

پیشگویی شاخص سمیت LD_{50} در مشتقات آنیلین (حاوی ترکیبات علف‌کش) با روش‌های محاسباتی

مرتضی رضایی^۱، عصمت محمدی نسب^{۱*}

۱- به‌ترتیب دانشجوی دکتری، استادیار، دانشگاه آزاد اسلامی، واحد اراک، گروه شیمی، اراک، ایران

چکیده

در دهه‌های گذشته، استفاده از روش‌های محاسباتی با پارامترهای اعتبارسنجی دقیق برای تعیین خواص فیزیکی-شیمیایی ترکیبات، به‌عنوان جایگزین اقتصادی و زیست‌محیطی باصرفه جویی در زمان و حذف هزینه‌های بالا موردتوجه بسیاری از پژوهشگران قرار گرفته است. در این مطالعه، به بررسی ارتباط مقادیر لگاریتمی سمیت $(\log LD_{50})$ با توصیف‌گرهای مولکولی برای ۶۰ نوع از مشتقات آنیلین (شامل ترکیبات علف‌کش) پرداخته شده است. بعد از ترسیم ساختار این ترکیبات با استفاده از نرم‌افزار Gauss View ۰۵ و بهینه‌سازی آن‌ها با کمک نرم‌افزار ۰۹ Gaussian با روش $6-311++G**$ / B3LYP توصیف‌گرهای مولکولی استخراج شدند. به کمک ژنتیک الگوریتم، توصیف‌گرهای نامناسب حذف شده و بهترین آن‌ها برای مدل‌های رگرسیون خطی چندگانه و شبکه عصبی مصنوعی مورد استفاده قرار گرفتند. نتایج حاصل از این مدل نشان داد که روش شبکه عصبی مصنوعی با کمترین خطا و بالاترین ضریب تعیین نسبت به روش رگرسیون خطی چندگانه برای پیش‌بینی لگاریتم سمیت $LD_{50}(\text{molkg}^{-1})$ مشتقات آنیلین از برتری بالایی برخوردار است.

واژه‌های کلیدی: "متوسط دوز کشنده"، "روش رگرسیون خطی چندگانه"، "شبکه عصبی مصنوعی"، "مشتقات آنیلین"

* نویسنده رابط، پست الکترونیکی: e-mohammadinasab@iau-arak.ac.ir

تاریخ دریافت مقاله: ۹۷/۸/۲۶ - تاریخ پذیرش مقاله: ۹۸/۲/۲۰



مقدمه

آمین‌های نوع اول دسته وسیعی از ترکیبات شیمیایی هستند که کاربردهای گوناگونی در صنایع دارویی، تولید حشره‌کش‌ها در کشاورزی و تولید آفت‌کش‌ها دارند (Hofmann, 1843). آنیلین و مشتقات آن از دسته آمین‌های نوع اول هستند. (Wojcik, 2007) آنیلین با فرمول $C_6H_5NH_2$ متشکل از یک حلقه ۶ عضوی است که یک گروه آمین به یکی از کربن‌های روی حلقه متصل است و به‌طور وسیعی، به‌عنوان یکی از محصولات حاصل از کارخانه‌های سموم زراعی از جمله علف‌کش‌ها در محیط پراکنده می‌شود. این ماده به‌عنوان ماده اولیه در صنایع شیمیایی مختلف در ساخت دارو، رنگ، لاستیک و پلاستیک مورد استفاده قرار می‌گیرد و همچنین یکی از محصولات جانبی در صنایع نساجی و کاغذ است. هنگام رها شدن آنیلین در محیط‌های آبی، تداخلی در چرخه زندگی موجودات آبرزی ایجاد می‌شود، به‌این‌علت که به‌محض وارد شدن در خون، به‌سادی واکنش می‌دهد، مانع جذب اکسیژن گردیده، هموگلوبین را به مت هموگلوبین تبدیل می‌کند و باعث بروز بیماری مت هموگلوبینی می‌گردد و برای انسان سرطان‌زاست. بسیاری از علف‌کش‌ها از خانواده دی نیترو آنیلین هستند که به‌منظور حذف علف‌های هرز در محصولات کشاورزی کاربرد داشته و در روزهای گرم توسط نور خورشید تجزیه می‌گردند. به دلیل سمیت و ویژگی خاص آن‌ها و کاربردهای وسیع در صنایع شیمیایی و کشاورزی، این مواد برای محیط زیست و سلامت موجودات زنده یک تهدید جدی به‌شمار می‌آیند (Benigni & Passerini, 2002; Oh *et al.*, 1991; Ward *et al.*, 1997). قرار گرفتن انسان در معرض آنیلین می‌تواند به دلیل تنفس هوای آلوده، دخانیات و یا نزدیک شدن به صنایع تولید شده باشد. اثرات حاد و مزمن آنیلین در انسان عمدتاً شامل سردرد، خستگی، التهاب دستگاه تنفس فوقانی، گرفتگی و نیز اثراتی بر روی خون می‌باشد. تحلیل گلبول‌های قرمز خون و مت هموگلوبین یکی از اثرات جانبی قرار گرفتن در معرض انواع مختلف مشتقات آریل آمین شامل مواد شیمیایی کشاورزی و صنعتی است. مطالعات اخیر با آنیلین، متابولیت‌های N-هیدروکسیل را به‌عنوان واسطه‌های گلبول قرمز (RBC) شناسایی کرده‌اند و نشان داده‌اند که آنیلین علت تومور مثانه است، درحالی‌که مطالعات حیوانی نشان می‌دهد که آنیلین باعث تومورهای طحال می‌شود. مشتقات آنیلین مانند فنول‌ها در واکنش‌های الکتروفیلی جایگزین، بسیار واکنش‌پذیر هستند. به‌عنوان مثال، اسید سولفانیلیک از سولفوناسیون آنیلین تولید می‌شود که می‌تواند منجر به تولید سولفانیل آمید گردد که در اوایل قرن بیستم به‌عنوان ضد باکتری مورد استفاده قرار گرفت (Singh, 2005).

در دهه‌های اخیر تلاش برای یافتن خواص ترکیبات شیمیایی جدید با استفاده از روش‌های محاسباتی، رشد چشمگیری داشته است. تکنیک محاسباتی می‌تواند برهمکنش بین دو مولکول را پیش‌بینی کند. بررسی و تولید ترکیباتی که در بدن، اثرات جانبی و سمیت نداشته باشند، مطالعه ترکیباتی که از لحاظ فارماکوسیتیکی در سیستم‌های بیولوژیک پایدار هستند، طراحی منطقی مواد پرمصرف در حوزه صنایع مانند عطرها، رنگ‌های شیمیایی و مواد شیمیایی، شناسایی مواد سمی زیان‌آور برای محیط‌زیست و صرفه‌جویی در زمان و نیز کاهش هزینه‌های مربوط به انجام آزمایش‌ها، می‌تواند از مزایای به‌کارگیری روش‌های محاسباتی در پیش‌بینی خواص شیمیایی و فیزیکی مواد شیمیایی جدید باشند که با بازدهی بالایی نیز همراه هستند (Eriksson, 2003). یکی از این روش‌ها، بررسی ارتباط کمی ساختار- خاصیت^۱ ساختار- فعالیت^۲ می‌باشد.

