب مرحله‌ای<sup>۱</sup> (SW) و الگوریتم کرم شب تاب<sup>۲</sup> (FFA) برای انتخاب مناسب‌ترین توصیف‌کننده‌ها استفاده شد. در روش مرحله‌ای با بررسی همه توصیف‌کننده‌ها، فرآیند انتخاب تا زمانی ادامه می‌یابد که مدلی با ضریب همبستگی بالا (معمولاً در حدود ۰/۷ تا ۰/۹) به دست آید. اگر ضریب همبستگی به مقدار بالای ۰/۹۵ برسد، می‌توان گفت که مدل خطی، مدل مناسبی برای توصیف سیستم مورد بررسی است و با استفاده از این مدل می‌توان پیش‌بینی را به نحو مطلوب انجام داد. با روش مرحله‌ای، از بین ۳۴۵ توصیف‌کننده باقی مانده تعداد ۵ توصیف‌کننده به عنوان مناسب‌ترین آنها انتخاب و توسط روش MLR مدل‌سازی انجام شد. لیست توصیف‌کننده‌های انتخاب شده توسط رگرسیون خطی چندگانه مرحله‌ای (SW-MLR) به همراه توصیف مختصری از آنها در جدول ۱ آورده شده است.

جدول ۱. توصیفگرهای انتخاب شده توسط رگرسیون خطی چندگانه مرحله‌ای (SW-MLR)

Descriptor symbol	Meaning	Type	MF <sup>a</sup>
D/DR05	Distance / Detour ring index of order 5.	topological descriptors.	۱/۵۵۳
RDF090u	Radial Distribution Function -9.0 /Unweighted.	RDF descriptors.	۱/۵۲۶
RDF080m	Radial Distribution Function -8.0 / Weighted by atomic masses.	RDF descriptors.	۱/۴۵۷
HATS8e	Leverage-weighted autocorrelation of lag 8 / weighted by atomic sanderson electronegativities.	GETAWAY descriptors.	۱/۱۴۲
H-050	H attached to heteroatom.	atom -centred fragments.	۱/۳۰۱

<sup>a</sup> اثر متوسط هر توصیف‌کننده

## ۲-۴. ارزیابی توصیف‌کننده‌های انتخاب شده

به منظور ارزیابی توصیف‌کننده‌های انتخاب شده مبنی بر مستقل بودن از یکدیگر، ضرایب همبستگی آنها نسبت به یکدیگر در جدول ۲ آورده شده است. همانطور که در این جدول مشاهده می‌شود ضریب همبستگی بین توصیفگرهای انتخاب شده همگی

<sup>1</sup> - Stepwise

<sup>2</sup> - Firefly Algorithm (FFA)

کمتر از ۰/۴۶۰۹ می‌باشد. لذا نتایج جدول نشان می‌دهد که بین توصیفگرهای انتخاب شده همبستگی چندانی وجود نداشته و توصیفگرها تقریباً مستقل از هم هستند.

جدول ۲. ماتریس ضرایب همبستگی توصیف‌کننده‌های انتخاب شده

	D/DR05	RDF090u	RDF080m	HATS8e	H-050
D/DR05	۱	۰	۰	۰	۰
RDF090u	۰/۴۶۰۹	۱	۰	۰	۰
RDF080m	۰/۴۱۲۰	۰/۴۴۸۰	۱	۰	۰
HATS8e	۰/۰۸۰۰	۰/۲۲۹۰	۰/۲۵۲۵	۱	۰
H-050	-۰/۴۳۷۰	-۰/۳۰۹۵	-۰/۱۲۵۴	۰/۰۸۶۷	۱

## ۲-۵. ایجاد مدل با استفاده از روش رگرسیون خطی چندگانه (MLR)

پس از انتخاب مناسب‌ترین توصیف‌کننده‌ها توسط روش افزایش مرحله‌ای، برقراری ارتباط مناسب بین توصیف‌کننده‌های منتخب و فعالیت ترکیبات دارویی صورت گرفت. معادله‌ی خطی ساخته شده جهت پیش‌بینی فعالیت‌های بازدارندگی مشتقات پیریدین به صورت زیر است. بین توصیف‌کننده‌ها و فعالیت‌های بازدارندگی مشتقات پیریدین برای سری آموزش با استفاده از روش MLR رابطه زیر به عنوان مدل خطی بدست آمد:

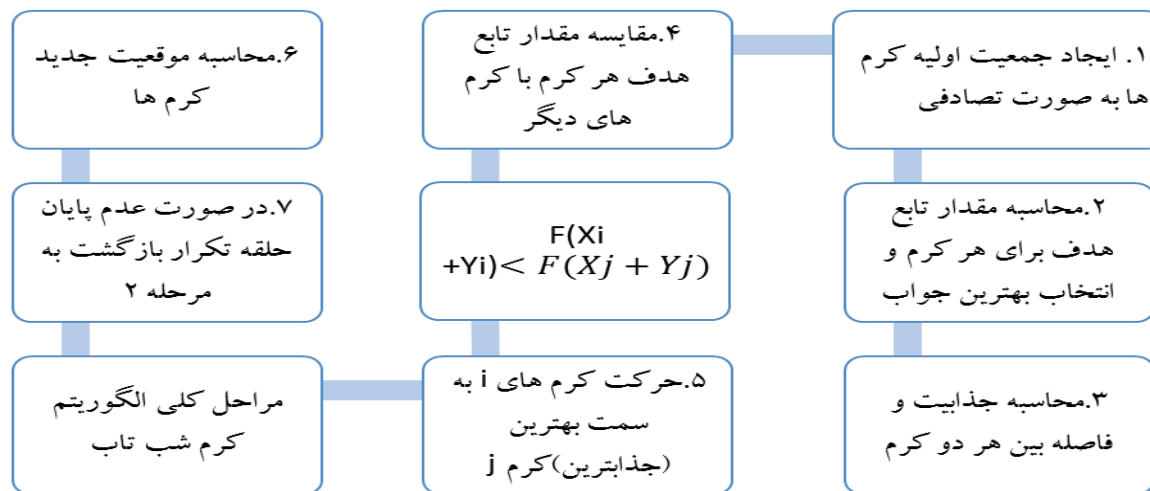
$$pIC_{50} = 3.8092 + 0.0117 (D/DR05) - 0.1117 (RDF090u) - 0.082 (RDF080m) + 5.0166 (HATS8e) + 1.0896 (H-050)$$

سپس از معادله بدست آمده برای پیش‌بینی فعالیت سری تست استفاده گردید. مقادیر واقعی و پیش‌بینی شده فعالیت‌ها برای کلیه ترکیبات مجموعه آموزش و تست در جدول ۴ آورده شده است.

