

ORIGINAL ARTICLE

Predicting the toxicity of some pesticides against five different species of birds using quantitative structure-toxicity relationships model

Shahin Ahmadi¹, Shahram Lotfi^{2*}, Leila Shamakhi², Ali Azimi³

¹Faculty of Pharmaceutical Chemistry, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran.

²Department of Chemistry, Payame Noor University (PNU), Tehran, Iran.

³Department of Chemistry, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Correspondence

Shahram Lotfi

Email: Sh.lotfei@gmail.com

How to cite

Ahmadi, Sh., Lotfi, Sh., Shamakhi, L., & Azimi, A. (2024). Predicting the toxicity of some pesticides against five different species of birds using quantitative structure-toxicity relationships model. *Experimental Animal Biology*, 12(48), 77-91.

ABSTRACT

The increasing use of pesticides following the rising in agricultural, demand has threatened non-target organisms such as avian species and disrupted the ecological system. Therefore, considering the application methods and the nature of chemical pesticides, testing their toxicity level on birds, and protecting various species of endangered birds is an essential requirement from the point of view of ecosystem safety. In this study, quantitative structure-toxicity relationships modeling was done for the first time to estimate the toxicity of 244 types of pesticides on five different species of birds consist of bobwhite quail (*C. virginianus*), mallard duck (*A. platyrhynchos*), house sparrow (*P. domesticus*), ring-necked pheasant (*P. colchicus*), and Japanese quail (*C. japonica*). All data were randomly divided into four series including active training, passive training, calibration, and test sets. Hybrid optimal descriptors, resulting from the combination of quasi-SMILES descriptors and hydrogen-suppressed graph (HSG) based on a new target function, were used to generate QSTR models. Four target functions (TF₀, TF₁, TF₂, TF₃) were used to develop QSTR models and the predictive potential of these models was evaluated using a validation set. The QSTR models designed using TF₃ target function with the range of $R^2 = 0.7218-0.8131$ and $Q^2 = 0.7031-0.7878$ for the validation set were statistically the best models. Statistically, the best model is model number six, with R^2 values for active training, passive training, calibration, and validation sets equal to 0.836, 0.852, 0.806, and 0.813, respectively. The mean absolute error (MAE) values for the sets of active training, passive training, calibration and validation are 0.371, 0.342, 0.409 and 0.362, respectively, indicating the accuracy of the model created to predict the toxicity of pesticides against five species of endangered birds. From the results of this modeling, important descriptors were identified for increasing and decreasing the average effective toxicity concentration (pLD₅₀) of pesticides. Using the QSTR models obtained from this study, it becomes possible to predict the toxicity (pLD₅₀) of new pesticides even before their synthesis by only having the SMILES symbol of the pesticides, which can help to reduce time, resources, costs and the need for laboratory animals.

KEYWORDS

Avian oral toxicity, LD50, Optimal descriptor, Pesticide, QSAR.

نشریه علمی

زیست‌شناسی جانوری تجربی

«مقاله پژوهشی»

پیش‌بینی سمیت برخی آفت‌کش‌ها علیه پنج گونه مختلف پرندگان با استفاده از الگوی ارتباط کمی ساختار-سمیت

شهین احمدی^۱، شهرام لطفی^{۲*}، لیلا شماخی^۳، علی عظیمی^۳

چکیده

افزایش تقاضای کشاورزی به استفاده از آفت‌کش‌ها، موجودات غیرهدف مانند گونه‌های پرندگان را تهدید کرده و در سیستم اکولوژی اختلال ایجاد کرده است. بنابراین، با توجه به روش‌های کاربرد و ماهیت آفت‌کش‌ها، آزمایش سمیت پرندگان و حفاظت گونه‌های مختلف پرندگان در معرض خطر از منظر ایمنی اکوسیستم یک نیاز ضروری است. در این مطالعه، مدل‌سازی رابطه کمی ساختار-سمیت برای تخمین سمیت ۲۴۴ نوع آفت‌کش بر روی پنج گونه پرنده شامل بلدرچین سفید، اردک، قرقاول، گنجشک خانگی و بلدرچین ژاپنی برای اولین بار انجام شده است. تمام داده‌ها به صورت تصادفی به چهار سری شامل سری آموزش فعال، آموزش غیرفعال، کالیراسیون و دسته آزمون تقسیم شدند. توصیف‌گرهای بهینه هیبریدی، حاصل از ترکیب توصیف‌گرهای شبه-اسملی (quasi-SMILES) و گراف مولکولی بدون هیدروژن (HSG) براساس یک تابع هدف جدید برای تولید مدل‌های QSTR استفاده شد. چهار تابع هدف (TF0، TF1، TF2، TF3) برای توسعه مدل‌های QSTR به کار گرفته شد و پتانسیل پیش‌بینی این مدل‌ها با استفاده از سری آزمون مورد ارزیابی قرار گرفت. از میان مدل‌های به دست آمده، مدل‌های QSTR طراحی شده با استفاده از تابع هدف TF3 با محدوده $R^2 = 0.7218 - 0.8131$ و $Q^2 = 0.7031 - 0.7878$ از لحاظ آماری، مدل‌ها بودند. بهترین مدل از نظر آماری، مدل شماره شش، با مقادیر R^2 برای دسته آموزش فعال، آموزش غیرفعال، کالیراسیون و اعتبارسنجی به ترتیب برابر 0.836 ، 0.852 ، 0.806 و 0.813 می‌باشد. مقادیر میانگین خطای مطلق (MAE) برای دسته آموزش فعال، آموزش غیرفعال، کالیراسیون و اعتبارسنجی به ترتیب برابر 0.371 ، 0.342 ، 0.409 و 0.362 بیانگر دقت مدل ایجادشده به منظور پیش‌بینی سمیت آفت‌کش‌ها بر علیه پنج گونه مختلف پرندگان می‌باشد. از نتایج حاصل از این مدل‌سازی، توصیف‌گرهای با اهمیت برای افزایش و کاهش متوسط غلظت مؤثر سمیت (pLD₅₀) شناسایی شدند. با استفاده از مدل‌های QSTR به دست آمده از این مطالعه، پیش‌بینی سمیت (pLD₅₀) آفت‌کش‌های جدید حتی قبل از سنتز آن‌ها تنها با داشتن نماد SMILES از آفت‌کش‌ها امکان‌پذیر می‌شود که می‌تواند به کاهش زمان، منابع، هزینه‌ها و نیاز به حیوانات آزمایشگاهی کمک کند.

واژه‌های کلیدی

QSTR، LD₅₀، آفت‌کش، توصیف‌گر بهینه، سمیت خوراکی پرندگان.

^۱گروه شیمی (محض-دارویی)، دانشکده شیمی دارویی، علوم پزشکی تهران، دانشگاه آزاد اسلامی، تهران، ایران.
^۲گروه شیمی، دانشگاه پیام نور، تهران، ایران.
^۳گروه شیمی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

نویسنده مسئول:

شهرام لطفی

ایمانامه: Sh.lotfei@gmail.com

استناد به این مقاله:

احمدی، شهین؛ لطفی، شهرام؛ شماخی، لیلا و عظیمی، علی (۱۴۰۳). پیش‌بینی سمیت برخی آفت‌کش‌ها علیه پنج گونه مختلف پرندگان با استفاده از الگوی ارتباط کمی ساختار-سمیت. فصلنامه زیست‌شناسی جانوری تجربی، ۱۲ (۴۸)، ۷۷-۹۱.

<https://eab.journals.pnu.ac.ir/>

مقدمه

استفاده از آفت‌کش‌های شیمیایی در سراسر دنیا در دهه‌های اخیر به‌طور تصاعدی افزایش یافته است، زیرا کاربرد آفت‌کش‌ها در حفاظت از محصولات مختلف و همچنین در انتقال بیماری افزایش پیدا کرده است. به‌ویژه مصرف این مواد شیمیایی در کشورهای در حال توسعه و مبتنی بر کشاورزی بیش‌تر است. آفت‌کش‌های شیمیایی مولکول‌های سمی طراحی‌شده برای اهداف خاص هستند که کاربرد غیرعلمی آن‌ها ممکن است برای گونه‌های غیرهدف، از جمله انسان و اکوسیستم، خطراتی ایجاد کند. با توجه به ماندگاری و ویژگی نیمه‌فرار آفت‌کش‌های سنتزی، بخش قابل‌توجهی از دوز مصرفی آن‌ها را می‌توان در دوره‌های طولانی به‌عنوان باقیمانده در محصولات کشاورزی، خاک و همچنین در اتمسفر شناسایی کرد (Ahtar et al., 2009; Singh et al., 2014).

گونه‌های پرندگان بخشی جدایی‌ناپذیر از اکوسیستم را تشکیل می‌دهند. پرندگان با انجام خدمات مهمی مانند گرده‌افشانی گیاهان، کنترل آفات، پراکندگی بذر و پایش تنوع زیستی به پایداری محیط زیست و کشاورزی کمک می‌کنند. بنابراین، نگرانی‌های عمده‌ای در مورد تأثیر سموم دفع آفات بر روی پرندگان چه در نتیجه اثرات منفی غیرمستقیم آن‌ها بر تنوع و چه در نتیجه اثرات مضر مستقیم بر سلامت پرندگان بیان شده است (Hallmann et al., 2017; Humann-; Guilleminot et al., 2019). آفت‌کش‌ها می‌توانند از طریق غذا یا پوست وارد بدن پرنده گان شوند، اما مسیر دهانی یا خوراکی به‌عنوان مهم‌ترین مسیر مواجهه شدن این سموم است. مواردی از مسمومیت گونه‌های پرندگان در سراسر جهان گزارش شده است (از جمله مواردی که توسط ارگانوفسفرها و کاربامات ایجاد می‌شود) (Ahtar et al., 2009). استیل کولین استراز پروتئینی است که به این ترکیبات سمی متصل می‌شود و در نتیجه به سیستم عصبی گونه‌های پرندگان آسیب جدی وارد می‌کند که شامل کوری، تاری دید، بی‌حالی، عدم هماهنگی حرکات عضلانی، تشنج، فلج و مشکل در تنفس می‌شود. بنابر این، قرارگرفتن در معرض این مواد شیمیایی دلیل انقراض چندین گونه پرندگان و سایر گونه‌های جانوری در دهه‌های گذشته شده است (Mitra et al., 2011; Podder et al., 2023; Singh et al., 2014). با توجه به موارد فوق، نهادهای نظارتی مختلف بر ارزیابی سمیت آفت‌کش‌های شیمیایی موجود و همچنین جدید تأکید کرده‌اند.