¹ Quantitative Structure Property Relationship

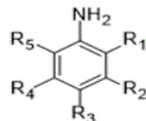
² Quantitative Structure Activity Relationship

(Polishchuk, 2017; Mohammadinasab, 2017; Shahri & Niazi, 2016) در این روش‌ها، ابتدا با استفاده از نرم‌افزارها و برنامه‌های مختلف کامپیوتری، با کمک توصیف‌گرهای مولکولی مانند توصیف‌گرهای کوانتومی، توپولوژیکی و هندسی، ساختار ترکیبات را با فرمول‌های ریاضی، به اعداد تبدیل می‌کنند. سپس با کمک مدل‌های ریاضی اقدام به پیش‌بینی خواص و یا فعالیت ترکیبات می‌شود. در روش‌های رگرسیون خطی و غیرخطی که در این مدل‌سازی‌ها استفاده می‌شوند، خواص و یا فعالیت به‌عنوان متغیر وابسته و توصیف‌گرهای مولکولی به‌عنوان متغیرهای مستقل در نظر گرفته می‌شوند. تاکنون تحقیقات بی‌شماری در کاربرد روش‌های محاسباتی برای تعیین خواص فیزیکی - شیمیایی ترکیبات مختلف انجام شده است

(Karbakhsh & Sabet, 2011; Roy & Mitra, 2011; Tao., *et al.*, 2002; Damborskyl & Schulz, 1996; Roberts, 1987). روش‌های محاسباتی شامل الگوهای خطی و غیرخطی می‌باشند. در مواقعی که روش خطی ارائه‌دهنده الگوی مناسبی برای رسیدن به هدف نیست، روش‌های غیرخطی مورد استفاده قرار می‌گیرند. (Namasivayam, 2001; Cross, 1995; Haykin, 1990; Lapedes, 1987). یکی از این روش‌ها، شبکه عصبی مصنوعی است که بر اساس نورون‌های عصبی پایه‌گذاری شده است و توانایی این را دارد که روش‌های آماری کلاسیک و مدرن را ترکیب کند. در یک شبکه عصبی مصنوعی، برای ایجاد ارتباط بین متغیرهای مستقل و وابسته از لایه‌هایی استفاده می‌شود که این لایه‌ها با یکدیگر در ارتباط هستند و شامل لایه ورودی، لایه خروجی و لایه‌های پنهان یا میانی است. لایه‌های پنهان به این دلیل استفاده می‌شوند که بر ارائه مدل نهایی تأثیرگذار هستند و دلیل به‌کار بردن یک لایه خروجی به خاطر ارائه یک نتیجه منطقی برای پیش‌بینی خاصیت مورد نظر می‌باشد (Gadzuric *et al.*, 2014; Miller & Miller, 2010; Maiellaro *et al.*, 2004). بررسی ارتباط بین متغیرهای وابسته و مستقل در روش‌های محاسباتی و ارائه مدل‌هایی که بتوانند قبل از هرگونه سنتزی، خواص ترکیبات را پیش‌بینی کنند، برای پژوهشگران علوم تجربی از ارزش بسیار بالایی برخوردار است. یکی از متغیرهای وابسته مهم شاخص LD₅₀ می‌باشد. LD₅₀ یک معیار آماری است، شامل دوز لازم از یک ترکیب شیمیایی که ۵۰ درصد حیوانات تحت آزمایش را پس از ۲۴ ساعت می‌کشد و از آن می‌توان برای مقایسه سمیت ترکیبات شیمیایی با یکدیگر استفاده نمود، به‌طوری‌که هرچه مقدار LD₅₀ کمتر باشد، سمیت ماده بیشتر است (Williams, 2003). در این تحقیق ارتباط لگاریتم شاخص سمیت LD₅₀ با توصیف‌گرهای مولکولی برای ۶۰ نوع از مشتقات آنیلین، با کمک روش‌های محاسباتی خطی و غیرخطی، بررسی گردیده و بهترین مدل برای پیش‌بینی سمیت مورد نظر ارائه گردیده است.

مواد و روش‌ها

برای انجام این تحقیق، تعداد ۶۰ ترکیب از مشتقات آنیلین با استفاده از وب‌سایت Sigma Aldrich استخراج گردید و مقادیر لگاریتمی مربوط به شاخص سمیت ((log(LD₅₀)(molkg⁻¹)) برای همه آن‌ها، با استفاده از وب‌سایت Chemicalize.org، در جدول (۱)، درج گردیده است.



شکل ۱- ساختار شیمیایی مشتقات آنیلین

Fig. 1- The chemical structure of aniline derivatives

جدول ۱- مشتقات آنیلین مورد مطالعه و مقادیر لگاریتمی LD₅₀Table 1- The studied aniline derivatives and their log (LD₅₀) values

No.	Name	R ₁	R ₂	R ₃	R ₄	R ₅	log (LD ₅₀)(molkg ⁻¹)
1	3-chloro-2,6-diethylaniline	H	Et	Cl	H	Et	2.23
2	3-chloro-2-fluoroaniline	H	F	Cl	H	H	2.56
3	3-chloro-4-methoxyaniline	H	H	Cl	OCH ₃	H	2.26
4	3-chloro-4-methylaniline	H	H	Cl	Me	H	2.50
5	3-chloroaniline	H	H	Cl	H	H	2.48
6	3-ethylaniline	H	H	Et	H	H	2.15
7	3-fluoro-4-(trifluoromethoxy)aniline	H	H	F	OCF ₃	H	1.80
8	3-methylaniline	H	H	Me	H	H	2.32
9	3-Nitroaniline	H	H	NHO ₂	H	H	2.29
10	4,4'-methylenedianiline	H	H	H	C ₇ H ₇ - NH ₂	H	2.59
11	4,6-dibromo-2,3-dichloroaniline	Cl	Cl	F	F	H	2.72
12	4-amino-2,3-difluoro phenol	H	F	F	OH	H	3.31
13	4-amino-2,6-dichlorophenol	H	Cl	H	OH	Cl	2.32
14	4-Aminophenol	H	H	H	OH	H	2.62
15	4-bromo-2-(trifluoromethyl)aniline	H	CF ₃	H	H	Br	1.81
16	4-bromo-2,6-dimethylaniline	H	Me	H	Br	Me	2.01
17	4-bromo-2-chloro-6-methylaniline	H	Cl	H	Br	Me	2.01
18	4-bromo-2-chloroaniline	H	Cl	H	Br	H	2.35
19	4-bromo-2-ethylaniline	H	Et	H	Br	H	2.80
20	4-bromo-2-methoxyaniline	H	OCH ₃	H	Br	H	2.67
21	4-bromo-2-methylaniline	H	Me	H	Br	H	2.01
22	4-bromo-3-(trifluoromethyl)aniline	H	H	CF ₃	Br	H	1.99
23	4-bromo-3-chloroaniline	H	H	Cl	Br	H	2.40
24	4-bromo-3-methoxyaniline	H	H	OCH ₃	Br	H	2.79
25	4-bromo-3-methylaniline	H	H	Me	Br	H	1.93
26	4-bromoaniline	H	H	H	Br	H	2.30
27	4-butoxyaniline	H	H	H	OC ₄ H ₉	H	2.13
28	4-butylaniline	H	H	H	C ₄ H ₉	H	2.13
29	4-chloro-2-(trifluoromethyl)aniline	H	CF ₃	H	Cl	H	1.96
30	4-chloro-2,5-dimethoxyaniline	H	OCH ₃	H	Cl	OCH ₃	1.99
31	4-chloro-2,6-difluoroaniline	H	F	H	Cl	F	2.03
32	4-chloro-3-methylaniline	H	H	Me	Cl	H	2.05
33	4-chloroaniline	H	H	H	Cl	H	2.33
34	4-cyclohexylaniline	H	H	H	C ₆ H ₁₂	H	2.36
35	4-decylaniline	H	H	H	C ₁₀ H ₂₁	H	2.38
36	4-Ethoxyaniline	H	H	H	OC ₂ H ₅	H	2.32
37	4-ethylaniline	H	H	H	Et	H	2.26
38	4-fluoro-2-(trifluoromethyl)aniline	H	CF ₃	H	F	H	2.08
39	4-fluoro-2,3-dimethylaniline	H	Me	Me	F	H	2.29
40	4-hexylaniline	H	H	H	C ₆ H ₁₂	H	2.21
41	4-hexyloxyaniline	H	H	H	OC ₆ H ₁₃	H	2.28
42	4-isopropylaniline	H	H	H	C ₃ H ₈	H	1.98
43	4-methoxy-2-methylaniline	H	Me	H	OCH ₃	H	2.18