## ۲-۶. انتخاب توصیفگرها به روش الگوریتم کرم شب تاب

هدف از اجرای الگوریتم کرم شب تاب (FFA) شناسایی و انتخاب ترکیبی از توصیفگرهاست، به گونه‌ای که برای پیش‌بینی  $pIC_{50}$  مشتقات پیریدین بوسیله رگرسیون خطی چندگانه مورد استفاده قرار گیرند. در الگوریتم کرم شب تاب، ابتدا تعدادی کرم شب تاب مصنوعی به طور تصادفی در دامنه مسئله پخش می‌شود. سپس به هریک از این کرم‌های شب تاب مصنوعی با استفاده از مقدار به دست آمده برای تابع هدف در آن نقطه، یک شدت نور مناسب نسبت داده می‌شود. نحوه انتساب شدت نور به هریک از کرم‌های شب تاب مصنوعی به گونه‌ای است که با افزایش میزان بهینگی نقطه‌ای که هر کرم شب تاب در آن واقع شده، شدت نور ساطع شده از آن افزایش می‌یابد. بدیهیست که بسته به میزان فاصله، این کرم‌های شب تاب یکدیگر را با نورهایی با شدت‌های مختلف مشاهده خواهند کرد. در مرحله بعد کرم‌های شب تاب کم نورتر مرتباً به سوی کرم‌های شب تاب پر نورتر جذب می‌شوند و

این فرآیند تکراری تا آنجا ادامه می‌یابد که همه کرم‌های شب‌تاب در یک نقطه که احتمالاً نقطه بهینه سراسری است تجمع یابند و یا اینکه تعداد از پیش تعیین شده‌ای از تکرارها انجام شود [۱۹-۲۰].



شکل ۲. مراحل کلی و شماتیک اجرای الگوریتم کرم شب‌تاب

جدول ۳. توصیفگرهای انتخاب شده توسط الگوریتم کرم شب‌تاب

۲-۷. مدل رگرسیون خطی چندگانه با استفاده از توصیفگرهای انتخاب شده توسط الگوریتم کرم شب‌تاب

Descriptor symbol	Meaning	Type	MF <sup>a</sup>
CIC3	Complementary information content (neighborhood symmetry of 3-order).	topological descriptors.	۱/۴۳۳
C-028	R—CR—X	atom-centred fragments.	-۰/۴۲۰
MS	mean electrotopological descriptors.	constitutional descriptors.	-۰/۱۵۶
R7m	R autocorrelation of lag 7 / weighted by atomic masses.	GETAWAY descriptors .	۰/۰۹۸
HOMA	Harmonic Oscillator model of Aromaticity index .	aromaticity indices.	۰/۰۵۰

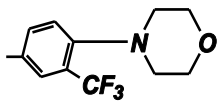
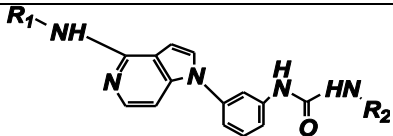
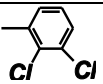
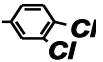
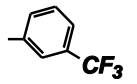
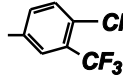
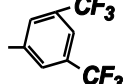
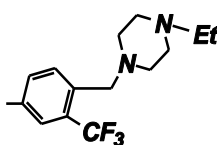
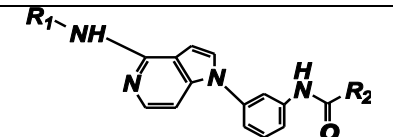
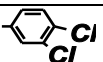
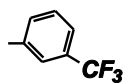
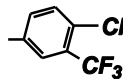
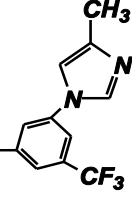
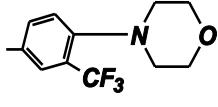
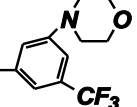
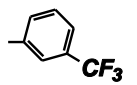
پس از انتخاب مناسب‌ترین توصیفگرها توسط الگوریتم کرم شب‌تاب، مدل سازی بین توصیفگرها و فعالیت‌های بازدارندگی مشتقات پیریدین برای سری آموزش با استفاده از روش MLR انجام و مدل خطی زیر بدست آمد:

$$pIC_{50} = -0.7263 + 0.8880 (CIC3) - 0.0854 (C-028) + 0.7184 (MS) - 0.8208 (R7m) + 9.0709 (HOMA)$$

سپس از معادله بدست آمده برای پیش‌بینی فعالیت سری تست استفاده گردید. مقادیر واقعی و پیش‌بینی شده فعالیت‌ها برای تمام ترکیبات در جدول ۴ آورده شده است.

جدول ۴. مقادیر تجربی و محاسبه شده مقادیر  $pIC_{50}$  برای ترکیبات مختلف برای مجموعه‌های آموزشی و پیش‌بینی در مدل‌های FFA و MLR