بنابراین، برای اطمینان از ایمنی مواد شیمیایی، نیاز به آزمایش سمیت جامع وجود دارد.

به‌طور کلی، بلدرچین سفید شمالی (*Colinus virginianus*)، اردک (*Anas platyrhynchos*)، بلدرچین ژاپنی، قرقاول گردن حلقه‌ای و گنجشک خانگی، گونه‌های آزمایشی رایج پرندگان هستند که توسط سازمان همکاری‌های اقتصادی و توسعه (OECD) و آژانس حفاظت از محیط زیست ایالات متحده (USEPA) توصیه شده است (Basant et al., 2015; Hilton et al., 2019). در مورد آزمایش سمیت پرندگان، ۵۰ درصد دوز مؤثر کشندگی (LD_{50}) مناسب در نظر گرفته می‌شود. اگرچه پروتکل‌های استاندارد برای آزمایش سمیت پرندگان (LD_{50}) تکامل یافته است، این آزمایش‌ها (هم *in vivo* و هم *in vitro*) اغلب گران، زمان‌بر و غیراخلاقی هستند. سم‌شناسی شیمیایی را می‌توان با استفاده از روش‌های محاسباتی ساده کرد. روش‌های محاسباتی می‌توانند تعداد آزمایش‌ها و هزینه‌های ارزیابی سمیت تجربی را کاهش دهند و روش‌هایی مناسب برای ارزیابی اولیه در توسعه مواد شیمیایی جدید را ارائه دهند (Cronin et al., 2003; Hill et al., 1984; Jaworska et al., 2003; Zhu et al., 2008).

مدل‌سازی QSAR/QSTR یک روش محاسباتی مهم است که برای ایجاد ارتباط کمی بین ساختار شیمیایی ترکیبات (یعنی توصیف‌گر) و فعالیت/سمیت (خواص فیزیکو-شیمیایی و بیولوژیکی) استفاده می‌شود (Jeschke et al., 2013; Lotfi et al., 2022; Toropov & Toropova, 2018). نرم‌افزار کورال (CORrelation And Logic CORAL) به‌عنوان ابزاری برای طراحی و انجام تجزیه و تحلیل مدل‌های پیش‌بینی فعالیت و یا سمیت توصیه شده است (Ahmadi et al., 2021; Kumar, 2018; Toropova & Toropov, 2014). نمادهای سیستم ورود خط ورودی مولکولی ساده‌شده (SMILES) مربوط به ساختارهای شیمیایی برای محاسبه وزن همبستگی توصیف‌گر (DCW) در نرم‌افزار کورال با استفاده از بهینه‌سازی مونت کارلو استفاده می‌شود. کمیت DCW به ساختار ترکیب و ویژگی مورد تجزیه بستگی دارد و به ساختار سه‌بعدی مولکول بستگی ندارد. به هر حال، مدل‌های QSAR/QSTR ساخته‌شده توسط نرم‌افزار کورال را می‌توان از سه نوع توصیف‌گر توسعه داد: توصیف‌گرهای مبتنی بر SMILES، مبتنی بر گراف مولکولی، و توصیف‌گرهای ترکیبی (ترکیبی از SMILES و گراف مولکولی). مدل‌های ایجادشده براساس توصیف‌گرهای ترکیبی از نظر آماری بهتر از مدل‌های ساخته‌شده توسط توصیف‌گرهای

ژاپنی، به‌ترتیب [VIRG]، [PLAT]، [DOM]، [COL] و [JAP] می‌باشد. سپس quasi-SMILES مربوطه از ترکیب SMILES سموم و نوع پرند به‌دست آمد که در فایل داده‌های همراه مقاله می‌توان مشاهده کرد. در مراحل اولیه مدل‌سازی داده‌های پرت (داده‌های با میزان خطای بیش‌تر از سه برابر انحراف استاندارد مدل) شناسایی و از سری داده‌ها حذف شدند. چهار quasi-SMILES به‌عنوان داده پرت شناسایی و حذف شدند و مدل‌سازی نهایی با ۲۴۰ داده انجام شد. قرارگرفتن هر quasi-SMILES در دسته آموزش فعال (+)، آموزش غیر فعال (-)، کالیبراسیون (#) و اعتبارسنجی (*) و نتایج تجربی میزان سمیت و مقادیر پیش‌بینی شده آن‌ها برای هر هفت مدل در جدول S1 به‌عنوان داده‌های همراه مقاله ارائه شده است. با هدف ایجاد مدل QSTR و با استفاده از نرم‌افزار کورال، به‌روش تصادفی، داده‌ها به چهار دسته آموزش فعال، آموزش غیر فعال، کالیبراسیون و آزمون تقسیم شدند (Ghaedi, 2015). به سری آموزش فعال تقریباً ۳۰ درصد، آموزش غیرفعال ۲۶ درصد، کالیبراسیون ۲۰ درصد و ارزیابی ۲۴ درصد از داده‌ها اختصاص یافت. به‌صورت تصادفی این کار هفت بار تکرار شد (Benfenati et al., 2011).

وظیفه مجموعه آموزشی فعال محاسبه وزن‌های همبستگی و ایجاد مدل، وظیفه مجموعه آموزشی غیرفعال، بررسی مناسب‌بودن وزن‌های همبستگی ساختارهای مولکولی که در مجموعه آموزشی فعال به‌کار نرفته‌اند (ارزیابی کیفیت مدل) می‌باشد. مجموعه کالیبراسیون باید شروع بیش‌برازش را تشخیص دهد. کار مجموعه اعتبارسنجی انجام بررسی نهایی از پتانسیل پیش‌بینی مدل است.

انتخاب توصیف‌گرهای مولکولی

نرم‌افزار CORAL براساس توصیف‌گرهای مختلف می‌تواند سه نوع مدل را ایجاد کند، (الف) مدل‌های QSTR براساس توصیف‌گرهای SMILES، (ب) مدل‌های QSTR براساس توصیف‌گرهای گراف مولکولی و (ج) مدل‌های QSTR براساس ترکیبی از توصیف‌گرهای SMILES و گراف مولکولی. هم‌چنین سه نوع گراف مولکولی در این نرم‌افزار می‌تواند ارائه شود؛ ۱- گراف مولکولی با هیدروژن (HFG)، ۲- گراف مولکولی بدون هیدروژن (HSG) و ۳- گراف اوربیتال‌های اتمی (GAO) (Das et al., 2023; Ghaedi, 2015). در این کار از توصیف‌گرهای ترکیبی بدون هیدروژن برای ایجا مدل استفاده شده است.

SMILES یا گراف مولکولی صورت جداگانه هستند (Ghiasi et al., 2021; Lotfi et al., 2021). برای اعمال شرایطی که نقطه پایانی در آن شرایط اندازه‌گیری شده، به‌عنوان مثال، دما، فشار، یا هر شرط دیگری مانند نوع پرند در این مطالعه به‌جای استفاده از SMILES به‌تنهایی از ترکیب SMILES و شرایط آزمایش استفاده می‌شود که در این صورت به آن quasi-SMILES می‌گویند (Ahmadi & Azimi, 2023).

هدف از این پژوهش، مطالعه رابطه کمی ساختار-فعالیت ۲۴۴ ترکیب از آفت‌کش‌ها به‌منظور پیش‌بینی سمیت آن‌ها علیه پرندگان براساس الگوریتم مونت‌کارلو و استفاده از quasi-SMILES ساختار آن‌ها می‌باشد. در ایجاد مدل‌های QSTR، توصیف‌گرهای بهینه ترکیبی، که ترکیبی از quasi-SMILES و گراف بدون هیدروژن است (HSG) استفاده شده است. برای توسعه مدل‌های QSTR از توابع هدف جدیدی به نام‌های شاخص ایده‌آل همبستگی (IIC)^۱ و شاخص شدت همبستگی (CII)^۲ به‌عنوان معیار قابل پیش‌بینی استفاده می‌شود. هفت تقسیم تصادفی برای ارزیابی قابلیت اطمینان و دقت مدل‌های طراحی شده QSTR اجرا شده است. در نهایت گروه‌های عاملی یا ساختارهای متصل‌شده به ساختار اصلی سموم که مسئول افزایش یا کاهش میزان سمیت آن‌هاست استخراج و برای درک بهتر مکانیسم مورد استفاده قرار گرفت.

مواد و روش‌ها

داده‌ها

داده‌های سمیت آفت‌کش‌ها علیه پنج گونه پرند شامل بلدر چین سفید، اردک، قرقاول، گنجشک خانگی و بلدرچین ژاپنی از مقاله Mukherjee et al. (2021) استخراج شد. از داده‌های تجربی سمیت علیه پرندگان به‌صورت LD₅₀ که میانگین غلظت کشنده را نشان می‌دهد به‌کار برده شد. برای مدل‌سازی QSTR از تابع (-log LD₅₀) pd₅₀ این ترکیبات به‌عنوان متغیر وابسته استفاده شد. ساختار مولکولی آفت‌کش‌های مورد مطالعه توسط نرم‌افزار BIOVIA ۲۰۲۰ رسم شد و صورت نمادهای SMILES تبدیل شد و با اضافه کردن نوع پرند به هر SMILES، quasi-SMILES به‌دست آمد و برای هر یک از پرندگان یک کد در نظر گرفته شد و به SMILES سموم اضافه شد. کدهای مربوطه برای بلدرچین سفید، اردک، قرقاول، گنجشک خانگی و بلدرچین