No.	Name	R ₁	R ₂	R ₃	R ₄	R ₅	log (LD ₅₀)(molkg ⁻¹)
44	4-methoxy-3,5-dimethylaniline	H	H	Me	OCH ₃	Me	2.43
45	4-methyl-3-(trifluoromethyl)aniline	H	H	CF ₃	Me	H	1.87
46	4-methylaniline	H	H	H	Me	H	2.37
47	4-n-butylaniline	H	H	H	C ₅ H ₉	H	2.13
48	4-nonylaniline	H	H	H	C ₉ H ₁₉	H	2.23
49	4-octylaniline	H	H	H	C ₈ H ₁₇	H	2.13
50	4-pentylaniline	H	H	H	C ₅ H ₁₁	H	2.61
51	4-pentyloxyaniline	H	H	H	OC ₅ H ₁₁	H	2.36
52	4-phenoxyaniline	H	H	H	C ₆ H ₅ OH	H	2.34
53	4-propylaniline	H	H	H	C ₃ H ₇	H	2.19
54	5-bromo-2-methoxyaniline	H	OCH ₃	H	H	Br	2.27
55	5-bromo-2-methylaniline	H	Me	H	H	Br	2.28
56	5-chloro-2,4-dimethoxyaniline	H	OCH ₃	H	OCH ₃	Cl	2.36
57	5-fluoro-2-methoxyaniline	H	OCH ₃	H	H	F	2.10
58	aniline	H	H	H	H	H	2.24
59	o-anisidine	OCH ₃	H	H	H	H	2.43
60	o-toluidine	Me	H	H	H	H	2.38

لازم به ذکر است که همه مقادیر مندرج در این مطالعه مربوط به شاخص سمیت تجویز دهانی (موش) LD₅₀ می‌باشند که به دلیل بزرگی برخی از این مقادیر در ترکیبات مورد مطالعه، از مقادیر لگاریتمی آن‌ها استفاده شده است. پس از جمع‌آوری ترکیبات و داده‌ها، ساختار اولیه تمامی ترکیبات با استفاده از نرم‌افزار Gauss View ۰۵ ترسیم و با کمک نرم‌افزار Gaussian ۰۹ به روش تئوری تابعیت چگالی DFT و خط دستور *G**++۳۱۱-۶/B3LYP اقدام به بهینه‌سازی آن‌ها گردید. به منظور محاسبه انواع توصیف‌گرهای مولکولی (Todeschini & Consonni, 2010; Todeschini & Consonni, 2000) نرم‌افزار Dragon به کار گرفته شد و به منظور غربالگری شاخص‌های به‌دست آمده، از برنامه MATLAB ۲۰۱۰a استفاده شد (Elisseeff & Guyo, 2003). سپس روش الگوریتم ژنتیک جهت انتخاب بهترین توصیف‌کننده‌ها به کار گرفته شد. از الگوریتم ژنتیک در شیمی به‌طور گسترده‌ای در زمینه مدل‌سازی QSPR و کمومتریکس استفاده شده است (Ghosh & Bagchi, 2009; Saxena & Prathipati, 2003). این الگوریتم از اصول انتخاب طبیعی داروین برای یافتن فرمول بهینه جهت پیش‌بینی یا تطبیق الگو استفاده می‌کند و اغلب گزینه خوبی برای تکنیک‌های پیش‌بینی بر مبنای رگرسیون است. کاربرد الگوریتم ژنتیک در مدل‌سازی QSPR، جستجو در میان توصیف‌کننده‌ها به منظور یافتن موثرترین آن‌ها جهت انجام مدل‌سازی است. الگوریتم ژنتیک، یک روش جستجوی هوشمند و تصادفی است که با به‌کارگیری عملگرهای ژنتیک از یک فرآیند تکامل تدریجی تبعیت می‌کند. پس از حذف توصیف‌کننده‌های غیرضروری، آن‌هایی که بیشترین میزان ارتباط با لگاریتم سمیت را داشتند، برای مدل‌سازی انتخاب گردیدند که برای این منظور از روش‌های رگرسیون خطی چند متغیره و شبکه عصبی مصنوعی استفاده گردید. هم‌چنین اهمیت شاخص‌های به کار رفته در مدل نهایی به ترتیب، با توجه به شاخص حساسیت جهانی^۱ مورد بررسی قرار گرفت. حساسیت جهانی، یک معیار عددی است و مقدار آن بستگی به درجه اهمیت شاخص به کار رفته در مدل دارد (Iooss, & Lemaitre, 2014)

¹ global sensitivity

به منظور برقراری ارتباط بین توصیف گره‌های مولکولی و خاصیت موردنظر و همچنین اعتبار سنجی نتایج به دست آمده، از روش ارزیابی خارجی استفاده شد. داده‌های ورودی، توسط نرم‌افزار Statistica V. ۱۲ به صورت تصادفی، در روش MLR به دو دسته ۴۵ آموزشی و ۱۵ آزمایشی (به نسبت ۷۵٪ به ۲۵٪)، و در روش MLP به سه دسته، شامل دسته آموزش به میزان ۷۰٪ از داده‌ها و دسته‌های آزمون و اعتبارسنجی هر کدام به میزان ۱۵٪ تقسیم‌بندی شدند. در نهایت با کمک نرم‌افزار فوق، اقدام به مدل‌سازی برای بررسی ارتباط بین متغیرها در دو روش گردید.

روش‌های محاسباتی به کار رفته

۱- رگرسیون خطی چندگانه MLR^۱

در روش رگرسیون خطی چندگانه، منظور از y (متغیر وابسته) شاخص سمیت LD₅₀ مشتقات آنیلین شامل برخی ترکیبات علف‌کش، منظور از x_1, x_2, \dots (متغیرهای مستقل) توصیف گره‌های مولکولی و a_1, a_2, \dots ضرایب رگرسیون در معادله (۱) می‌باشند (Kutner et al., 2004; Randic & Basak, 2000).

$$y = a_1x_1 + a_2x_2 + \dots + a_kx_k \quad (1)$$

۲- شبکه عصبی مصنوعی ANN^۲

یکی از روش‌های آماری پرکاربرد در تجزیه و تحلیل اطلاعات و داده‌های آماری در مطالعه رابطه ساختار-خاصیت و ساختار-فعالیت، روش شبکه عصبی مصنوعی است. بررسی ارتباط غیرخطی بین متغیرهای مستقل با متغیر وابسته به کمک این روش امکان‌پذیر است. اعمال روش‌های غیرخطی، استفاده گسترده‌ای در حیطه مطالعات ارتباط ساختار-فعالیت و ساختار-خاصیت دارد. در صورتی که معلوم شود ارتباط خطی مناسبی میان شاخص سمیت با توصیف گره‌های مولکولی ترکیبات مورد مطالعه وجود ندارد، به بررسی رابطه غیرخطی از طریق شبکه عصبی مصنوعی پرداخته می‌شود.