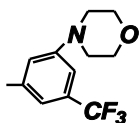
No.	R1	R2	R3	pIC <sub>50</sub>	SW-MLR	FFA-MLR
1	H		PhCO	۳/۸۵	۴/۳۸۱۵	۵/۵۷۸۱
2*	H		PhCO	۶/۵۴	۵/۷۶۲۶	۶/۰۴۰۷
3	H		PhCO	۴/۲	۴/۲۹۳۹	۵/۲۳۱۸
4	H		PhCO	۵/۸۵	۶/۱۶۸۳	۵/۹۷۷۹
5	H		H	۴/۲۱	۴/۲۹۸۹	۵/۳۳۴۵
6	H		H	۴/۳۸	۴/۵۵۳۶	۵/۳۳۳۱
7*	H		H	۵/۴۸	۴/۷۹۵	۵/۹۷۱۶
8*	H		PhCO	۵	۵/۵۲۶۴	۶/۰۵۷
9*	H		PhCO	۴/۹۴	۴/۳۵۹	۵/۲۰۷۲
10	H		PhCO	۵/۳	۵/۴۵۵۸	۵/۱۵۲۷
11	H		PhCO	۵/۴۸	۵/۹۲۷۷	۵/۲۷۶۵
12	H		PhCO	۵/۴۸	۴/۷۹۵	۵/۲۷۶۵
13	CH <sub>3</sub>		PhCO	۵/۴۸	۵/۳۳۴۴	۵/۲۷۶۵
14	H		PhCO	۵/۶	۵/۴۴۵۷	۵/۵۷۴۱
15	H		H	۵/۷۸	۶/۲۸۱۶	۶/۲۳۴۸

16	H		H	۶	۶/۱۱۲۱	۵/۱۴۹۳
						
17	PhCO		-	۶/۲۶	۶/۳۹۶۶	۶/۱۵۷
18	PhCO		-	۶/۳۸	۶/۷۹۸	۶/۰۹۶۱
19	PhCO		-	۶/۴۳	۵/۹۱۷۴	۵/۰۷۶۵
20	PhCO		-	۶/۴۳	۶/۲۰۶۸	۵/۰۷۶۵
21*	PhCO		-	۶/۹	۷/۵۳۹۱	۷/۱۴۱۳
22*	PhCO		-	۵	۴/۳۷۷۱	۵/۳۱۲۳
						
23	PhCO		-	۶/۷۳	۷/۱۳۷۲	۷/۱۳۲۷
24	PhCO		-	۶/۷۹	۶/۴۷۴۳	۶/۴۳۳۹
25	PhCO		-	۶/۸۵	۶/۲۵۲۸	۶/۰۸۶۸
26	PhCO		-	۷/۱۸	۷/۰۵۹۷	۶/۳۸۵۹
27	PhCO		-	۴/۱۸	۴/۶۰۴۳	۵/۵۱۸۴۴
28	PhCO		-	۴/۹۸	۴/۵۲۵۳	۵/۶۵۲۹
29	H		-	۵/۳	۵/۳۹۷۴	۵/۱۵۲۷



30

H



۵/۹

۵/۷۸۰۸

۵/۲۷۸۱

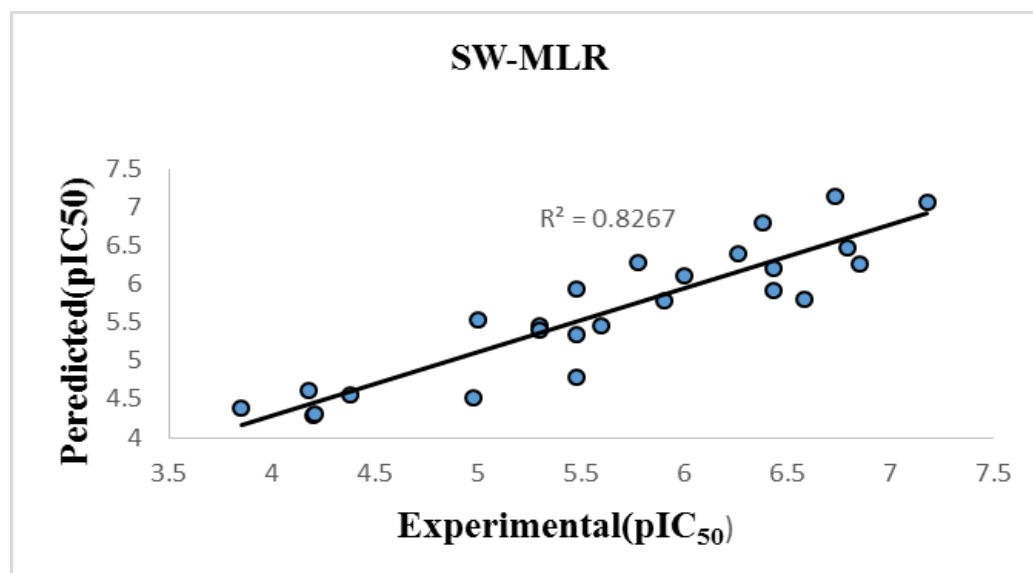
\*ترکیباتی که دارای ستاره می‌باشند در مجموعه تست و ترکیبات بدون ستاره در مجموعه آموزش قرار دارند.

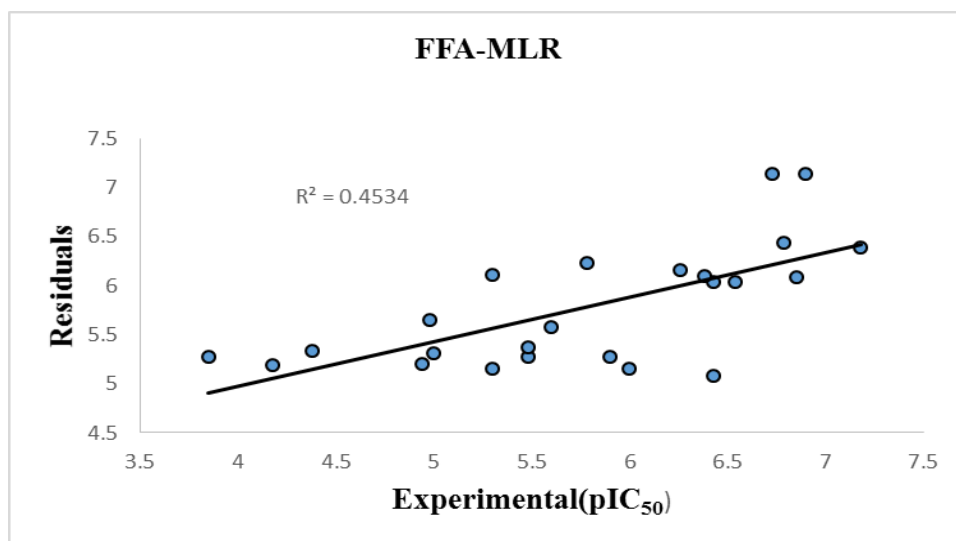
### ۳. بررسی نتایج

اعتبار و اهمیت معادلات پیش بینی و مدل های ساخته شده وقتی مشخص می گردد که فعالیت مولکول هایی که در سری تست هستند را به خوبی و بطور رضایت بخش و قابل قبول پیش بینی کند. در اینجا چندین روش به منظور ارزیابی توانایی مدل های ارائه شده در پیش بینی مقادیر  $pIC_{50}$  ذکر شده است.