داده می‌شود. هم‌چنین حضور یا عدم حضور چهار عنصر فلوئور، کلر، برم و ید در ساختار ترکیبات توسط شاخص HALO نشان داده می‌شود. HARD، مجموع BOND، NOSP، HALO و PAIR حضور هر جفت از توصیف‌گرهای مولکولی BOND، NOSP و HALO با هم در SMILES می‌باشد. به‌عنوان مثال، باند دوگانه و کلر، نیتروژن و برم، باند سه‌گانه و فسفر و غیره. $e0_k$ و $e1_k$ اتصالات بسط‌یافته مورگان مرتبه صفر و ۱ هستند. وجود مولکول‌هایی با حلقه‌های پنج ضلعی و شش ضلعی توسط C5 و C6 نشان داده می‌شود. $pt3_k$ و $pt4_k$ نشان‌دهنده تعداد مسیرهای به طول ۳ و ۴ که از رأس k شروع می‌شود، $S3_k$ نشان‌دهنده لایه‌های ظرفیت درجه ۳ که از رأس k شروع می‌شوند، می‌باشند. nn_k ، کد نزدیک‌ترین همسایه‌ها می‌باشد. Type Bird، نوع پرند را نشان می‌دهد که کدهای مربوطه برای بلدرچین سفید، اردک، قرقاول، گنجشک خانگی و بلدرچین ژاپنی، به‌ترتیب [VIRG]، [PLAT]، [DOM]، [COL] و [JAP] در نظر گرفته شد. $CW(X)$ وزن‌های همبستگی ویژگی‌های خاص مبتنی بر SMILES یا گراف مولکولی را نشان می‌دهد. داده‌های عددی آن‌ها توسط الگوریتم مونت‌کارلو به‌دست می‌آید. داده‌های عددی به‌دست‌آمده برای تعیین LD_{50} آفت‌کش‌ها با روش حداقل مربعات توسط مدل تک‌متغیر زیر استفاده می‌شود:

$$pLD_{50} = C_0 + C_1 \times DCW(T, N) \quad \text{(رابطه ۴)}$$

C_0 ضرایب رگرسیون و C_1 شیب محاسبه‌شده با روش حداقل مربعات است. DCW، جمع جبری ضرایب توصیف‌گرهای مولکولی بهینه مشتق‌شده از نمادهای HSG و SMILES می‌باشد. T و N به‌ترتیب ضریب آستانه و تعداد سیکل‌های بهینه‌شده روش مونت‌کارلو را نشان می‌دهد.

بهینه‌سازی مونت‌کارلو

برای ایجاد مدل‌های پیش‌بینی QSTR، چهار تابع هدف TF0، TF1، TF2 و TF3 برای بهینه‌سازی مونت‌کارلو موردبررسی قرار گرفت و نتایج آماری آن‌ها مقایسه شد. معادلات هر یک از توابع هدف در زیر آورده شده است.

تابع هدف اول (TF0)

TF0 با استفاده از روش تعادل همبستگی محاسبه‌شده که با معادله زیر نشان داده شده است:

$$TF_0 = R_{ATRN} + R_{PTRN} - |R_{ATRN} - R_{PTRN}| \times dr_{weight} \quad \text{(رابطه ۵)}$$

توصیف‌گر بهینه مجموع وزن‌های همبستگی ویژگی‌های مولکولی مستخرج از SMILES و HSG است. وزن‌های همبستگی از طریق بهینه‌سازی و فراهم کردن بالاترین مقدار تابع هدف که از داده‌های مجموعه آموزشی که شامل آموزش فعال، آموزش غیرفعال و مجموعه‌های کالیبراسیون است محاسبه می‌شوند.

با داشتن داده‌های عددی وزن‌های همبستگی، می‌توان توصیف‌گر بهینه برای هر SMILES را طبق فرمول محاسبه کرد. برای محاسبه توصیف‌گر بهینه ترکیبی، DCW از معادله ریاضی زیر استفاده می‌شود (Ahmadi & Azimi, 2023).

$$HybridDCW(T, N_{epoch}) = SMILES_{DCW}(T, N_{epoch}) + HSG_{DCW}(T, N_{epoch}) \quad \text{(رابطه ۱)}$$

مدل‌های مونت‌کارلو، توابع ریاضی از تعداد سیکل‌های مونت‌کارلو (N_{epoch}) و ضریب آستانه (T) بهینه‌سازی مونت‌کارلو است. قوی‌ترین و پیشگوترین مدل‌ها با بهترین T و N برای ضریب همبستگی مجموعه آزمون خارجی که تابع ریاضی این دو پارامتر است به‌دست می‌آید. براساس جستجوی جامع برای هر مجموعه تصادفی از داده‌ها، با هدف انتخاب توصیف‌گرهای مناسب از بین توصیف‌گرهای ترکیبی چندین مدل مختلف با دسته‌های مختلف توصیف‌گرها ایجاد و بهترین مجموعه توصیف‌گرها انتخاب شدند.

معادله ریاضی زیر توصیف‌گرهای بهینه مولکولی را برای مدل pLD_{50} آفت‌کش‌ها نشان می‌دهد:

$$DCW(T, N) = \sum CW(S_k) + \sum CW(SS_k) + \sum CW(SSS_k) + CW(BOND) + CW(NOSP) + CW(HALO) + CW(HARD) + CW(C5) + CW(C6) + \sum CW(e0_k) + \sum CW(e1_k) + \sum CW(pt3_k) + \sum CW(pt4_k) + \sum CW(nn_k) + \sum CW(S3_k) + CW(Bird Type) \quad \text{(رابطه ۲)}$$

نمادهای S_k ، SS_k و SSS_k ویژگی SMILES هستند که ترکیبی از قطعات مولکولی تک‌عضوی، دو‌عضوی و سه‌عضوی نماد SMILES را نمایش می‌دهند. در معادله (۲)، حضور و یا عدم حضور پیوندهای مختلف دوگانه (=)، سه‌گانه (#) و استروشمیایی (@)، (@@) توسط شاخص BOND نشان داده می‌شود و حضور و یا عدم حضور چهار عنصر نیتروژن، اکسیژن، گوگرد و فسفر در ساختار ترکیبات توسط شاخص NOSP نشان

CII نشان‌دهنده شاخص شدت همبستگی می‌باشد و با استفاده از داده‌های مجموعه کالیبراسیون برای بهبود کیفیت بهینه‌سازی مونت کارلو برای ایجاد مدل‌های QSTR/QSAR محاسبه شده است. CII با معادله‌های ریاضی زیر محاسبه می‌شود.

$$CII_{CAL} = 1 - \sum Protest_K \quad (\text{رابطه } ۱۲)$$

$$Protest_K = \begin{cases} R_K^2 - R^2, & \text{if } R_K^2 - R^2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

R^2 ضریب تعیین برای مجموعه‌ای است که حاوی n ماده است و R_{2k} ضریب همبستگی $n-1$ ماده برای مجموعه پس از حذف k th ماده است.

تابع هدف چهارم (TF3)

در مورد TF3، از هر دو معیار شاخص شدت همبستگی (CII) و هم شاخص ایده‌آل همبستگی (IIC) استفاده شده است که توسط معادله زیر محاسبه می‌شود:

$$TF_3 =$$

$$TF_0 + IIC_{CAL} \times IIC_{weight} + CII_{CAL} \times CII_{weight}$$

IIC و CII دو شاخص هستند که برای قضاوت در مورد پتانسیل پیش‌بینی مدل‌های توسعه‌یافته QSTR به کار می‌رود و به کاربردن هم‌زمان آن‌ها نتایج بهتری نشان داده است. CII_{weight} و IIC_{weight} مقادیر ثابتی هستند که به صورت تجربی به دست می‌آیند.

دامنه کاربرد مدل^۱

بررسی دامنه کاربرد (AD) فرد را قادر می‌سازد تا پیش‌بینی یک ماده شیمیایی را بسته به این که چقدر با مولکول‌های استفاده شده در آموزش شباهت دارد را در توسعه مدل ارزیابی کند و به محدوده محیط شیمیایی مدل توسعه یافته QSTR/QSAR اشاره دارد. هر مدلی که استفاده می‌شود باید بتواند فضای توصیف‌گری را که در آن عمل می‌کند، توصیف کند. حضور توصیف‌گر مولکولی منحصر به فرد نادر، منجر به افزایش احتمال پرت بودن یک ترکیب می‌شود. از این نقطه نظر هر ترکیب می‌تواند بالقوه یک داده پرت باشد. با توجه به تفاوت توزیع یک توصیف‌گر مولکولی در مجموعه‌های آموزش و کالیبراسیون، می‌توان «نقص»^۲ یک توصیف‌گر مولکولی را

R_{PTRN} و R_{ATRN} ضرایب همبستگی بین نقاط پایانی مشاهده شده pLD_{50} و پیش‌بینی شده برای مجموعه‌های آموزش فعال و غیرفعال هستند. dr_{weight} به طور تجربی به دست می‌آید و مقدار ثابتی دارد.

تابع هدف دوم (TF1)

از شاخص ایده‌آل بودن همبستگی برای محاسبه TF1 با اضافه کردن وزن IIC به TF0 استفاده شد. از معادله زیر برای نشان دادن TF1 استفاده می‌شود.

$$TF_1 =$$

$$TF_0 + IIC_{CAL} \times \text{Weight For IIC (IIC}_{weight})$$

IIC نمایانگر شاخص ایده‌آل بودن همبستگی مجموعه کالیبراسیون است و با معادله زیر محاسبه می‌شود.

$$IIC_{CAL} = R_{CAL} \times \frac{\min(-MAE_{CAL}, +MAE_{CAL})}{\max(-MAE_{CAL}, +MAE_{CAL})} \quad (\text{رابطه } ۷)$$

R_{CAL} ضریب همبستگی بین مقادیر تجربی و محاسبه شده pLD_{50} برای مجموعه کالیبراسیون می‌باشد و MAE نشان‌دهنده میانگین خطای مطلق است که می‌تواند با معادله‌های ریاضی زیر محاسبه شود:

$$-MAE_{CLB} =$$

$$-\frac{1}{N} \sum_{y=1}^N |\Delta_k| \quad \Delta_k < 0, \quad N \text{ is the number of } \Delta_k < 0$$

$$+MAE_{CLB} =$$

$$+\frac{1}{N} \sum_{y=1}^N |\Delta_k| \quad \Delta_k \geq 0, \quad N \text{ is the number of } \Delta_k \geq 0$$

$$\Delta_k = Obs_k - Calc_k \quad (\text{رابطه } ۱۰)$$

در اینجا، شاخص K از ۱ تا N می‌تواند متغیر باشد و Obs_k و $Calc_k$ مقادیر تجربی و پیش‌بینی شده نقطه پایانی (pLD_{50}) می‌باشد.