نتایج و بحث

پس از رسم و بهینه‌سازی ترکیبات مورد مطالعه با نرم‌افزار گوسین، فایل‌های خروجی حاصل از محاسبات کوانتومی برای همه ترکیبات مورد مطالعه، به برنامه Dragon انتقال داده شد و ۱۴۲۲ توصیف‌گر مولکولی به دست آمد. یکی از مشکلاتی که بعد از استفاده از این نرم‌افزار به وجود می‌آید، ایجاد تعداد زیاد توصیف‌کننده به عنوان متغیرهای مستقل است که در اغلب موارد تعداد آن‌ها از تعداد مولکول‌ها بسیار بیشتر است. باید توجه داشت، توصیف‌گرهای محاسبه‌شده در یک ماتریکس داده‌ای جمع‌آوری شدند که در این ماتریکس، تعداد مولکول‌ها در ردیف و تعداد توصیف‌گرها در ستون قرار گرفتند (Iman & Davood, 2013). از میان تمام توصیف‌گرها، متغیرهای دارای مقادیر یکسان و نیز متغیرهای تکراری مشخص شده و حذف شدند. از این تعداد شاخص‌های اولیه با حذف شاخص‌های غیرضروری با استفاده از نرم‌افزار متلب، ۹۸۱ شاخص باقی‌ماند. سپس ارتباط توصیف‌گرها با یکدیگر و با فعالیت مولکول‌ها مورد بررسی قرار گرفت. توصیف‌گرهایی که در یک نام و گروه توصیف‌کننده‌های محاسبه‌شده، به همراه بعد توصیف‌کننده و تعداد آن‌ها که توسط نرم‌افزار متلب محاسبه شده‌اند به طور کامل در جدول (۲)، نشان داده شده‌اند.

^۱ Multiple Linear Regression

^۲ Artificial Neural Network

جدول ۲- انواع توصیف‌کننده‌های محاسبه‌شده توسط نرم‌افزار متلب

Table 2- Types of descriptors calculated by MATLAB software

No.	Name	Descriptors dimension	Descriptors no.
1	constitutional descriptors	0	11
2	topological descriptors	2	69
3	walk and path counts	2	29
4	connectivity indices	2	32
5	information indices	0	46
6	2D autocorrelations	2	55
7	edge adjacency indices	3	86
8	Burden eigenvalues	2	49
9	topological charge indices	0	8
10	eigenvalue-based indices	0	44
11	Randic molecular profiles	3	49
12	geometrical descriptors	3	23
13	RDF descriptors	3	61
14	3D-MoRSE descriptors	3	159
15	WHIM descriptors	3	98
16	GETAWAY descriptors	3	151
17	molecular properties	0	11
Total of indices			981

پس از حذف توصیف‌گرهای غیرضروری و استخراج توصیف‌کننده‌های مناسب، بایستی توصیف‌کننده‌هایی انتخاب شوند که ضمن داشتن اطلاعات مفید، از قدرت پیش‌بینی قابل قبولی برای مدل انتخابی برخوردار باشند. در این مطالعه، پس از اعمال ژنتیک الگوریتم توسط نرم‌افزار MATLAB ۲۰۱۰a برای انتخاب بهترین توصیف‌کننده‌ها، تعداد توصیف‌گرهای نهایی برای مدل‌سازی به ۸ شاخص رسید. لازم به ذکر است، این توصیف‌گرها به همراه تعریف آن‌ها در جدول (۳) ارائه شده‌اند و در روش رگرسیون خطی برای مشخص شدن الگوی بهتر به کار گرفته شدند. سپس با توجه به مدل نهایی به‌دست‌آمده، ۴ توصیف‌گر به روش غیرخطی، مورد استفاده قرار گرفتند. تعریف و نوع آن‌ها نیز در جدول (۳) مشخص گردیده است.

جدول ۳- شاخص‌های نهایی به‌کاررفته در روش رگرسیون خطی چندگانه

Table3- The final descriptors used on MLR method

Descriptors	Block	classification
GATS5p	2D autocorrelations	Geary autocorrelation - lag 5 / weighted by atomic polarizabilities
GATS5v	2D autocorrelations	Geary autocorrelation - lag 5 / weighted by atomic van der Waals volumes
RTp+	GETAWAY descriptors	R maximal index / weighted by atomic polarizabilities
H5m	GETAWAY descriptors	H autocorrelation of lag 5 / weighted by atomic masses
Mor16m	3D-MoRSE descriptors	3D-MoRSE - signal 16 / weighted by atomic masses
HATS2e	GETAWAY descriptors	leverage-weighted autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities
G1m	WHIM descriptors	1st component symmetry directional WHIM index / weighted by atomic masses
Mor26m	3D-MoRSE descriptors	3D-MoRSE - signal 26 / weighted by atomic masses

مقادیر مختلف پارامترهای آماری به‌دست‌آمده برای ۴۵ ترکیب از دسته آموزش و ۱۵ ترکیب از دسته آزمون به روش MLR، در جدول (۴) ارائه شده است.

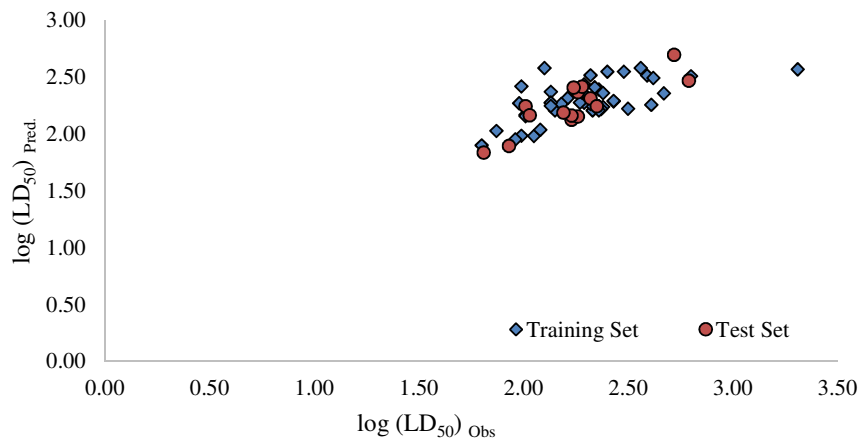
جدول ۴- پارامترهای آماری به‌دست آمده برای دسته آموزش و آزمون در روش رگرسیون خطی چندگانه

Table 4- The obtained statistical parameters of training and test set on MLR method

	Pro.	N	R	R ²	R ² _{adj}	RMSE	p-value	D-W
Train	LD ₅₀	45	0.637	0.406	0.347	0.002	0.0002	2.10
Test	LD ₅₀	15	0.753	0.567	0.495	0.02	0.0005	2.92

در جدول (۴)، تعداد ترکیبات (N)، ضریب رگرسیون (R)، ضریب تعیین (R²)، ضریب رگرسیون تعدیل‌شده (R²_{adj})، خطای جذر میانگین مربعات (RMSE)، سطح معناداری (P-value)، آماره دوربین-واتسون (D-W)، در دسته‌های آموزش و آزمون در روش رگرسیون خطی چندگانه نشان داده شده‌اند.