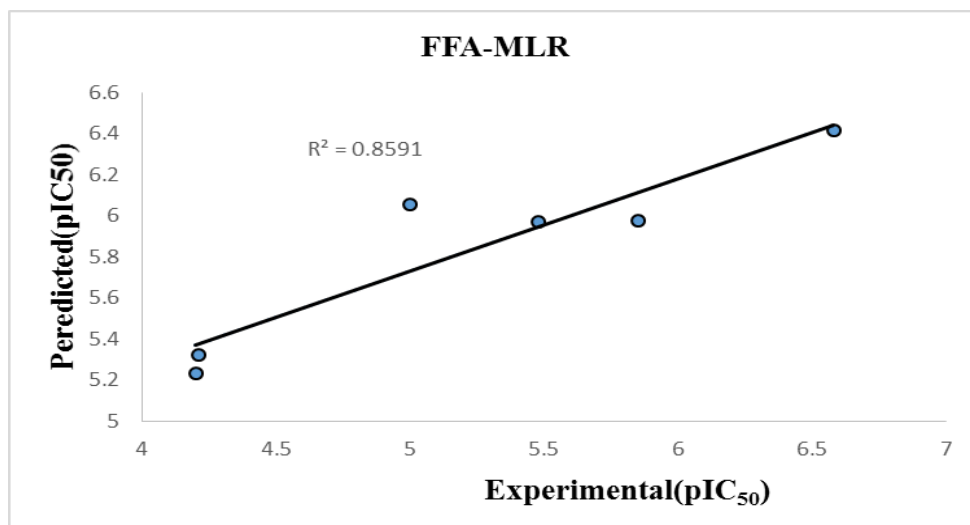
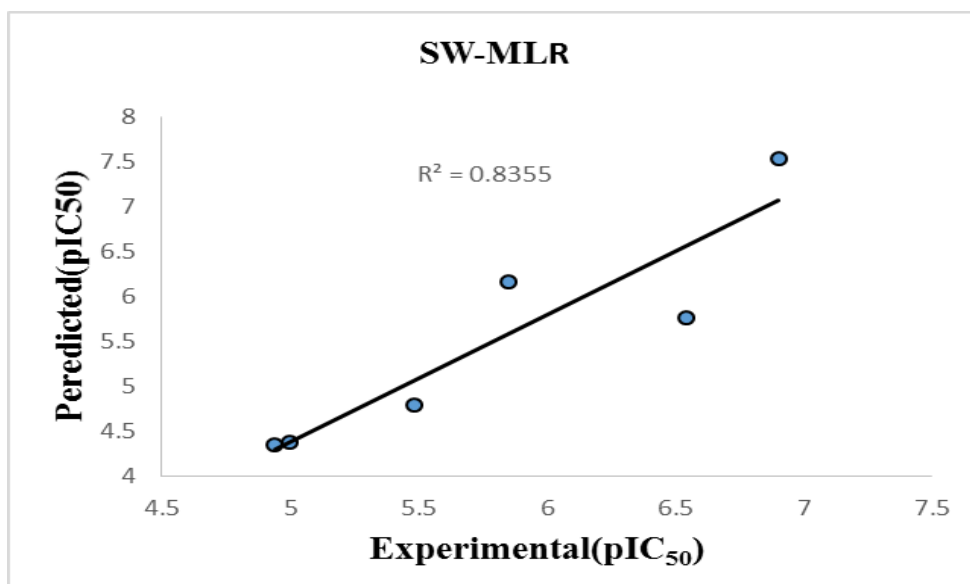
#### ۳-۱. ارزیابی مدل ها با استفاده از نمودار برگشتی

در نمودار برگشتی مقادیر پیش بینی شده برحسب مقادیر تجربی رسم می گردد و با توجه به مقدار ضریب تعیین ( $R^2$ ) به دست آمده از نمودار، پراکنندگی نقاط در اطراف خط برگشت تعیین می شود. هر چه مقدار ضریب تعیین به یک نزدیک تر باشد، مدل ساخته شده، مدل بهتری است. نتایج حاصل از نمودارهای برگشتی مربوط به دو روش انجام شده در این کار برای سری های آموزش و تست در شکل های (۳) و (۴) آورده شده است.