تابع هدف سوم (TF2)

در مورد TF2، معیار شاخص شدت همبستگی (CII)، به جای IIC استفاده شده است. تابع هدف که در اینجا با CII محاسبه شده با معادله زیر نشان داده می‌شود.

$$TF_2 =$$

$$TF_0 + CII_{CAL} \times \text{Weight For CII (CII}_{weight})$$

اعتبارسنجی خارجی با استفاده از دسته آزمون و با ارزیابی آماره‌های $Q_{F1}^2, Q_{F2}^2, Q_{F3}^2, Cr_p^2, MAE, s, \Delta r_m^2$ و $\bar{r}_m^2$ انجام می‌شود. آزمون تصادفی‌سازی برای بررسی قدرتمندی مدل به کار می‌رود. برای یک مدل QSTR مناسب، میانگین ضریب همبستگی مدل‌های تصادفی ایجادشده (R_r) باید کم‌تر از ضریب همبستگی مدل‌های غیر تصادفی (R) باشد. رابطه ریاضی پارامترهای آماری ذکرشده در مقالات قبلی ارائه شده است (Ahmadi et al., 2022).

بحث و بررسی یافته‌ها

مدل‌های QSTR

در این کار، برای به دست آوردن نتایج آماری قابل اعتماد، هفت مدل برای هر نوع تابع هدف (TF0, TF1, TF2, TF3) با استفاده از توصیف‌گرهای ترکیبی بهینه ایجادشده است. توصیف‌گرهای بهینه با استفاده از وزن‌های همبستگی (DCW) از طریق بهینه‌سازی مونت کارلو برای هر تابع هدف محاسبه شدند. نتایج آماری مدل‌های CORAL محاسبه‌شده با TF0, TF1, TF2 و TF3 در جدول (۱) نشان داده شده است.

برای ایجاد مدل‌های QSTR با تابع هدف TF0 در نرم‌افزار کورال روش توازن همبستگی اجرا شد ($W_{IIC} = W_{CII} = 0$) و برای توسعه مدل‌های مبتنی بر تابع هدف TF1 ($W_{CII} = 0$) و $W_{IIC} = 0.5$ ، روش توازن همبستگی با افزودن تنها IIC به TF0 اجرا شد و نتایج نشان داد که افزودن IIC به TF0 کیفیت آماری R^2 و RMSE را برای مجموعه کالیبراسیون و اعتبارسنجی بهبود می‌بخشد. همچنین برای ایجاد مدل‌های مبتنی بر تابع هدف TF2 ($W_{IIC} = 0$ & $W_{CII} = 0.3$) فقط CII به روش توازن همبستگی اضافه شد که نتایج جدول (۱) تأیید می‌کند که افزودن CII به TF0 مقدار عددی R^2 را برای همه زیرمجموعه‌ها (آموزش فعال، آموزش غیرفعال مجموعه‌های کالیبراسیون و اعتبارسنجی) افزایش داد. در نهایت، در مدل بهینه‌سازی‌شده با تابع هدف TF3، IIC و CII به‌طور همزمان به روش توازن همبستگی ($W_{IIC} = 0.5$ و $W_{CII} = 0.3$) اضافه شد. گزارش‌های قبلی ثابت کرده‌اند که با گنجاندن IIC و CII به‌طور همزمان به تابع هدف بهینه‌سازی (TF3)، مدل حاصل می‌تواند کارآمدتر از مدلی باشد که در تابع هدف آن تنها از IIC یا CII وارد شود (Toropov et al., 2022). بنابر این، مدل‌های بهبودیافته با TF3 از نظر آماری بر همه مدل‌های دیگر برتری داشتند، پس

برآورد کرد. در حالت ایده‌آل احتمال این که به صورت تصادفی یک توصیف‌گر در مجموعه آموزش قرار بگیرد ($P_T(A_K)$) با احتمالی که به طور تصادفی در مجموعه کالیبراسیون ($P_C(A_K)$) قرار بگیرد یکسان است. این احتمالات با فرمول‌های زیر محاسبه می‌شوند:

$$P_T(A_K) = N_T(A_K)/N_T \quad \text{رابطه ۱۴}$$

$$P_C(A_K) = N_C(A_K)/N_C \quad \text{رابطه ۱۵}$$

$N_T(A_K)$ تعداد ترکیباتی که دارای توصیف‌گر A_K در دسته آموزش و N_T تعداد کل ترکیبات در دسته آموزش می‌باشد. $N_C(A_K)$ تعداد ترکیباتی که دارای توصیف‌گر A_K در دسته کالیبراسیون هستند و N_C ، تعداد کل ترکیبات در دسته کالیبراسیون می‌باشد. با این فرض پرت‌بودن یک مولکول را می‌توان به صورت زیر تعریف کرد:

$$\text{Defect}(AF_K) = \frac{|P_T(A_K) - P_C(A_K)|}{N_T(A_K) + N_C(A_K)} \quad \text{رابطه ۱۶}$$

می‌توان با داشتن $\text{Defect}(A_K)$ برای هر توصیف‌گر، شاخص

D را برای همه مولکول‌ها محاسبه کرد:

$$D = \sum_{K=1}^{NF} \text{Defect}(A_K) \quad \text{رابطه ۱۷}$$

NF، تعداد توصیف‌گرهای فعال (غیربلاک) هر ترکیب است.

یک مولکول به عنوان داده پرت شناخته می‌شود اگر:

$$D > 2\bar{D} \quad \text{رابطه ۱۹}$$

$\bar{D}$ ، میانگین D برای مولکولی است که در دسته یادگیری است. توزیعی که مولکول‌های آن دارای D کوچک است بهتر از توزیعی است که مولکول‌های آن دارای D بزرگ باشد. مولکول‌های با D بزرگ را مولکول‌های مشکوک می‌نامند که می‌توانند در دسته داده‌های پرت قرار گیرند.

اعتبارسنجی مدل

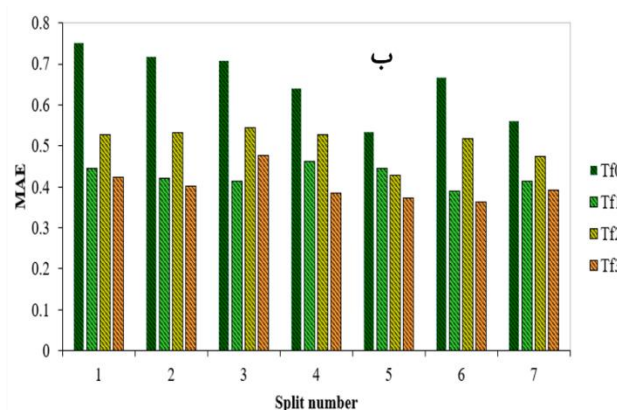
هدف اصلی مدل محاسباتی ایجاد یک مدل قابل اطمینان و دارای ظرفیت پیش‌گویی ویژگی ترکیبات جدید می‌باشد. سه روش اصلی برای اعتبارسنجی مدل وجود دارد. (الف) اعتبارسنجی متقابل (اعتبارسنجی داخلی)^۱، (ب) اعتبارسنجی خارجی و (ج) تصادفی سازی داده‌ها^۲.

اعتبارسنجی متقابل تک به تک^۳ به عنوان اعتبارسنجی داخلی با ارزیابی آماره‌های مختلف CCC، IIC، R^2 ، Q^2 و F حاصل از مدل‌های ایجادشده برای دسته آموزش انجام می‌شود. به‌منظور

1. Cross validation
2. Data randomization or Y-scrambling
3. Leave-one-out cross validation

مقالات پیشین موافقت دارند. بررسی نتایج مقدار نقص^۱ نشان داد که بیش از ۸۷ درصد ترکیبات در دامنه کاربرد مدل‌ها قرار دارند که نشان‌دهنده آن است که مدل‌های پیش‌بینی طراحی شده در این مطالعه، توانایی پیش‌بینی بیش از ۸۷ درصد از داده‌های جدید را دارند.

نمایش گرافیکی داده‌های به‌دست‌آمده pLD₅₀ در مقابل مقادیر محاسباتی pLD₅₀ برای مدل‌های به‌دست‌آمده از هفت شکاف در شکل (۲) نشان داده شده است. همان‌طور که در شکل می‌توان دید، همبستگی بسیار خوبی بین داده‌های تجربی و محاسبه‌شده در مدل مورد بحث وجود دارد. این نتایج بیان‌کننده قدرتمندی مدل‌های QSTR طراحی شده و برآزش خوب آن‌هاست.



شکل ۱. مقایسه ضرایب تعیین و میانگین خطای مطلق مجموعه اعتبارسنجی شماره ۷ مدل طراحی شده با چهار تابع هدف مختلف

آن‌ها را به یک مدل قابل اعتماد برای پیش‌بینی LD₅₀ تبدیل کرد.

مدل‌های ایجادشده برای pLD₅₀ به‌دست‌آمده با تابع هدف TF3 توسط معادلات زیر ارائه می‌شوند:

$$\text{pLD}_{50} = \text{رابطه ۱۹ شکاف ۱} \\ -4.0546 (\pm 0.0272) + 0.0808 (\pm 0.0005) \times \text{DCW} (7, 35)$$

$$\text{pLD}_{50} = \text{رابطه ۲۰ شکاف ۲} \\ -3.2106 (\pm 0.0230) + 0.0641 (\pm 0.0004) \times \text{DCW} (7, 35)$$

$$\text{pLD}_{50} = \text{رابطه ۲۱ شکاف ۳} \\ -2.2635 (\pm 0.0200) + 0.0632 (\pm 0.0005) \times \text{DCW} (7, 35)$$