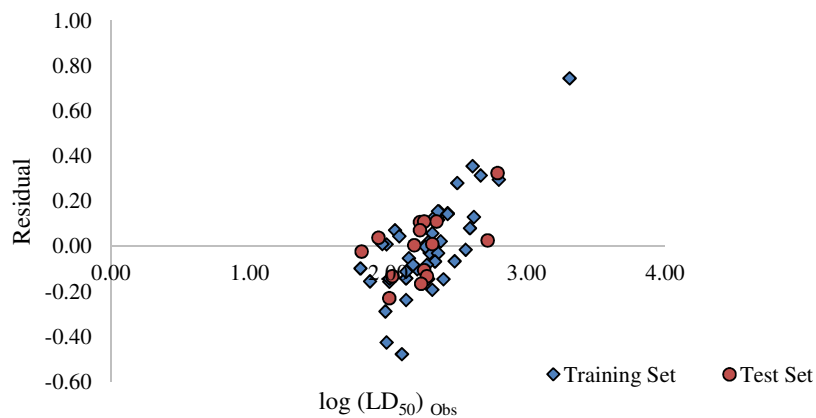
در شکل (۲)، نمودار پراکنندگی مقادیر پیش‌بینی برحسب مشاهده شده، نشان داده شده است.



شکل ۲- نمودار مقادیر پیش‌بینی برحسب مشاهده شده log(LD₅₀) در روش رگرسیون خطی چندگانه

Fig2- The curve of predicted versus observed log (LD₅₀) on MLR method

هم‌چنین در شکل (۳) مقادیر باقی‌مانده برحسب مشاهده شده برای لگاریتم سمیت با روش MLR نشان داده شده است.



شکل ۳- نمودار مقادیر باقی‌مانده برحسب مشاهده شده log(LD₅₀) در روش رگرسیون خطی چندگانه

Fig 3- The curve of the residual versus observed log (LD₅₀) values on MLR method

تجزیه و تحلیل با استفاده از روش MLR

مقادیر مختلف پارامترهای آماری به دست آمده به روش MLR، ضمن تاکید بر عدم همبستگی خوب بین متغیرها، نامناسب بودن این روش در ارائه مدل، جهت پیش‌بینی لگاریتم سمیت ترکیبات را نشان می‌دهد.

مقدار $R=0/637$ میزان رابطه خطی بین سمیت و متغیرهای مستقل را بیان می‌کند. مقدار ضریب تعیین $0/406 = R^2$ مطابق با جدول (۴) نشان می‌دهد که ۴۰ درصد تغییرات متغیر وابسته به وسیله متغیرهای مستقل تبیین می‌شود. R^2_{adj} مقدار ضریب تعیین را با توجه به متغیرهای مستقل اضافه شده به خط رگرسیون و با توجه به عرض از مبدهای جدید، تعدیل و اصلاح می‌کند. RMSE تفاوت میان مقدار پیش‌بینی شده توسط مدل با مقدار مشاهده شده است. با توجه به مقدار سطح معنی‌داری کمتر از ۰/۰۵، می‌توان نتیجه گرفت که رابطه مورد نظر معنی‌دار می‌باشد. آماره D-W برای آشکارسازی وجود خودهمبستگی بین باقی مانده‌ها مورد استفاده قرار می‌گیرد. به‌طور کلی، مقدار این آماره در بازه بین ۰ تا ۴ در نوسان است و مقدار بهینه آن برابر ۲ می‌باشد، بنابراین مقدار ۲/۱ برای داده‌های آموزش مطابق جدول (۴) نشانه عدم خودهمبستگی میان باقی مانده هاست.

معادله نهایی به دست آمده (معادله (۲)) در روش خطی، شامل ۴ نوع توصیف‌گر نهایی و ضرایب تاثیر آن‌ها، به شرح

زیر می‌باشد:

$$\log(LD_{50}) = 219.55 - 0.44(HATS 2e) + 3.46(G1m) - 0.04(GATS 5p) + 3.42(RTp +) \quad (2)$$

در شکل (۲)، نمودار پراکنندگی مقادیر مشاهده شده برحسب پیش‌بینی شده مطابق با معادله (۲)، گواه بر این است که روش MLR، روش مناسبی برای پیش‌بینی لگاریتم سمیت مشتقات آنیلین نمی‌باشد.

روش ANN پارامترهای مربوط به اعتبارسنجی در هر سه دسته آموزش، آزمون و اعتبارسنجی در روش ANN، مطابق جدول (۵) ارائه شده است. تعداد لایه‌های ورودی و پنهان و خروجی نیز در جدول (۵) نشان داده شده است.

جدول ۵- پارامترهای آماری به دست آمده برای دسته‌های آموزش، آزمون و اعتبارسنجی در روش شبکه عصبی مصنوعی

Table 5- The obtained statistical parameters of training, test and validation sets on ANN method

ANN(4-5-1)	N	R	R ²	RMSE
Training	42	0.953	0.908	0.12
Test	9	0.885	0.783	0.17
Validation	9	0.807	0.651	0.72

همچنین اهمیت شاخص‌های به کاررفته در مدل، با توجه به شاخص حساست جهانی در جدول (۶) نشان داده شده است. ترتیب کاهش اهمیت توصیف‌گرهای به دست آمده در جدول (۶)، با توجه به مقادیر عددی آن‌ها در روش شبکه عصبی مصنوعی، نشان می‌دهد که میزان تاثیرگذاری توصیف‌گر Rtp+ بر روی سمیت مشتقات آنیلین از همه بیشتر و G1m کمتر از بقیه توصیف‌گرها می‌باشد. این توصیف‌گر از دسته توصیف‌گرهای متعلق به گروه GETAWAY¹ می‌باشد که مجموعه‌ای از توصیف‌گرهای توپولوژی، هندسی و وزن‌های اتمی است و سعی می‌کند که ساختار هندسی ۳ بعدی مولکول و توپولوژی آن را با اطلاعات شیمیایی به دست آمده از وزن‌های اتمی تطبیق نماید.

¹ GEometry, Topology and Atom Weights Asembly

جدول ۶- مقادیر حساسیت جهانی برای توصیف‌گرهای نهایی در روش شبکه عصبی مصنوعی

Table 6- The global sensitivity values of final descriptors on ANN method

MLP(4-5-1)	RTp+	GATS5p	HATS2e	G1m
	2.025	1.901	1.490	1.106

همچنین مقادیر پیش‌بینی و باقی‌مانده مربوط به لگاریتم شاخص LD₅₀ در روش ANN مطابق جدول (۷) گزارش شده است.