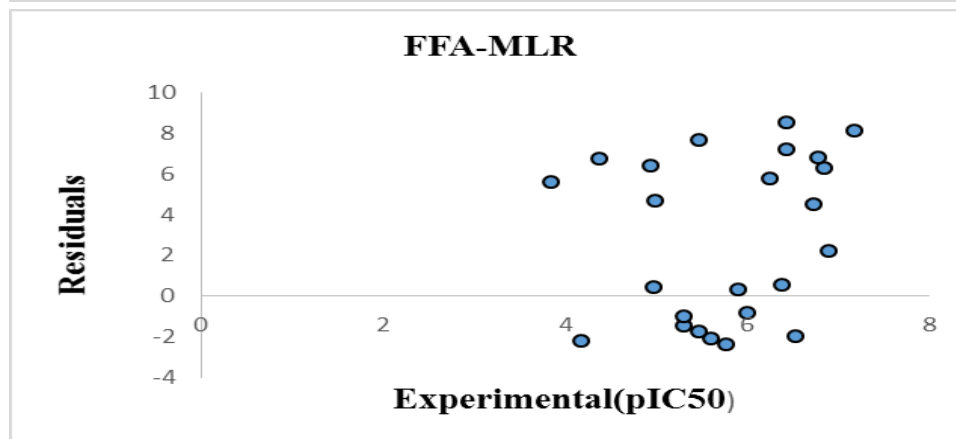
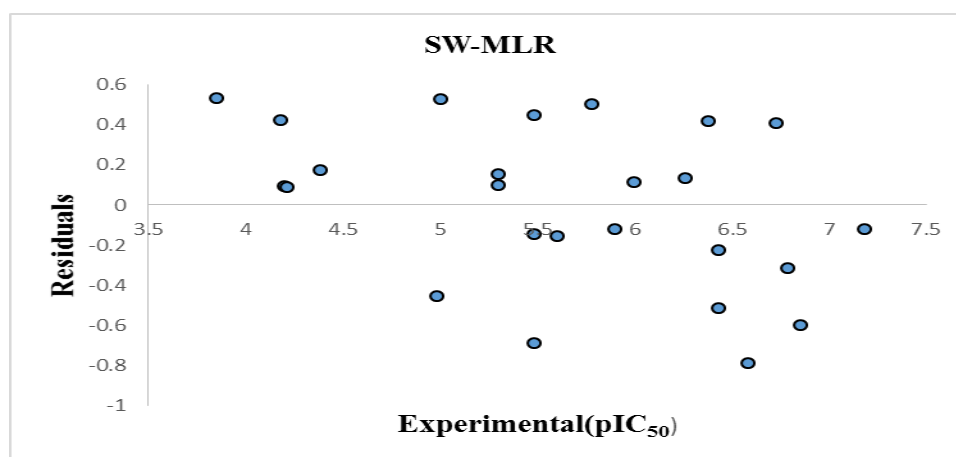
شکل ۳. نمودار مقادیر پیش بینی شده  $pIC_{50}$  بر حسب مقادیر تجربی برای سری آموزش



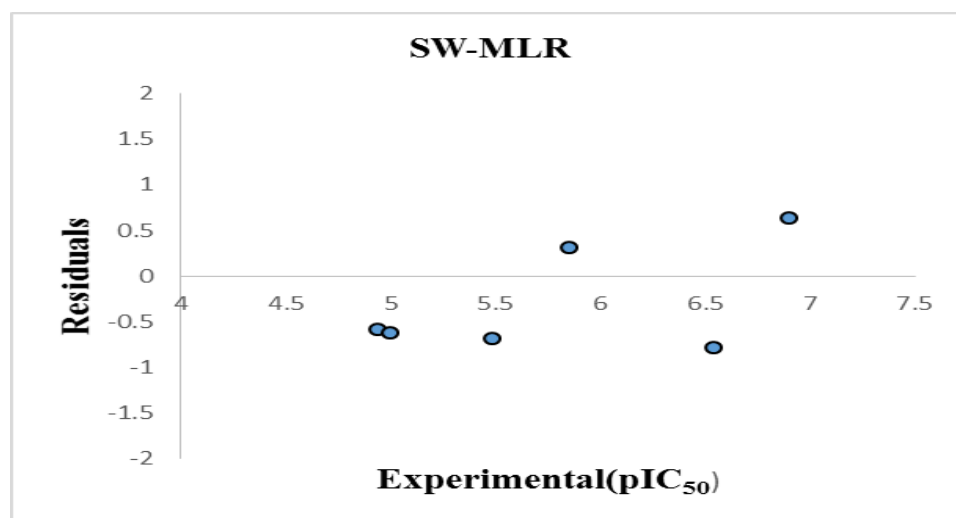
شکل ۴. نمودار مقادیر پیش بینی شده  $pIC_{50}$  بر حسب مقادیر تجربی برای سری تست

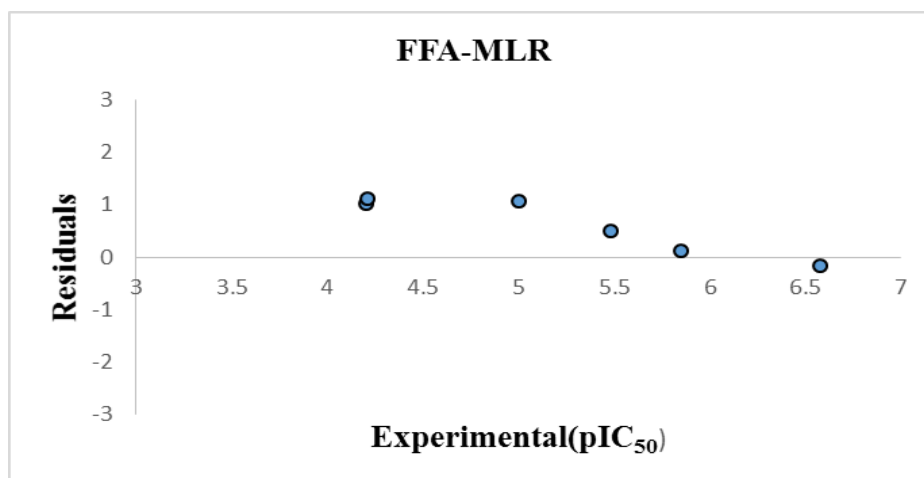
## ۲-۳. ارزیابی مدل‌ها با استفاده از نمودار خطای باقیمانده

اختلاف مقادیر پیش بینی شده و مقادیر تجربی، خطای باقیمانده نامیده می‌شود. پراکندگی یکنواخت نقاط، حول محور افقی که بیانگر خطای باقیمانده‌ی صفر است، نشان‌دهنده‌ی آن است که خطای سیستماتیک در مدل سازی وجود ندارد. نمودار خطای باقیمانده بر حسب مقادیر تجربی، برای مدل‌های ذکر شده در شکل‌های (۵) و (۶) نشان داده شده است.



شکل ۵. نمودار مقادیر باقیمانده  $pIC_{50}$  بر حسب مقادیر تجربی برای سری آموزش





شکل ۶. نمودار مقادیر باقیمانده  $pIC_{50}$  بر حسب مقادیر تجربی برای سری تست

### ۳-۳. ارزیابی مدل ها با استفاده از پارامترهای آماری

مطابق جدول (۵) پنج پارامتر آماری، جهت ارزیابی توانایی پیش بینی مدل های ساخته شده به روش های FF-MLR و SW-MLR به کار گرفته شد. این نتایج نشان دهنده آن است که الگوریتم کرم شب تاب به همراه مدلسازی MLR می تواند به طور نسبتاً رضایت بخشی فعالیت دارویی این سری ترکیبات را پیش بینی نماید.