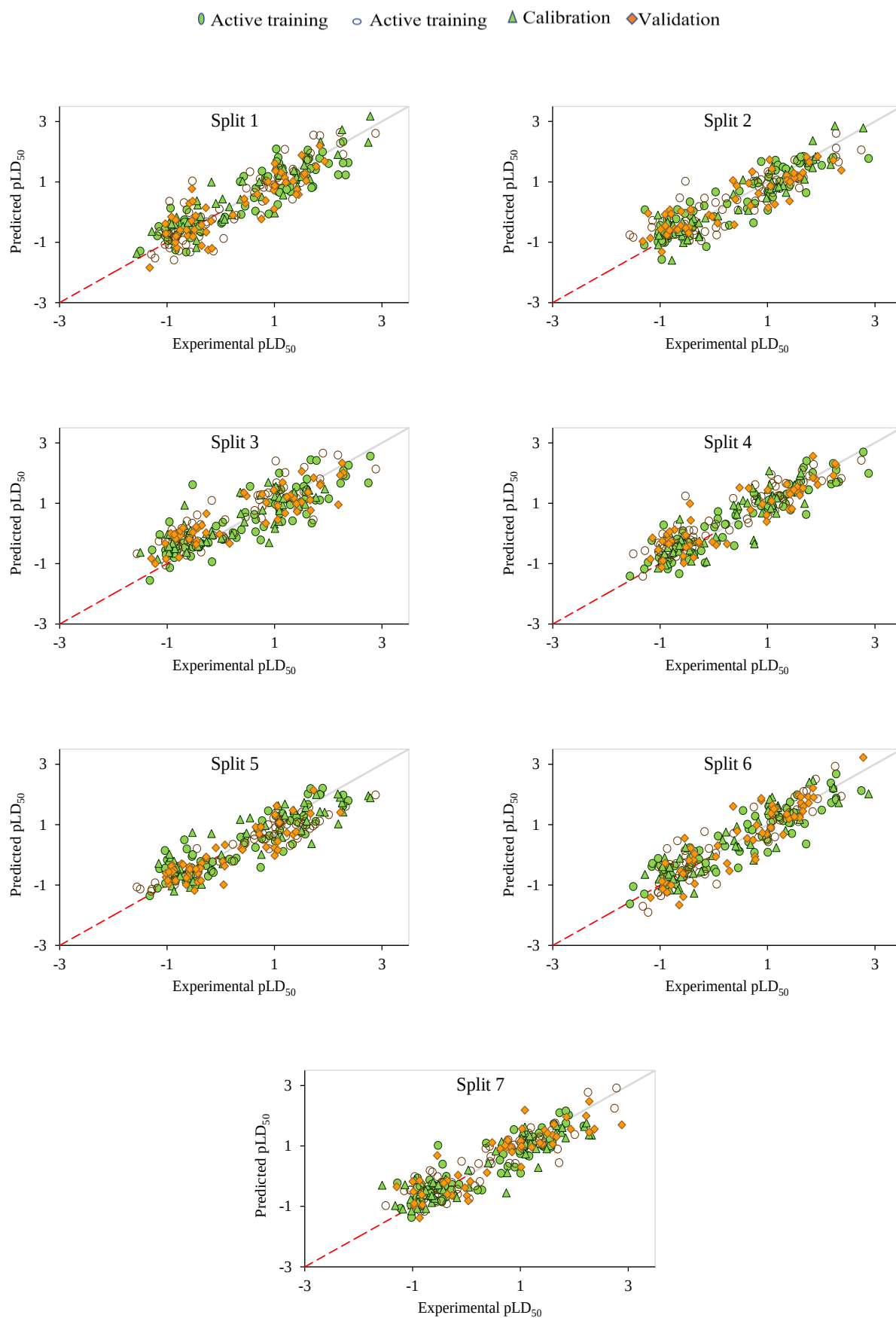
$$\text{pLD}_{50} = \text{رابطه ۲۲ شکاف ۴} \\ -3.7238 (\pm 0.0236) + 0.0765 (\pm 0.0004) \times \text{DCW} (7, 35)$$

$$\text{pLD}_{50} = \text{رابطه ۲۳ شکاف ۵} \\ -3.4817 (\pm 0.0209) + 0.0857 (\pm 0.0005) \times \text{DCW} (7, 35)$$

$$\text{pLD}_{50} = \text{رابطه ۲۴ شکاف ۶} \\ -4.4277 (\pm 0.0254) + 0.0824 (\pm 0.0004) \times \text{DCW} (7, 35)$$

$$\text{pLD}_{50} = \text{رابطه ۲۵ شکاف ۷} \\ -3.6019 (\pm 0.0187) + 0.0740 (\pm 0.0004) \times \text{DCW} (7, 35)$$

تمام مدل‌های ایجادشده با توجه به معیارهای آماری در جدول (۱) رضایت‌بخش هستند، با این حال، توانایی پیش‌بینی مدل‌ها براساس نتایج آماری مجموعه اعتبارسنجی ارزیابی می‌شود. نتایج مدل شماره ۶ با توجه به این که مقدار R² بالاتر و MAE کم‌تر برای مجموعه اعتبارسنجی دارد، نسبت به سایر مدل‌ها بهتر و دارای قدرت پیش‌بینی بالاتری است. مقایسه ضریب تعیین (R²) و میانگین خطای مطلق (MAE) مجموعه اعتبارسنجی برای مدل‌های طراحی شده با چهار تابع هدف برای هفت شکاف در شکل ۱- الف و ب به ترتیب ارائه شده است. هم‌چنین مدل ایجادشده توسط شکاف شش دارای IIC=0.8912 می‌باشد که نسبت به سایر مدل‌ها بیش‌تر است. بنابراین توانایی پیش‌بینی بالاتری دارد. علاوه بر این، مجموعه اعتبارسنجی تمام مدل‌های ایجادشده دارای مقدار عددی R²_m رضایت‌بخش بود که با معیارهای بیان‌شده در



شکل ۲. نمایش گرافیکی همبستگی بین داده‌های تجربی و پیش‌بینی‌شده pLD_{50} برای مدل‌های ۱ تا ۷

جدول ۱. مشخصات آماری برای مدل‌های QSTR مربوط به سمیت آفت‌کش‌ها براساس چهار تابع هدف برای شکاف‌های مختلف

شکاف	تابع هدف	Set	n	R ²	CCC	HC	CII	Q ²	Q ² _{F1}	Q ² _{F2}	Q ² _{F3}	RMSE	MAE	F	Average R ² m	ΔR^2m	Y-Test	مدل QSTR
۱	TF0	آموزش فعال	78	0.980	0.990	0.990	0.985	0.979				0.146	0.095	3662				pLD ₅₀ = -5.816
		آموزش غیر فعال	65	0.977	0.988	0.832	0.984	0.976				0.169	0.120	2672				(0.0109) + 0.1492832
		کالیبراسیون	51	0.338	0.562	0.580	0.726	0.288	0.012	-0.153	-0.015	1.060	0.821	25	0.312	0.007	0.028	(0.0003) × DCW (7, 35)
	TF1	آموزش فعال	78	0.827	0.906	0.780	0.904	0.819				0.425	0.338	364				pLD ₅₀ = -3.5336
		آموزش غیر فعال	65	0.842	0.907	0.726	0.909	0.834				0.439	0.340	336				(0.0222) + 0.0852
		کالیبراسیون	51	0.699	0.829	0.836	0.876	0.676	0.735	0.691	0.728	0.550	0.455	114	0.655	0.003	0.030	(0.0005) × DCW (7, 35)
	TF2	اعتبارسنجی	50	0.781	0.858	0.571	0.879	0.764				0.568	0.445	171	0.696	0.086	0.019	
		آموزش فعال	75	0.965	0.982	0.957	0.979	0.963				0.211	0.160	2026				pLD ₅₀ = -5.7486
		آموزش غیر فعال	67	0.962	0.948	0.696	0.976	0.959				0.395	0.338	1637				(0.0139) + 0.1024
		کالیبراسیون	46	0.893	0.942	0.538	0.969	0.881	0.886	0.885	0.883	0.379	0.321	366	0.831	0.036	0.019	(0.00029) × DCW (7, 35)
	TF3	اعتبارسنجی	51	0.746	0.805	0.497	0.853	0.719				0.714	0.528	144	0.529	0.040	0.018	
		آموزش فعال	75	0.805	0.892	0.668	0.897	0.795				0.499	0.412	302				pLD ₅₀ = -4.0545850
۲	TF0	آموزش غیر فعال	67	0.822	0.903	0.868	0.887	0.812				0.496	0.353	300				(0.0271850) + 0.0807803
		کالیبراسیون	46	0.827	0.906	0.909	0.903	0.813	0.824	0.822	0.820	0.470	0.390	210	0.759	0.065	0.018	(0.0005085) * DCW (7, 35)
		اعتبارسنجی	51	0.722	0.846	0.764	0.841	0.703				0.532	0.423	127	0.667	0.057	0.018	
	TF0	آموزش فعال	73	0.974	0.987	0.651	0.981	0.973				0.176	0.116	2692				pLD ₅₀ = -9.9277
		آموزش غیر فعال	67	0.974	0.981	0.953	0.981	0.972				0.195	0.144	2404				(0.0155) + 0.1277
		کالیبراسیون	47	0.558	0.732	0.696	0.766	0.521	0.337	0.331	0.321	0.892	0.721	57	0.506	0.057	0.022	(0.0002) × DCW (7, 35)
	TF1	اعتبارسنجی	52	0.538	0.705	0.480	0.768	0.501				0.959	0.717	58	0.501	0.017	0.022	
		آموزش فعال	73	0.839	0.913	0.799	0.912	0.830				0.440	0.338	371				pLD ₅₀ = -4.2397548
		آموزش غیر فعال	67	0.813	0.892	0.797	0.884	0.802				0.461	0.364	283				(0.0283) + 0.0687
		کالیبراسیون	47	0.869	0.928	0.932	0.925	0.859	0.851	0.850	0.848	0.423	0.351	299	0.788	0.044	0.020	(0.0004) × DCW (7, 35)
	TF2	اعتبارسنجی	52	0.770	0.866	0.851	0.869	0.755				0.512	0.420	167	0.742	0.053	0.020	
		آموزش فعال	73	0.958	0.979	0.952	0.972	0.956				0.225	0.167	1618				pLD ₅₀ = -6.6666
		آموزش غیر فعال	67	0.958	0.968	0.735	0.971	0.956				0.254	0.198	1481				(0.0186) + 0.1220
۳	TF2	کالیبراسیون	47	0.872	0.926	0.753	0.962	0.860	0.834	0.833	0.830	0.446	0.377	308	0.784	0.043	0.023	(0.0003) × DCW (7, 35)
		اعتبارسنجی	52	0.613	0.776	0.376	0.778	0.584				0.737	0.532	79	0.560	0.085	0.023	
	TF3	آموزش فعال	73	0.813	0.897	0.629	0.890	0.803				0.475	0.363	309				pLD ₅₀ = -3.2106
		آموزش غیر فعال	67	0.784	0.875	0.781	0.884	0.771				0.495	0.405	236				(0.0230) + 0.0641
		کالیبراسیون	47	0.875	0.933	0.936	0.939	0.864	0.868	0.867	0.865	0.398	0.320	316	0.822	0.042	0.021	(0.0004) × DCW (7, 35)
	TF0	اعتبارسنجی	52	0.783	0.866	0.792	0.867	0.768				0.502	0.402	180	0.719	0.058	0.021	
		آموزش فعال	77	0.739	0.850	0.838	0.849	0.725				0.577	0.432	212				pLD ₅₀ = -2.2635
		آموزش غیر فعال	64	0.767	0.821	0.432	0.886	0.751				0.606	0.499	204				(0.0200) + 0.0632
		کالیبراسیون	47	0.641	0.765	0.801	0.807	0.612	0.612	0.610	0.705	0.599	0.465	80	0.482	0.118	0.018	(0.0005) × DCW (7, 35)
	TF1	اعتبارسنجی	51	0.778	0.839	0.679	0.877	0.764				0.564	0.476	172	0.579	0.119	0.018	
		آموزش فعال	77	0.793	0.885	0.742	0.885	0.782				0.513	0.402	288				pLD ₅₀ = -4.4024
		آموزش غیر فعال	64	0.796	0.856	0.498	0.886	0.783				0.561	0.457	242				(0.0352) + 0.0734604
۳	TF1	کالیبراسیون	47	0.682	0.796	0.826	0.798	0.661	0.653	0.651	0.736	0.567	0.416	97	0.522	0.108	0.020	(0.0005) × DCW (7, 35)
		اعتبارسنجی	51	0.799	0.873	0.647	0.876	0.785				0.526	0.413	195	0.641	0.097	0.020	
	TF2	آموزش فعال	77	0.944	0.971	0.768	0.968	0.941				0.268	0.220	1257				pLD ₅₀ = -5.6692
		آموزش غیر فعال	64	0.944	0.946	0.455	0.964	0.940				0.363	0.287	1041				(0.0173028) + 0.1331
		کالیبراسیون	47	0.719	0.807	0.377	0.924	0.694	0.605	0.602	0.699	0.605	0.499	115	0.521	0.112	0.021	(0.0004) × DCW (7, 35)
	TF3	اعتبارسنجی	51	0.742	0.839	0.675	0.854	0.725				0.661	0.545	141	0.620	0.106	0.021	
		آموزش فعال	77	0.739	0.850	0.838	0.849	0.725				0.577	0.432	212				pLD ₅₀ = -2.2635
		آموزش غیر فعال	64	0.767	0.821	0.432	0.886	0.751				0.606	0.499	204				(0.0200) + 0.06322
		کالیبراسیون	47	0.641	0.765	0.801	0.807	0.612	0.612	0.610	0.705	0.599	0.465	80	0.482	0.118	0.018	(0.0005) × DCW (7, 35)
	TF3	اعتبارسنجی	51	0.778	0.839	0.679	0.877	0.764				0.564	0.476	172	0.579	0.119	0.018	