جدول ۷- مقادیر پیش‌بینی و باقی‌مانده مربوط به لگاریتم شاخص LD₅₀ در روش شبکه عصبی مصنوعی

Table 7- The predicted and residual LD₅₀ values on ANN method

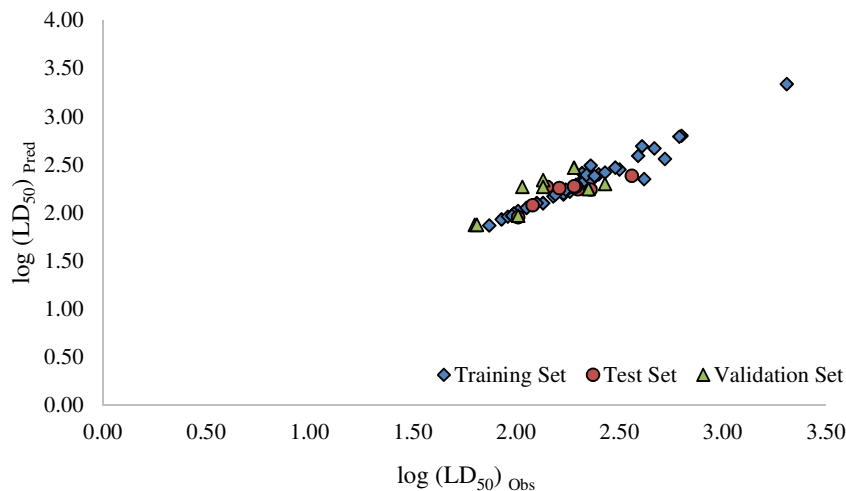
No	Name	Predicted LD ₅₀	Residual LD ₅₀	No	Name	Predicted LD ₅₀	Residual LD ₅₀
1	3-chloro-2,6-diethylaniline***	2.19	0.03	31	4-chloro-2,6-difluoroaniline**	2.27	-0.24
2	3-chloro-2-fluoroaniline*	2.38	0.18	32	4-chloro-3-methylaniline	2.05	0.00
3	3-chloro-4-methoxyaniline***	2.24	0.02	33	4-chloroaniline	2.36	-0.03
4	3-chloro-4-methylaniline	2.45	0.05	34	4-cyclohexylaniline	2.49	-0.13
5	3-chloroaniline	2.47	0.02	35	4-decylaniline	2.37	0.01
6	3-ethylaniline*	2.27	-0.12	36	4-Ethoxyaniline	2.31	0.01
7	3-fluoro-4-(trifluoromethoxy)aniline**	1.88	-0.08	37	4-ethylaniline***	2.22	-0.04
8	3-methylaniline***	2.31	0.01	38	4-fluoro-2-(trifluoromethyl)aniline*	2.08	0.00
9	3-Nitroaniline	2.29	0.00	39	4-fluoro-2,3-dimethylaniline*	2.27	0.02
10	4,4'-methylenedianiline***	2.59	0.00	40	4-hexylaniline*	2.26	-0.05
11	4,6-dibromo-2,3-dichloroaniline***	2.56	0.16	41	4-hexyloxyaniline*	2.27	0.00
12	4-amino-2,3-difluoro phenol	3.34	-0.03	42	4-isopropylaniline***	1.97	0.01
13	4-amino-2,6-dichlorophenol	2.41	-0.09	43	4-methoxy-2-methylaniline	2.17	0.02
14	4-Aminophenol	2.35	0.07	44	4-methoxy-3,5-dimethylaniline	2.42	0.01
15	4-bromo-2-(trifluoromethyl)aniline**	1.88	-0.07	45	4-methyl-3-(trifluoromethyl)aniline	1.87	0.00
16	4-bromo-2,6-dimethylaniline*	1.95	0.02	46	4-methylaniline	2.36	0.01
17	4-bromo-2-chloro-6-methylaniline	2.02	-0.01	47	4-n-butylaniline	2.10	0.02
18	4-bromo-2-chloroaniline***	2.25	0.10	48	4-nonylaniline***	2.20	0.03
19	4-bromo-2-ethylaniline***	2.80	0.00	49	4-octylaniline	2.10	0.03
20	4-bromo-2-methoxyaniline	2.67	0.00	50	4-pentylaniline	2.69	-0.01
21	4-bromo-2-methylaniline**	1.97	0.04	51	4-pentyloxyaniline	2.34	0.02
22	4-bromo-3-(trifluoromethyl)aniline	1.99	0.00	52	4-phenoxyaniline	2.39	-0.05
23	4-bromo-3-chloroaniline	2.40	0.00	53	4-propylaniline***	2.19	0.00
24	4-bromo-3-methoxyaniline***	2.79	0.00	54	5-bromo-2-methoxyaniline	2.27	0.00
25	4-bromo-3-methylaniline***	1.93	0.00	55	5-bromo-2-methylaniline**	2.47	-0.19
26	4-bromoaniline*	2.24	0.06	56	5-chloro-2,4-dimethoxyaniline*	2.24	0.12
27	4-butoxyaniline**	2.34	-0.21	57	5-fluoro-2-methoxyaniline	2.10	0.00
28	4-butylaniline**	2.27	-0.14	58	aniline***	2.24	0.00
29	4-chloro-2-(trifluoromethyl)aniline	1.96	0.00	59	o-anisidine**	2.30	0.13
30	4-chloro-2,5-dimethoxyaniline***	1.99	0.00	60	o-toluidine	2.38	0.00

*دسته آزمون در روش شبکه عصبی مصنوعی

**دسته اعتبارسنجی در روش شبکه عصبی مصنوعی

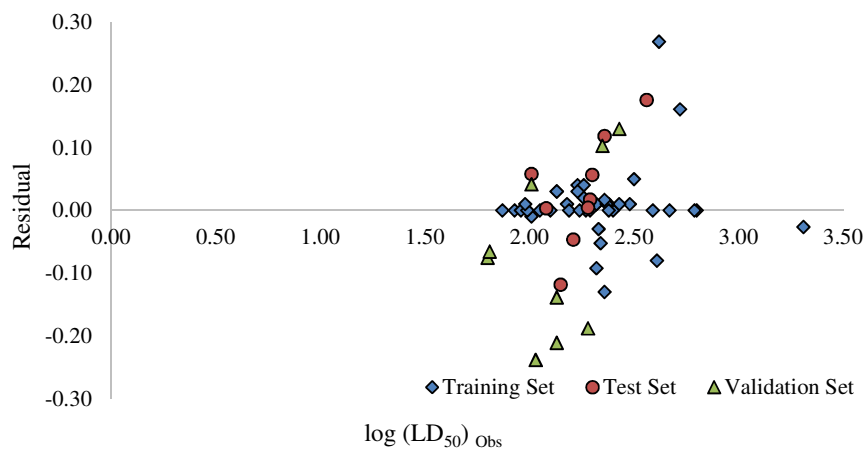
***دسته آزمون در روش رگرسیون خطی چندگانه

میزان پراکندگی نقاط برای سه دسته آموزش و آزمون و اعتبارسنجی در روش شبکه عصبی مصنوعی در شکل (۴) نشان داده شده است.



شکل ۴- نمودار مقادیر پیش‌بینی برحسب مشاهده شده $\log(LD_{50})$ در روش شبکه عصبی مصنوعی
Fig 4- The curve of the predicted versus observed $\log(LD_{50})$ values on ANN method

نمودار مقادیر LD_{50} مشاهده شده برحسب باقی مانده در مدل ANN در شکل (۵) نشان داده شده است.



شکل ۵- نمودار مقادیر باقی مانده برحسب مشاهده شده $\log(LD_{50})$ در روش شبکه عصبی مصنوعی
Fig 5- The curve of the residual versus observed $\log(LD_{50})$ values on ANN method

تجزیه و تحلیل با استفاده از روش ANN

مقادیر نسبتاً بالای R برای هر سه دسته آموزش، آزمون و اعتبارسنجی هم‌چنین مقدار کم RMSE در آنها مطابق جدول (۵)، کارایی روش شبکه عصبی مصنوعی برای پیش‌بینی لگاریتم شاخص سمیت LD_{50} را تأیید می‌نماید، به طوری که با میزان درصد خطای ۱۲ و ضریب رگرسیون ۰/۹۵۳، می‌توان بیش از ۹۰ درصد از تغییرات شاخص سمیت در ترکیبات آموزش را به وسیله توصیف‌گرهای مولکولی تبیین نمود.