جدول ۵. پارامترهای آماری برای مدل های انتخاب شده

		SW-MLR	FF-MLR
Press	سری آموزش	۳/۷۸۲۳	۱۰/۵۱۱۹
	سری تست	۲/۳۰۸۹	۳/۷۰۹
RMSEP	سری آموزش	۰/۳۸۹	۰/۶۶۱۸
	سری تست	۰/۶۲۰۲	۰/۷۸۶۲
REP	سری آموزش	۶/۹۰۸۷	۱۱/۴۵۵
	سری تست	۱۰/۷۲۳۲	۱۰/۵۲۶
SEP	سری آموزش	۰/۳۸۹	۰/۶۶۱۸
	سری تست	۰/۶۲۰۲	۰/۷۸۶۲
R <sup>2</sup> pred	سری آموزش	۰/۸۲۶۷	۰/۶۷۱۳
	سری تست	۰/۸۳۵۵	۰/۸۵۹۱

### ۳-۴. ارزیابی مدل ها توسط روش رد مرحله ای تک تک و گروهی

به منظور بررسی بیشتر قدرت پیش بینی مدل ها، تکنیک رد مرحله ای تک تک و گروهی مورد استفاده قرار گرفت. در روش رد مرحله ای تک تک، هر بار یکی از ترکیبات به طور تصادفی از سری داده ها حذف شدند و در روش رد مرحله ای گروهی، هر بار یک گروه از ترکیبات (۵ ترکیب) به طور تصادفی از سری داده ها حذف شدند. سپس با استفاده از مدل ساخته شده توسط بقیه ترکیبات، فعالیت دارویی ترکیب یا ترکیبات حذف شده، پیش بینی شدند. این فرایند برای تمام اعضای سری داده ها تکرار شد. نتایج حاصل از رد مرحله ای و گروهی در جدول (۶) ارائه شده است.

جدول ۶. پارامترهای آماری برای مدل های انتخاب شده

		SW-MLR	FF-MLR
$Q^2_{LOO}$	کل داده ها	۰/۷۲۴۷	۰/۸۵۷
$Q^2_{LGO}$	کل داده ها	۰/۶۷۲۹	۰/۸۳۷

## ۳-۵. ارزیابی مدل های ارائه شده با استفاده از آزمون Y-تصادفی

آزمون Y-تصادفی به عنوان یک تکنیک ارزیابی مدل، با هدف بررسی هر گونه ارتباط تصادفی بین داده ها انجام شد. در این آزمون، متغیر وابسته بطور تصادفی بهم ریخته شد. مدل QSAR جدید با استفاده از ماتریکس متغیرهای مستقل اصلی و مقادیر تصادفی از متغیر وابسته توسعه یافت. اگر در مدل اصلی هیچ گونه ارتباط تصادفی وجود نداشته باشد، تفاوت قابل توجهی بین مقدار ضریب تعیین مدل اصلی و مدل QSAR که با پاسخ تصادفی توسعه یافته، وجود خواهد داشت. نتایج حاصل از چندین بار اجرای آزمون Y-تصادفی در جدول (۷) نشان داده شده است. مقادیر کوچک ضریب تعیین ( $R^2$ ) بیانگر عدم ارتباط شانس در مدل توسعه یافته توسط رگرسیون خطی چندگانه می باشد.

جدول ۷. نتایج حاصل از ۱۰ بار آزمون Y-تصادفی

تکرار	$R^2_{test}$	
	SW-MLR	FF-MLR
۱	۰/۱۹۸	۰/۰۹۹
۲	۰/۰۹۷	۰/۰۲۶
۳	۰/۳۰۹	۰/۰۶۸
۴	۰/۲۶۶	۰/۰۸۷
۵	۰/۰۶۴	۰/۰۸۰
۶	۰/۰۶۸	۰/۰۷۷
۷	۰/۰۳۴	۰/۰۳۶
۸	۰/۰۰۲	۰/۰۵۵
۹	۰/۱۲۶	۰/۰۶۰
۱۰	۰/۰۶۵	۰/۱۲۵

## ۳-۶. بررسی ارتباط توصیفگرهای وارد شده در مدل با اثر بازدارندگی

در این قسمت به طور خلاصه ارتباط بین توصیفگرهای وارد شده در مدل و اثر بازدارندگی ترکیبات، مورد بررسی قرار خواهد گرفت. با توجه به نتایج به دست آمده در مدل برتر، توصیفگرهای انتخاب شدند که هر کدام بیانگر خصوصیات متفاوتی از مولکول مورد بررسی است. توصیفگرهای انتخاب شده توسط روش رگرسیون مرحله ای شامل (HATS8e, RDF080m, RDF090u, D/DR05) و (H-050) و توصیفگرهای انتخاب شده توسط الگوریتم کرم شب تاب شامل (MS, C-028, CIC3) و (HOMA و R7m) می باشند.

۳-۶-۱. توصیفگرهای RDF<sup>۱</sup>

می توان گفت RDF یا تابع توزیع شعاعی مربوط به یک دسته از اتم ها، معادل توزیع احتمال یافتن یک اتم در فضای کروی به شعاع R است. رابطه زیر نحوه محاسبه تابع RDF را نشان می دهد.

$$g(R) = \frac{1}{f} \sum_i^{N-1} \sum_{j>i}^N A_i \cdot A_j \cdot e^{-B \cdot (R-r_{ij})^2}$$

که  $f$  یک فاکتور مقیاس، و  $N$  تعداد اتم های مولکول است. همچنین  $r_{ij}$  فاصله ی بین دو اتم  $i$  و  $j$  و  $A$  یک ویژگی اتمی (وزن دار نشده  $u$ ، جرم اتمی  $m$ ، حجم واندروالس  $v$ ، الکترونگاتیویته  $e$ ، قطبش پذیری  $p$ ) است.  $\beta$  یک فاکتور تسهیل کننده است که توزیع احتمال فاصله بین اتمی را مشخص می کند و می توان از آن به فاکتور دما برای تعریف جنبش اتمی تعبیر کرد.  $g(R)$  در نقاط گسسته-ای با فواصل معین محاسبه و با مجموعه ای از کدهای RDF با ویژگی های مختلف اتمی می توان ساختار سه بعدی یک مولکول را به طور واضح توصیف کرد. این توصیفگرها اهمیت توزیع اتم ها در مولکول روی اثر بازدارندگی آنرا نشان می دهند. همچنین این توصیفگرها علاوه بر فاصله اتمی، اطلاعات با ارزشی در مورد فاصله پیوند، انواع حلقه، سیستم های مسطح و غیر مسطح و نوع اتم ها فراهم می کند.