ادامه جدول ۱. مشخصات آماری برای مدل‌های QSTR مربوط به سمیت آفت‌کش‌ها براساس چهار تابع هدف برای شکاف‌های مختلف

شکاف	تابع هدف	Set	n	R ²	CCC	IIC	CII	Q ²	Q ² _{F1}	Q ² _{F2}	Q ² _{F3}	RMSE	MAE	F	Average R ² m	ΔR^2m	Y-Test	مدل QSTR
۴	TF0	آموزش فعال	75	0.989	0.995	0.968	0.992	0.988				0.118	0.071	6560				pLD ₅₀ = -7.5229 (0.0114) + 0.1293 (0.0002) × DCW (7, 35)
		آموزش غیر فعال	66	0.983	0.988	0.742	0.987	0.981				0.169	0.127	3586				
		کالیبراسیون	47	0.319	0.518	0.529	0.701	0.260	-0.582	-0.601	0.027	1.110	0.855	21	0.282	0.044	0.026	
		اعتبارسنجی	51	0.580	0.719	0.374	0.734	0.547				0.933	0.640	68	0.470	0.129	0.026	
	TF1	آموزش فعال	75	0.861	0.925	0.903	0.924	0.854				0.421	0.334	452				pLD ₅₀ = -4.1934 (0.0239) + 0.0811 (0.0004) × DCW (7, 35)
		آموزش غیر فعال	66	0.861	0.899	0.627	0.921	0.852				0.484	0.374	396				
		کالیبراسیون	47	0.687	0.824	0.829	0.817	0.665	0.626	0.621	0.770	0.541	0.423	99	0.647	0.080	0.022	
		اعتبارسنجی	51	0.774	0.869	0.809	0.886	0.757				0.537	0.463	168	0.681	0.078	0.022	
	TF2	آموزش فعال	75	0.962	0.981	0.955	0.978	0.960				0.221	0.165	1842				pLD ₅₀ = -6.0202 (0.0184) + 0.1176 (0.0004) × DCW (7, 35)
		آموزش غیر فعال	66	0.961	0.980	0.963	0.977	0.958				0.220	0.168	1573				
		کالیبراسیون	47	0.639	0.793	0.606	0.917	0.606	0.589	0.584	0.747	0.567	0.487	80	0.564	0.097	0.023	
		اعتبارسنجی	51	0.637	0.784	0.528	0.787	0.609				0.750	0.527	86	0.542	0.112	0.023	
	TF3	آموزش فعال	75	0.856	0.923	0.810	0.921	0.848				0.429	0.330	434				pLD ₅₀ = -3.7237730 (0.0236) + 0.0765 (0.0004) × DCW (7, 35)
		آموزش غیر فعال	66	0.860	0.895	0.552	0.925	0.852				0.487	0.379	394				
		کالیبراسیون	47	0.752	0.859	0.867	0.858	0.733	0.699	0.695	0.815	0.485	0.380	137	0.692	0.071	0.023	
		اعتبارسنجی	51	0.802	0.884	0.885	0.886	0.788				0.515	0.385	198	0.681	0.076	0.023	
۵	TF0	آموزش فعال	76	0.989	0.995	0.763	0.993	0.989				0.108	0.070	6875				pLD ₅₀ = -5.7072 (0.0096) + 0.1031 (0.0002) × DCW (7, 35)
		آموزش غیر فعال	66	0.992	0.995	0.812	0.994	0.992				0.106	0.079	7954				
		کالیبراسیون	46	0.481	0.681	0.602	0.760	0.488	0.316	0.293	0.099	1.020	0.730	41	0.408	0.087	0.018	
		اعتبارسنجی	51	0.593	0.751	0.669	0.766	0.557				0.721	0.534	71	0.578	0.025	0.018	
	TF1	آموزش فعال	76	0.834	0.909	0.913	0.912	0.825				0.428	0.345	372				pLD ₅₀ = -4.8356 (0.0292163) + 0.0807 (7, 35) (0.0005) × DCW
		آموزش غیر فعال	66	0.879	0.896	0.703	0.924	0.872				0.435	0.335	466				
		کالیبراسیون	46	0.798	0.883	0.893	0.868	0.783	0.798	0.792	0.734	0.554	0.425	173	0.760	0.027	0.023	
		اعتبارسنجی	51	0.642	0.797	0.558	0.820	0.613				0.552	0.445	88	0.621	0.020	0.023	
	TF2	آموزش فعال	76	0.963	0.981	0.883	0.980	0.961				0.201	0.158	1947				pLD ₅₀ = -4.0358 (0.0109) + 0.0977 (0.0003) × DCW (7, 35)
		آموزش غیر فعال	66	0.968	0.970	0.833	0.981	0.966				0.243	0.195	1922				
		کالیبراسیون	46	0.784	0.847	0.736	0.940	0.765	0.770	0.763	0.698	0.591	0.481	160	0.642	0.108	0.019	
		اعتبارسنجی	51	0.679	0.820	0.667	0.828	0.654				0.542	0.429	104	0.619	0.035	0.019	
	TF3	آموزش فعال	76	0.801	0.890	0.806	0.887	0.793				0.468	0.361	299				pLD ₅₀ = -3.4817 (0.0209) + 0.0856 (0.0005) × DCW (7, 35)
		آموزش غیر فعال	66	0.893	0.902	0.604	0.936	0.887				0.430	0.338	534				
		کالیبراسیون	46	0.754	0.852	0.868	0.873	0.735	0.759	0.751	0.682	0.606	0.493	135	0.707	0.078	0.020	
		اعتبارسنجی	51	0.760	0.862	0.603	0.894	0.739				0.452	0.372	155	0.692	0.010	0.020	
۶	TF0	آموزش فعال	74	0.983	0.992	0.715	0.988	0.982				0.147	0.097	4243				pLD ₅₀ = -7.8438 (0.0136089) + 0.1478 (0.0003) × DCW (7, 35)
		آموزش غیر فعال	66	0.983	0.985	0.602	0.990	0.982				0.198	0.148	3795				
		کالیبراسیون	46	0.698	0.828	0.684	0.834	0.673	0.605	0.603	0.628	0.677	0.534	102	0.673	0.020	0.021	
		اعتبارسنجی	53	0.585	0.715	0.478	0.761	0.557				0.888	0.667	72	0.513	0.073	0.021	
	TF1	آموزش فعال	74	0.876	0.934	0.887	0.926	0.870				0.401	0.319	508				pLD ₅₀ = -6.1218 (0.0312) + 0.0965 (0.0005) × DCW (7, 35)
		آموزش غیر فعال	66	0.876	0.933	0.717	0.933	0.869				0.397	0.317	453				
		کالیبراسیون	46	0.767	0.875	0.876	0.866	0.747	0.749	0.747	0.763	0.540	0.416	145	0.733	0.048	0.027	
		اعتبارسنجی	53	0.819	0.894	0.836	0.891	0.807				0.498	0.390	231	0.748	0.074	0.027	
	TF2	آموزش فعال	74	0.968	0.984	0.984	0.977	0.966				0.205	0.145	2152				pLD ₅₀ = -6.2982 (0.0139) + 0.1276 (0.0003) × DCW (7, 35)
		آموزش غیر فعال	66	0.968	0.968	0.456	0.982	0.966				0.279	0.219	1910				
		کالیبراسیون	46	0.847	0.907	0.598	0.955	0.831	0.808	0.807	0.819	0.472	0.404	243	0.738	0.043	0.023	
		اعتبارسنجی	51	0.667	0.801	0.660	0.815	0.644				0.678	0.518	102	0.605	0.052	0.023	
	TF3	آموزش فعال	74	0.836	0.911	0.866	0.904	0.829				0.461	0.371	368				pLD ₅₀ = -4.4277 (0.0254) + 0.0824 (0.0004) × DCW (7, 35)
		آموزش غیر فعال	66	0.852	0.919	0.549	0.914	0.844				0.442	0.342	368				
		کالیبراسیون	46	0.806	0.896	0.898	0.913	0.788	0.787	0.786	0.799	0.497	0.409	183	0.775	0.041	0.025	
		اعتبارسنجی	53	0.813	0.891	0.731	0.883	0.800				0.500	0.362	222	0.775	0.068	0.025	