میزان پراکندگی نقاط برای سه دسته آموزش و آزمون و اعتبارسنجی در نمودار مشاهده‌شده برحسب پیش‌بینی‌شده، مطابق شکل (۴)، حاکی از همبستگی خوب بین توصیف‌گرهای نهایی با لگاریتم سمیت و در نتیجه کار آبی روش به‌کار رفته در پیش‌بینی LD₅₀ مشتقات آنیلین می‌باشد. مقادیر LD₅₀ مشاهده شده برحسب باقی‌مانده نشان می‌دهد که خطاها در اطراف محور x توزیع یکسانی دارند و این نمایانگر این است که الگوی مذکور برای پیش‌بینی این شاخص سمیت ترکیبات مورد مطالعه از دقت بالا و بسیار خوبی برخوردار می‌باشد.

تاکنون مطالعات زیادی در زمینه مدل‌سازی به روش QSAR، صورت گرفته است. Jackel و همکار وی با مطالعه ارتباط ساختار-فعالیت به روش خطی، به پیش‌بینی $\log(I/LD_{50})$ در ۳۳ ترکیب از مشتقات آنیلین بر روی پستانداران و چگونگی ارتباط آن با $\log P_{ow}$ پرداختند (Jackel & Klein, 1991). در طی مطالعه‌ای که توسط گونگ و همکاران روی جلبک‌ها انجام گرفت، رابطه میان $\log 1/EC_{50}$ با ضریب توزیع آب-اکتانال و گپ انرژی هومو-لومو توسط مدل‌سازی به روش خطی، برای پیش‌بینی سمیت ۲۰ ترکیب از مشتقات فنل و آنیلین مورد بررسی قرار گرفت (GUANG, 2008). در حالی که، این مطالعه برای اولین بار به کمک محاسبات کوانتومی توسط نرم‌افزار گوسین جهت بهینه‌سازی مولکول‌ها و نیز استخراج توصیف‌گرهای مولکولی توسط ژنتیک الگوریتم، ضمن مقایسه روش‌های محاسباتی خطی و غیرخطی، توانست، رابطه میان شاخص سمیت تجویز خوراکی (موش) LD₅₀ برای ۶۰ ترکیب از مشتقات آنیلین با توصیف‌گرهای مولکولی به روش شبکه عصبی راه، با دقت بسیار خوبی نشان دهد. مقایسه روش‌های رگرسیون خطی با شبکه عصبی، نشان داد که بین مقادیر سمیت مشتقات آنیلین با توصیف‌گرهای مولکولی، رابطه غیرخطی وجود دارد. نتایج این مطالعه، نشان داد که مدل رگرسیون خطی به‌کار گرفته شده، علی‌رغم اینکه چارچوب وسیعی را در برمی‌گیرد که با تحلیل‌های زیادی همراه هست، ولی نمی‌تواند الگوی مناسبی برای هدف این تحقیق باشد، زیرا در بعضی مواقع، متغیرهای رگرسیونی با توابع غیرخطی به هم مربوط می‌شوند. اگرچه روابط غیرخطی بین پارامترها، می‌توانند به‌وسیله مدل‌های متنوع رگرسیونی مدل‌سازی شوند، ولی طبیعت و ذات این روابط به‌گونه‌ای است که همه این مدل‌ها نمی‌توانند چنین روابطی را به‌خوبی تجزیه و تحلیل کنند. بنابراین، شبکه عصبی مصنوعی به‌عنوان یک مدل تشخیص روندهای نا پارامتری و غیرخطی، می‌تواند روابط غیرخطی پیچیده یا لایه‌های پنهان بین متغیرهای وابسته و مستقل را بیابد و با دقت بهتری نسبت به روش‌های رگرسیونی خطی عمل نماید. روش ANN رابطه خوب بین ورودی و خروجی ایجاد نموده و از حساسیت کمتری نسبت به وجود خطاها در اطلاعات ورودی برخوردار است، لذا از نظر میزان توانایی در پیش‌بینی ارتباط بین متغیر وابسته و متغیرهای مستقل و قابلیت تعمیم نسبت به روش‌های دیگر مدل‌سازی، نتیجه قابل قبول‌تری ارائه داده است.

نتیجه‌گیری

بسیاری از آمین‌های نوع اول، سمی و برای سلامتی بشر مضرند. آنیلین به‌عنوان یک آمین نوع اول ماده بسیار سمی است و در صورت تماس با پوست و چشم باعث سوزش می‌شود و نباید مورد استنشاق قرار گیرد. اندازه‌گیری سمیت بسیاری از مشتقات آنیلین شامل علف‌کش‌های زراعی از روش‌های آزمایشگاهی به دلایل مختلف مانند اشتعال‌پذیری و یا سمیت آن‌ها زیان‌آور می‌باشد، به همین منظور در این مطالعه، به‌پیش‌بینی نسبتاً دقیق مقادیر لگاریتم سمیت LD₅₀ مشتقات آنیلین شامل برخی علف‌کش‌های مورد استفاده در زراعت با روش‌های محاسباتی پرداخته شده است. برای مطالعه QSAR سمیت ترکیبات مشتق از آنیلین از هر دو روش محاسباتی خطی و غیرخطی استفاده گردید. پارامترهای آماری در مدل نهایی برتری قابل توجه مدل MLP را نسبت به مدل MLR در شبکه عصبی ANN، برای پیش‌بینی لگاریتم سمیت ترکیبات

مورد مطالعه نشان دادند. نتایج این تحقیق نشان داد که می‌توان گام‌های بلندی را جهت پیش‌بینی مقادیر سمیت بسیاری از مشتقات جدید آنیلین با تکیه بر مدل‌های محاسباتی با پارامترهای ارزیابی مناسب برداشت.