از این گروه، توصیفگرها RDF080m و RDF090u در مدل ارائه شده برای پیش بینی اثر بازدارندگی انتخاب شده اند. اثر متوسط توصیفگرهای RDF080m و RDF090u منفی ذکر شده یعنی با افزایش مقدار این توصیفگرها  $pIC_{50}$  کاهش می یابد.

۳-۶-۲. توصیفگر ACFC<sup>۲</sup>

کدهای اجزای با مزیت اتمی (ACFC) یک کد با مزیت اتمی با دامنه ی کوتاه است که هر اتم را با نوع اتم، انواع پیوند و انواع اتم های همسایه توصیف می کند. هر مولکول کاملاً با یک کد اجزا که تعداد اتم های غیر هیدروژن است معرفی می شود. این توصیفگرها اطلاعات شیمیایی زیادی را در رابطه با گروه های عاملی در اطراف یک اتم مرکزی و مولکول ارائه می دهند. ثابت های اتمی آب گریز خوزه-کریپن<sup>۳</sup> (توصیف کننده های چربی دوستی) میزانی از توزیع ترکیبات بین یک فاز آلی و یک فاز آبی را نشان می دهد که معمولاً با ضرایب جداسازی  $P$  به صورت توزیع غلظت یک ترکیب در فازهای آلی و آبی از تحت شرایط تعدلی زیر تعریف می شود:

$$P = \frac{[C]_{org}}{[C]_{aq}}$$

که  $[C]_{org}$  و  $[C]_{aq}$ ، غلظت های حل شده در فازهای آلی و آبی هستند.

توصیفگر H-050 (H متصل به کربن با هیبریداسیون  $sp^3$  و عدد اکسایش صفر)، یکی دیگر از توصیفگرهای از نوع اجزای با مرکزیت اتمی می باشد که اولین همسایه اتم کربن، اتم هیدروژن می باشد. همچنین، این توصیفگر نمایانگر تعداد اتم های هیدروژن متصل به اتم کربن با هیبریداسیون  $sp^3$  می باشد. اثر متوسط این توصیفگر مثبت است و نشان دهنده ی این است که با افزایش مقدار این

<sup>1</sup>-Radial Distribution Function Descriptors

<sup>2</sup>-Atom-Centered Fragments Code

<sup>3</sup>-Ghose Crippen

توصیفگر،  $pIC_{50}$  افزایش می‌یابد. توصیفگر دیگر از این نوع C-028 می‌باشد. اثر متوسط این توصیفگر منفی است که نشان می‌دهد با افزایش این توصیفگر،  $pIC_{50}$  کاهش می‌یابد.

### ۳-۶-۳. توصیفگرهای گروه Topological

این توصیفگرها بر اساس نمایش گراف مولکول می‌باشند. در این گراف‌ها هر نقطه نشان‌دهنده‌ی یک اتم بوده و خطوط بین نقاط نیز نشان‌دهنده‌ی پیوند شیمیایی بین اتم‌ها می‌باشد. معمولاً در گراف‌های مولکولی اتم هیدروژن را نشان نمی‌دهند. این توصیفگرها اطلاعاتی راجع به ساختمان، اندازه، شکل، تقارن، شاخه‌دار شدن، نحوه اتصال اتم‌ها و نوع اتم‌های موجود در یک مولکول را در اختیار ما قرار می‌دهند. محاسبه این توصیفگرها به سادگی از روی ساختمان دو بعدی مولکول‌ها امکان پذیر می‌باشد. از این گروه توصیفگرها در مدل ارائه شده توصیفگر D/DR05 و CIC3 انتخاب شدند. اثر متوسط برای هر دو توصیفگر CIC3 و D/DR05 مثبت است که نشان می‌دهد با افزایش هر یک از آنها مقدار  $pIC_{50}$  افزایش خواهد یافت.

### ۳-۶-۴. توصیفگرهای Charge

این توصیفگرها اطلاعاتی در مورد بار الکتریکی مولکول و بار جزئی روی اتم‌ها می‌دهند و قادرند اطلاعاتی از جمله ماکزیمم بار مثبت، ماکزیمم بار منفی، کل بارهای مثبت و منفی و ... را نشان دهند. توصیفگر PCWTe در این گروه از توصیفگرها قرار می‌گیرد. این توصیفگر اطلاعاتی راجع به میزان برهم کنش‌های الکتریکی بین آن اتم، یا سایر اتم‌های موجود در مولکول می‌باشد. اثر متوسط توصیفگر HATS8e مثبت بوده و نشان‌دهنده‌ی این است که با افزایش مقدار این توصیفگر،  $pIC_{50}$  افزایش می‌یابد. همچنین اثر متوسط R7m نیز مثبت بوده و نشان‌دهنده‌ی این است که با افزایش مقدار این توصیفگر،  $pIC_{50}$  افزایش می‌یابد.

### ۴. نتیجه گیری

در این مقاله، مدل QSAR برای پیش بینی فعالیت بازدارندگی برخی از مشتقات پیریدین در درمان سرطان پوست بکار گرفته شد. دو مدل خطی چند متغیره با ۵ توصیفگر انتخاب شده بوسیله روشهای گام به گام (SW) و الگوریتم کرم شب‌تاب (FFA) ایجاد شد. مقایسه پارامترهای آماری بدست آمده برای مدل‌های SW-MLR و FFA-MLR نشان داد که هر دو مدل توانسته است نتایج قابل قبولی ارائه دهد. در نهایت با استفاده از تکنیکهای ارزیابی مشخص شد که مدل FFA-MLR ایجاد شده با کیفیت آماری بالاتر و همچنین خطای پیش بینی کمتر می‌تواند فعالیت بازدارندگی این نوع ترکیبات را پیش بینی کند.