ادامه جدول ۱. مشخصات آماری برای مدل‌های QSTR مربوط به سمیت آفت‌کش‌ها براساس چهار تابع هدف برای شکاف‌های مختلف

شکاف	تابع هدف	Set	n	R ²	CCC	IIc	CII	Q ²	Q ² _{Fi}	Q ² _{Fe}	Q ² _{Fi}	RMSE	MAE	F	Average R ² m	ΔR ² m	Y-Test	مدل QSTR
TF0	آموزش فعال	80	0.981	0.990	0.990	0.985	0.980	0.980	0.319	0.302	0.144	0.142	0.081	3930				pLD ₅₀ = -5.2964 (0.0062) + 0.1200 (0.0001) × DCW (7, 35)
	آموزش غیر فعال	62	0.981	0.921	0.400	0.984	0.980	0.980	0.390	0.302	0.144	0.374	0.299	3030				
	کالیبراسیون	48	0.438	0.658	0.651	0.721	0.390	0.390	0.390	0.302	0.144	0.959	0.720	36	0.379	0.072		
	اعتبارسنجی	49	0.549	0.740	0.449	0.781	0.517	0.517	0.390	0.302	0.144	0.768	0.561	57	0.476	0.039	0.026	
TF1	آموزش فعال	80	0.860	0.925	0.721	0.920	0.854	0.854	0.685	0.677	0.604	0.382	0.298	479				pLD ₅₀ = -5.0191 (0.0227) + 0.0859 (0.0004) × DCW (7, 35)
	آموزش غیر فعال	62	0.860	0.926	0.858	0.928	0.851	0.851	0.662	0.677	0.604	0.393	0.322	369				
	کالیبراسیون	48	0.696	0.830	0.835	0.832	0.662	0.662	0.662	0.677	0.604	0.652	0.484	106	0.654	0.076		
	اعتبارسنجی	49	0.756	0.869	0.848	0.867	0.733	0.733	0.662	0.677	0.604	0.551	0.414	146	0.708	0.017	0.026	
TF2	آموزش فعال	80	0.962	0.981	0.933	0.976	0.960	0.960	0.844	0.840	0.804	0.200	0.140	1958				pLD ₅₀ = -4.9039 (0.0123) + 0.1090 (0.0003) × DCW (7, 35)
	آموزش غیر فعال	62	0.962	0.961	0.908	0.972	0.959	0.959	0.844	0.840	0.804	0.279	0.229	1508				
	کالیبراسیون	48	0.851	0.919	0.617	0.954	0.838	0.838	0.844	0.840	0.804	0.459	0.395	263	0.790	0.044		
	اعتبارسنجی	49	0.715	0.841	0.811	0.857	0.690	0.690	0.844	0.840	0.804	0.615	0.474	118	0.655	0.104	0.029	
TF3	آموزش فعال	80	0.821	0.902	0.906	0.896	0.813	0.813	0.812	0.808	0.764	0.433	0.336	357				pLD ₅₀ = -3.6019 (0.0187) + 0.0740 (0.0004) × DCW (7, 35)
	آموزش غیر فعال	62	0.825	0.905	0.834	0.903	0.815	0.815	0.812	0.808	0.764	0.442	0.348	283				
	کالیبراسیون	48	0.817	0.887	0.904	0.894	0.803	0.803	0.812	0.808	0.764	0.504	0.380	205	0.789	0.013		
	اعتبارسنجی	49	0.784	0.881	0.854	0.893	0.765	0.765	0.812	0.808	0.764	0.509	0.393	171	0.738	0.057	0.026	

مقایسه مدل ایجادشده با مدل‌های پیشین

جستجوی جامع در انتشارات معتبر علمی در مورد مدل‌سازی QSTR به منظور پیش‌بینی سمیت حاد خوراکی آفت‌کش‌ها، نشان می‌دهد مقالات زیادی در مورد سمیت آفت‌کش‌ها علیه انواع گوناگون پرندگان هر ساله منتشر شده است. اما در میان آن‌ها تنها چند مطالعه مشاهده می‌شود که در مورد پرندگان مورد مطالعه در این کار، مدل‌سازی انجام شده است. در این مطالعات تأثیر سمیت آفت‌کش‌ها بر روی تنها ۱، ۲، ۳ یا ۴ پرنده بررسی شده (Podder *et al.*, 2023; Zhang, Cheng, 2021; Banjare *et al.*, 2015; *et al.*) که برای هر پرنده به شکل انفرادی مدل شده است و تاکنون در مورد مطالعه‌ی همزمان پنج پرنده مورد مطالعه در این کار، مقاله‌ای منتشر نشده است. در کنار این مطالعات تنها یک مطالعه وجود دارد که بر روی این پنج پرنده بررسی مدل‌سازی سمیت انجام شده است که آن هم تأثیر آفت‌کش‌ها برای هر پرنده به‌طور جداگانه مدل شده است (Basant *et al.*, 2015). Basant *et al.* (2015) مدل‌سازی QSTR را برای ۱۳۱ آفت‌کش علیه بلدرچین سفید شمالی، ۴۶ آفت‌کش علیه اردک، ۲۷ آفت‌کش علیه قرقاول گردن حلقه‌ای، ۱۹ آفت‌کش علیه بلدرچین ژاپنی، نه آفت‌کش علیه گنجشک خانگی به‌طور انفرادی انجام دادند.

بنابراین به دلیل تفاوت در تعداد پرندگان مورد مطالعه و هم‌چنین تفاوت در مجموعه‌های آموزشی و آزمون، مقایسه مستقیم امکان‌پذیر نیست، با این وجود در کنار تفاوت در رویکرد مدل‌سازی، ما سعی کرده‌ایم مطالعه حاضر را با مطالعات منتشرشده قبلی ارزیابی کنیم. در این مطالعه، یک مدل QSTR جامع برای هر پنج پرنده ایجادشده است و نسبت به مدل‌های قبلی برتری دارد، در اینجا کل مجموعه داده‌ها به چهار مجموعه (آموزش، آموزش نامرئی، مجموعه کالیبراسیون و اعتبارسنجی) و هفت شکاف تصادفی تقسیم شده، درحالی‌که در گزارش‌های قبلی، مجموعه داده به یک یا دو مجموعه (مجموعه آموزشی و آزمون) و یک شکاف تقسیم شده‌اند. در اینجا چهار تابع هدف (TF0, TF1, TF2, TF3) برای ایجاد مدل‌های QSTR به کار گرفته شده و پارامترهای آماری این مدل‌ها با استفاده از سری آزمون مورد ارزیابی و مقایسه قرار گرفته است. در مطالعه حاضر تنها یک متغیر DCW، که جمع جبری ضرایب توصیف‌گرهای افزاینده و کاهنده می‌باشد، برای تولید مدل‌های QSTR استفاده شده است. تأثیر دو تابع هدف مهم و جدید (شاخص همبستگی ایده‌آل (IIC) و شاخص شدت همبستگی (CII)) نیز در این مطالعه مورد پژوهش و بررسی قرار گرفته‌اند که در کار قبلی مورد مطالعه قرار نگرفته‌اند.

برای همه اجراها منفی است، ج) ویژگی‌های ساختاری تعریف نشده، توصیف‌گرهایی که وزن همبستگی آن‌ها در سه یا چند اجرا مثبت و منفی باشد.

ویژگی‌های کلیدی quasi-SMILES که مسئول افزایش یا کاهش pLD_{50} هستند، با تجزیه و تحلیل وزن‌های همبستگی به‌دست‌آمده در پنج مرحله بهینه‌سازی مونت‌کارلو برای شکاف‌های ۶ و ۷ که نتایج آماری بهتری داشتند شناسایی شدند. نمونه‌هایی از ویژگی ساختاری افزایشنده و کاهشنده مقدار pLD_{50} در جدول (۲) نشان داده شده است.

از جمله توصیف‌گرهای افزایشنده pLD_{50} مشترک در شکاف ۶ و ۷، توالی کربن آلیفاتیک، باند دوگانه و شاخه‌دار شدن $(C...=...C...)$ ، وجود نیتروژن آلیفاتیک و کربن آلیفاتیک با شاخه‌دار شدن $(N...C...)$ ، ترکیبی از اکسیژن، شاخه‌دار شدن و نیتروژن غیرآروماتیک $(O...N...)$ ، ترکیبی از فسفر، اکسیژن و کربن آلیفاتیک $(P...O...C...)$. هم‌چنین توصیف‌گرهای کاهشنده pLD_{50} مشترک در این دو شکاف عبارتند از: وجود پیوند دوگانه $(=.....)$ ، ترکیب متوالی کربن آلیفاتیک، اکسیژن و کربن آلیفاتیک $(C...O...C...)$ ، ترکیب یک حلقه، نیتروژن آلیفاتیک و شاخه‌دار شدن $(1...N...)$.

هم‌چنین، تفسیر مکانیکی مدل پیشنهادی در مقایسه با سایر مدل‌های گزارش شده برای pLD_{50} آفت‌کش‌ها بسیار آسان است. تمام پارامترهای آماری در گستره قابل قبول قرار دارند بنابر این، مدل‌های QSTR ارائه شده، مدل‌های جامع، قوی‌تر و قابل پیش‌بینی‌تر هستند.