Reference

- Ajdukovic, N. and Kovacevic, S. Z. 2014. Gadzuric, S. B., Podunavac-Kuzmanovic, S. O., Jokic, A. I., Vranes, M. B., Ajdukovic, N. and Kovacevic, S. Z. 2015.** Chemometric estimation of post-mortem interval based on Na⁺ and K⁺ concentrations from human vitreous humour by linear least squares and artificial neural networks. *Australian Journal of Forensic Sciences*, 46: 166-179.
- Bagheban Shahri, F. and Niazi, A. 2016.** Quantitative structure activity relationship study of inhibitory activities of 5-lipoxygenase and design new compounds by different chemometrics methods. *Iranian Journal of mathematical Chemistry*, 7: 47-59.
- Benigni, R. and Passerini, L. 2002.** Carcinogenicity of the aromatic amines: from structure-activity relationships to mechanisms of action and risk assessment. *Mutation Research*, 511: 191-206.
- Cross, S. S., Harrison, R. F. and Kennedy, R. L. 1995.** Introduction to neural networks, *Lancet*. 346: 1075-9.
- Damborsky, J. and Schulz, T. W. 1996.** comparison of the QSAR models for toxicity and biodegradability of anilines and phenols. *Chemosphere*, 34: 429-446.
- Elisseeff, A. and Guyon, I. 2003.** An introduction to variable and feature selection. *J. of Machine Learning Research*, 19: 1157-1182.
- Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M. and Gramatica, P. 2003.** Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect.* 111(10):1361-1375.
- Ghosh, P. and Bagchi, M. C. 2009.** QSAR modeling for quinoxaline derivatives using genetic algorithm and simulated annealing based feature selection. *Current Medicinal Chemistry*, 16: 4032-4048.
- Guang, h., wang, C. and Xiao-Ling, G. 2008.** Prediction of Toxicity of Phenols and Anilines to Algae by Quantitative Structure-activity Relationship. *Biomedical and environmental sciences* 21: 193-196.
- Haykin, S. 1990.** *Neural networks -a comprehensive foundation*, 2nd ed. New Jersey: Prentice-Hall.
- Hofmann, A. W. 1843.** Chemical investigation of organic bases in coal tar oil. *Annalen der Chemie und Pharmacie*, 47: 37-87.
- Iman, M. and Davood, A. 2013.** QSAR and QSTR study of selenocyanate derivatives to improve their therapeutic index as anti-leishmanial agents. *Med. Chem. Res.* 23(2):818-26.
- Iman, M. and Davood, A. 2013.** QSAR and QSTR study of pyrimidine derivatives to improve their therapeutic index as antileishmanial agents. *Med Chem. Res.*; 22(10):5029-35.
- Iooss, B. and Lemaitre, P. 2014.** A review on global sensitivity analysis methods. *Global sensitivity analysis*. HAL.
- Jackel, H. and Klein, W. 1991.** Prediction of Mammalian Toxicity by Quantitative Structure Activity Relationships: Aliphatic Amines and Anilines. *Quant. Struct.-Act. Relat.* 10: 198-204.
- Karbakhsh, R. and Sabet, R. 2011.** Application of different chemometric tools in QSAR study of azoloadamantanes against influenza A virus. *Reserch. in Pharmaceutical. Science*, 6: 23-33.
- Kutner, M. K., Nachtsheim, C. J. and Neter, J. 2004.** *Applied linear regression models*, Boston: McGraw-Hill.
- Lapedes, A., Farber, R. 1987.** Non-linear signal processing using neural networks. Technical Report no. LAUR- 2662 Los Alamos National Laboratory, Los Alamos, NM.

- Maiellaro, P. A., Cozzolongo, R. and Marino, P. 2004.** Artificial neural networks for the prediction of response to interferon plus ribavirin treatment in patients with chronic hepatitis C. *Cuurent Pharmaceutical Design*, 10: 2101-2109.
- Miller, J. N. and Miller, J. C. 2010.** Statistics and chemometrics for analytical chemistry. Harlow: Pearson Education Limited, 6th edn.
- Mohammadinasab, E. 2017.** Determination of Critical Properties of Alkanes Derivatives using Multiple Linear Regression. *Iranian J. Mathematical Chemistry*, 8: 199–220.
- Namasivayam, A., Waldemar, A. 2001.** Comparison of the prediction of extremely low birth weight neonatal mortality by regression analysis and by neural networks. *Early Hum Dev.* 65: 123-37.
- Oh, S. W., Kang, M. N., Cho, C. W. and Lee. M. W. 1997.** Detection of carcinogenic amines from dyestuffs or dyed substrates. *Dyes Pigments*, 33: 119-135.
- Polishchuk, P. 2017.** Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future. *J. Chemical Information Model*, 57: 2618-2639.
- Randic, M. and Basak, S. C. 2000.** Multiple regression analysis with optimal molecular descriptors. *SAR & QSAR Research*, 11: 1-23.
- Roberts, D. W. 1987.** An analysis of published data on fish toxicity of nitrobenzene and aniline derivatives, in *QSAR in Environmental Toxicology*. Edited by K.L. Kaiser. Vol. 2. Dordrecht: Springer Netherlands.
- Roy, K. and Mitra, I. 2011.** On Various Metrics Used for Validation of Predictive QSAR Models with Applications in Virtual Screening and Focused Library Design. *Combinatorial Chemistry & High Throughput Scree*, 14: 450-474.
- Saaidpour, S., Bahmani, A. and Rostami, A. 2015.** Prediction the Normal Boiling Points of Primary ,Secondary and Tertiary Liquid Amines from their Molecular Structure Descriptors. *Computational Methods Science & Technology*, 21: 201-210.
- Saxena, A. K. and Prathipati, P. 2003.** Comparison of MLR, PLS and GA-MLR in QSAR analysis. *SAR and QSAR in Environmental Research*, 14: 433-445.
- Singh, H. and Purnell, E. T. 2005.** Aniline derivative-induced methemoglobin in rats. *J Environ Pathol Toxicol Oncol*, 24(1):57-65.
- Tao, S., Xi, X., Xu, F. and Dawson, R. 2002.** A QSAR model for predicting toxicity (LC₅₀) to rainbow trout. *Water Research*, 36: 2926-2930.
- Todeschini, R. and Consonni, V. 2000.** Handbook of Molecular Descriptors. Weinheim: Wiley-VCH.
- Ward, E., Carpenter, A., Markowitz, S., Roberts, D., and Halperin, W. 1991.** Excess number of bladder cancers in workers exposed to ortho-toluidine and aniline. *National Concer Institute*, 83: 501-506.
- Web search engine developed by ChemAxon;** software available at “[http:// WWW. Chemicalize. Org](http://WWW.Chemicalize.Org)”.
- Williams, P. L., James, R. C., Roberts. S. M. 2003.** Ph.D., Principles of Toxicology: Environmental and Industrial Applications, Second Edition. John Wiley & Sons, Inc.
- Wójcik, G. M. 2007.** Structural Chemistry of Anilines in Anilines (Patai's Chemistry of Functional Groups). S. Patai, Ed. Weinheim.: Wiley-VCH.
- Yadav, R. and Nandi, S. 2014.** QSAR and anticancer drug design of β-Carboline compounds utilizing computed molecular descriptors. *Journal of Computational Methods in Molecular Design*, 4: 92-105.

Prediction of LD₅₀ for aniline derivatives (Including some herbicide compounds) using computational methods

M. Rezaei¹, E. Mohammadinasab¹

1- Respectively, Ph.d student and Assistant Profesor, Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran

Abstract

In recent decades, computational methods with regard to accurate validation parameters for the determination of the physical- chemistry properties of compounds have been considered by many researchers and have been used as an economic and environmental alternative to saving time and eliminating high costs. In this study, the relationship between the logarithmic values of LD₅₀, ($\log(\text{LD}_{50})(\text{molkg}^{-1})$) and molecular descriptors has been investigated for 60 types of aniline derivatives (including some herbicides compounds). At first, the structure of the compounds were drawn by Gauss view05 software and optimized using Gaussian 09 software with B3LYP/6-311++G** method, and then were extracted molecular descriptors. Then inappropriate descriptors were eliminated by genetic algorithm method and the best ones were used for multiple linear regression (MLR) and artificial neural networks (ANN) models. The results showed that the ANN method with the lowest error and the highest coefficient of determination was higher than the MLR method to predicting the $\log(\text{LD}_{50})(\text{molkg}^{-1})$ of studied aniline derivatives.

Key words: LD₅₀, Multiple linear regression method, Artificial neural network, Aniline derivatives

* Corresponding Author, E-mail: e-mohammadinasab@iau-arak.ac.ir

Received: 17 Nov. 2018– Accepted: 10 May. 2019