### ۵. مراجع

- [1] Melagraki, G., Afantitis, A., Makridima, K., Sarimveis, H. and Igglessi-Markopoulou, O., Prediction of toxicity using a novel RBF neural network training methodology. *Journal of molecular modeling*, 12(3) (2006) 297-305.
- [2] Pourbasheer, E., Riahi, S., Ganjali, M.R. and Norouzi, P., QSAR study of C allosteric binding site of HCV NS5B polymerase inhibitors by support vector machine. *Molecular diversity*, 15(3) (2011) 645-653.

- [3] Habibi-Yangjeh, A., Pourbasheer, E. and Danandeh-Jenagharad, M., Application of principal component-genetic algorithm-artificial neural network for prediction acidity constant of various nitrogen-containing compounds in water. *Monatshefte für Chemie-Chemical Monthly*, 140(1) (2009) 15-27.
- [4] Beheshti, A., Pourbasheer, E., Nekoei, M. and Banaei, A., Quantitative structure-activity relationship study of amino acid derivatives as histone deacetylase inhibitors using the genetic algorithm–Multiple linear regression. *Analytical Chemistry Letters*, 2(1) (2012) 33-43.
- [5] Adimi, M., Salimi, M., Nekoei, M., Pourbasheer, E. and Beheshti, Abolghasem, A quantitative structure-activity relationship study on histamine receptor antagonists using the genetic algorithm-multi-parameter linear regression method. *J Serb Chem Soc*, 77(5) (2012) 639-650.
- [6] Dolatabadi, M., Nekoei, M. and Banaei, A., Prediction of antibacterial activity of pleuromutilin derivatives by genetic algorithm–multiple linear regression (GA–MLR). *Monatshefte für Chemie-Chemical Monthly*, 141(5) (2010) 577-588.
- [7] Nekoei, M., Salimi, M., Dolatabadi, M. and Mohammadhosseini, M., Prediction of antileukemia activity of berbamine derivatives by genetic algorithm–multiple linear regression. *Monatshefte für Chemie-Chemical Monthly*, 142(9) (2011) 943.
- [8] Nekoei, M., Salimi, M., Dolatabadi, M. and Mohammadhosseini, M., A quantitative structure-activity relationship study of tetrabutylphosphonium bromide analogs as muscarinic acetylcholine receptors. *Journal of the Serbian Chemical Society*, 76(8) (2011) 1117-1127.
- [9] Todeschini, R., Consonni, V. and Mannhold, R., 2000. Methods and principles in medicinal chemistry. *Kubinyi H, Timmerman H (Series eds) Handbook of molecular descriptors*. Wiley-VCH, Weinheim.
- [10] Nekoei, M., Mohammadhosseini, M. and Zarei, K., Simultaneous Kinetic Determination of Phosphate and Silicate by Spectrophotometric H-Point Standard Addition Method. *Journal of the Chinese Chemical Society*, 55(2) (2008) 362-368.
- [11] Zarei, K., Atabati, M. and Nekoei, M., Simultaneous spectrophotometric determination of phosphate and silicate by using principal component artificial neural network. *Annali di Chimica: Journal of Analytical, Environmental and Cultural Heritage Chemistry*, 97(8) (2007) 723-731.
- [12] Nekoei, M. and Mohammadhosseini, M., Simultaneous spectrophotometric determination of iron and cobalt in micellar medium by using a principal component artificial neural network and multivariate calibration. *Journal of the Chinese Chemical Society*, 54(2) (2007) 383-390.
- [13] Habibi-Yangjeh, A. and Nooshyar, M., Prediction of normalized polarity parameter in binary mixed solvent systems using artificial neural networks. *Physics and Chemistry of Liquids*, 43(3) (2005) 239-247.
- [14] Meiler, J., Meusinger, R. and Will, M., Fast determination of <sup>13</sup>C NMR chemical shifts using artificial neural networks. *Journal of chemical information and computer sciences*, 40(5) (2000) 1169-1176.
- [15] Vidal Mosquera, M., 2016. *Chemical modulation of the nociceptive receptor TRPV1: Synthetic, biological and computational studies* (Doctoral dissertation, Universitat Ramon Llull).
- [16] Garbe, C., Hauschild, A., Volkenandt, M., Schadendorf, D., Stolz, W., Reinhold, U., Kortmann, R.D., Kettelhack, C., Frerich, B., Keilholz, U. and Dummer, R., Evidence and interdisciplinary consensus-based German guidelines: diagnosis and surveillance of melanoma. *Melanoma research*, 17(6) (2007) 393-399.
- [17] El-Gamal, M.I., Jung, M.H., San Lee, W., Sim, T., Yoo, K.H. and Oh, C.H., Design, synthesis, and antiproliferative activity of new 1H-pyrrolo [3, 2-c] pyridine derivatives against melanoma cell lines. *European journal of medicinal chemistry*, 46(8) (2011) 3218-3226.



[18] Jung, M.H., El-Gamal, M.I., Abdel-Maksoud, M.S., Sim, T., Yoo, K.H. and Oh, C.H., Design, synthesis, and antiproliferative activity of new 1H-pyrrolo [3, 2-c] pyridine derivatives against melanoma cell lines. Part 2. *Bioorganic & medicinal chemistry letters*, 22(13) (2012) 4362-4367.

[۱۹] م صنّعی آّاده، ز جبل عاملیان - "الگوریتم های تکاملی و محاسبات زیستی" جلد اول، ویراست دوم، ناشر نیاز دانش، تهران. ص ۲۴۰

[۲۰] ف مریخ بیات - "الگوریتم های بهینه سازی (الهام گرفته از طبیعت)"، جلد اول، چاپ دوم، ناشر نص، تهران. ص ۱۸۰-۱۹۱