تفسیر مدل

همان‌طور که در اصل پنجم سازمان توسعه و همکاری اقتصادی (OECD) ذکر شده است، هدف از تفسیر مکانیکی بررسی تعاملات مکانیکی بین توصیف‌گرهای مورد استفاده در مدل‌سازی QSTR و هم‌چنین نقطه پایانی است و این تعاملات باید در صورت امکان آشکار شود. توصیف‌گرها یا ویژگی‌های ساختاری (SA_k) به‌دست آمده از مدل‌های QSTR برای تفسیر مکانیکی استفاده می‌شوند. با داشتن مقادیر وزن همبستگی توصیف‌گرها در چندین اجرا از بهینه‌سازی مونت‌کارلو، می‌توان سه دسته از توصیف‌گرهای مولکولی quasi-SMILES و HSG استخراج کرد؛ الف) ویژگی‌های ساختاری افزایشنده یعنی توصیف‌گرهایی که وزن همبستگی این توصیف‌گرها برای همه اجراها مثبت است، ب) ویژگی‌های ساختاری کاهشنده یعنی توصیف‌گرهایی که وزن همبستگی

جدول ۲. توصیف‌گرهای مولکولی افزایشنده و کاهشنده pLD_{50} مشترک در شکاف‌های ۶ و ۷ به‌عنوان مدل‌های برتر

شماره	توصیف‌گر	شماره شکاف	وزن همبستگی در اجرای اول تا پنجم	N_{AT}^a	N_{PT}^b	N_C^c	نقص	توضیحات
توصیف‌گرهای افزایشنده								
۱	$C...=...C...$	6	0.1732	0.3429	0.8612	1.1486	1.4390	توالی کربن آلیفاتیک، باند دوگانه و شاخه‌دار شدن
		7	0.5415	0.6899	1.0311	0.9857		
۲	$N...C...C...$	6	0.5957	1.0748	1.0452	1.1669	0.1041	وجود نیتروژن آلیفاتیک و کربن آلیفاتیک با شاخه‌دار شدن
		7	0.3406	0.8897	0.2958	1.4324	0.1095	
۳	$O...N...$	6	0.3494	0.8926	0.6072	1.2715	1.7071	ترکیبی از اکسیژن، شاخه‌دار شدن و نیتروژن غیر آروماتیک
		7	1.0664	0.4444	0.1526	1.1619	0.8404	
۴	$P...O...C...$	6	2.4024	0.7099	0.2290	1.9617	1.4962	ترکیبی از فسفر، اکسیژن و کربن آلیفاتیک
		7	2.9905	2.0150	1.8877	3.3121	0.9540	
توصیف‌گرهای کاهشنده								
۱	$=.....$	6	-0.3349	-0.2676	-0.1593	-0.3516	-0.4689	وجود پیوند دوگانه
		7	-0.3396	-0.2265	-0.3709	-0.4431	-0.4406	
۲	$C...O...C...$	6	-1.1142	-0.7705	-0.8112	-1.2076	-0.2536	ترکیب متوالی کربن آلیفاتیک، اکسیژن و کربن آلیفاتیک
		7	-0.0735	-0.3633	-0.222	-0.7109	-0.0185	
۳	$1...N...C...$	6	-0.6025	-0.3482	-0.2085	-0.5031	-0.5011	ترکیب یک حلقه، نیتروژن آلیفاتیک و شاخه‌دار شدن
		7	-0.6217	-0.0045	-0.6167	-0.4660	-0.4543	

a, b, c فرکانس ویژگی‌های مولکولی SMILES در آموزش فعال، آموزش غیر فعال، و کالبراسیون، به ترتیب

نتیجه‌گیری

مقدار نسبت به سایر شکاف‌ها بود، بنابراین به‌عنوان بهترین مدل انتخاب شد. با استفاده از پارامترهای اعتبارسنجی داخلی و خارجی و آزمون تصادفی‌سازی Y ، قابلیت پیش‌بینی و قابلیت اطمینان مدل‌های QSTR مورد ارزیابی قرار گرفت. همه مدل‌ها از نظر آماری رضایت‌بخش و قوی بودند و با معیارهای آماری پیشنهادی موافقت داشتند. به‌منظور شناسایی نقاط پرت در مدل‌های QSTR ساخته‌شده، دامنه کاربرد (AD) به کار گرفته شد. تفسیر مکانیکی بهترین مدل‌های QSTR از نظر آماری (شکاف‌های ۶ و ۷)، با شناسایی عوامل افزایشنده و کاهشنده pLD_{50} انجام شد و از این موارد به‌منظور طراحی یا پیش‌بینی سمیت مولکول‌های جدید می‌تواند استفاده شود.

تعارض منافع

هیچ‌گونه تعارض منافی توسط نویسندگان وجود ندارد.

References

- Ahmadi, S., & Azimi, N. (2023). Quasi-SMILES-Based QSPR/QSAR Modeling. In *QSPR/QSAR Analysis Using SMILES and Quasi-SMILES* (pp. 191-210): Springer.
- Ahmadi, S., Ghanbari, H., Lotfi, S., & Azimi, N. (2021). Predictive QSAR modeling for the antioxidant activity of natural compounds derivatives based on Monte Carlo method. *Molecular diversity*, 25, 87-97.
- Ahmadi, S., Lotfi, S., & Kumar, P. (2022). Quantitative structure-toxicity relationship models for predication of toxicity of ionic liquids toward leukemia rat cell line IPC-81 based on index of ideality of correlation. *Toxicology Mechanisms and Methods*, 32(4), 302-312.
- Aktar, M. W., Sengupta, D., & Chowdhury, A. (2009). Impact of pesticides use in agriculture: their benefits and hazards. *Interdisciplinary toxicology*, 2(1), 1.
- Basant, N., Gupta, S., & Singh, K. P. (2015). Predicting toxicities of diverse chemical pesticides in multiple avian species using tree-based QSAR approaches for regulatory purposes. *Journal of Chemical Information and Modeling*, 55(7), 1337-1348.
- Benfenati, E., Toropov, A. A., Toropova, A. P., Manganaro, A., & Gonella Diaza, R. (2011). CORAL software: QSAR for anticancer agents. *Chemical biology & drug design*, 77(6), 471-476.
- Cronin, M. T., Jaworska, J. S., Walker, J. D., Comber, M. H., Watts, C. D., & Worth, A. P. (2003). Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environmental health perspectives*, 111(10), 1391-1401.
- Das, N. R., Sharma, T., Mallick, A., Toropova, A. P., Toropov, A. A., & Achary, P. (2023). Computational Approach in Designing and Development of Novel Inhibitors of AKR1C1. In *Ambient Intelligence in Health Care* (pp. 325-337): Springer.
- Ghaedi, A. (2015). Predicting the cytotoxicity of ionic liquids using QSAR model based on SMILES optimal descriptors. *Journal of Molecular Liquids*, 208, 269-279.
- Ghiasi, T., Ahmadi, S., Ahmadi, E., Talei Babil Olyai, M., & Khodadadi, Z. (2021). The index of ideality of correlation: QSAR studies of hepatitis C virus NS3/4A protease inhibitors using SMILES descriptors. *SAR and QSAR in Environmental Research*, 32(6), 495-520.
- Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., ... Hörrén, T. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PloS one*, 12(10), e0185809.
- Hill, E. F., Camardese, M. B., Heinz, G. H., Spann, J. W., & Debevec, A. B. (1984). Acute toxicity of diazinon is similar for eight stocks of bobwhite. *Environmental Toxicology and Chemistry: An International Journal*, 3(1), 61-66.
- Hilton, G. M., Odenkirchen, E., Panger, M., Waleko, G., Lowit, A., & Clippinger, A. J. (2019). Evaluation of the avian acute oral and sub-acute dietary toxicity test for pesticide registration. *Regulatory Toxicology and Pharmacology*, 105, 30-35.
- Humann-Guillemot, S., de Montaigne, C. T., Sire, J., Grünig, S., Gning, O., Glauser, G., ... Helfenstein, F. (2019). A sublethal dose of the neonicotinoid insecticide acetamiprid reduces sperm density in a songbird. *Environmental research*, 177, 108589.
- Jaworska, J. S., Comber, M., Auer, C., & Van Leeuwen, C. (2003). Summary of a workshop on regulatory acceptance of (Q) SARs for human health and environmental endpoints. *Environmental health perspectives*, 111(10), 1358-1360.
- Jeschke, P., Krämer, W., Schirmer, U., & Witschel, M. (2013). *Modern methods in crop protection research*: John Wiley & Sons.

- Kumar, P., & Kumar, A. (2018). Monte Carlo method based QSAR studies of Mer kinase inhibitors in compliance with OECD principles. *Drug Research*, 68(04), 189-195.
- Lotfi, S., Ahmadi, S., & Kumar, P. (2021). The Monte Carlo approach to model and predict the melting point of imidazolium ionic liquids using hybrid optimal descriptors. *RSC advances*, 11(54), 33849-33857.
- Lotfi, S., Ahmadi, S., & Kumar, P. (2022). Correction: Ecotoxicological prediction of organic chemicals toward *Pseudokirchneriella subcapitata* by Monte Carlo approach. *RSC advances*, 12(53), 34567-34567.
- Mitra, A., Chatterjee, C., & Mandal, F. B. (2011). Synthetic chemical pesticides and their effects on birds. *Res J Environ Toxicol*, 5(2), 81-96.
- Mukherjee, R. K., Kumar, V., & Roy, K. (2021). Ecotoxicological QSTR and QSTTR modeling for the prediction of acute oral toxicity of pesticides against multiple avian species. *Environmental Science & Technology*, 56(1), 335-348.
- Podder, T., Kumar, A., Bhattacharjee, A., & Ojha, P. K. (2023). Exploring Regression-based QSTR and i-QSTR Modeling for Ecotoxicity Prediction of Diverse Pesticides on Multiple Avian Species. *Environmental Science: Advances*.
- Singh, K. P., Gupta, S., Basant, N., & Mohan, D. (2014). QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. *Chemical Research in Toxicology*, 27(9), 1504-1515.
- Toropov, A. A., & Toropova, A. P. (2018). Application of the Monte Carlo method for building up models for octanol-water partition coefficient of platinum complexes. *Chemical Physics Letters*, 701, 137-146.
- Toropov, A. A., Toropova, A. P., Achary, P. G. R., Raškova, M., & Raška, I. (2022). The searching for agents for Alzheimer's disease treatment via the system of self-consistent models. *Toxicology Mechanisms and Methods*, 32(7), 549-557.
- Toropova, A. P., & Toropov, A. A. (2014). CORAL software: prediction of carcinogenicity of drugs by means of the Monte Carlo method. *European Journal of Pharmaceutical Sciences*, 52, 21-25.
- Zhu, H., Tropsha, A., Fourches, D., Varnek, A., Papa, E., Gramatica, P., ... Tetko, I. V. (2008). Combinatorial QSAR modeling of chemical toxicants tested against *Tetrahymena pyriformis*. *Journal of chemical information and modeling*, 48(4), 766-784.