

نمذجة العلاقة الكمية بين التركيب والسمية المدفوعة بالتعلم الآلي لتقييم السمية الحادة لبعض مبيدات الفطريات المثبطة للتنفس وتخليق الإرغوستيرول

د. محسن عبود*

* مدرس، قسم وقاية النبات، كلية الهندسة الزراعية، جامعة حمص.

المخلص

تُشكل المبيدات الفطرية من مثبطات التنفس وتخليق الإرغوستيرول مجموعة مهمة من مبيدات الفطريات على نطاق عالمي. بسبب التطبيق المكثف لهذه المواد في الزراعة القدرة على إحداث تأثيرات ضارة على الكائنات الحية غير المستهدفة وخاصة الإنسان. خلال هذه الدراسة، تم بناء نماذج علاقة كمية بين البنية والسمية (QSTR) للتنبؤ بالسمية الحادة لـ 30 مركباً تنتمي لمجموعتي مثبطات التنفس و التخليق الحيوي للإرغوستيرول. تم استخدام الخوارزمية الجينية (GA) مع الانحدار الخطي المتعدد (MLR)، وخوارزمية آلة المتجهات الداعمة (SVM) لبناء هذه النماذج. أظهر النموذج المنشأ باستخدام خوارزمية آلة المتجهات الداعمة (SVM) قدرة تنبؤية جيدة ($R^2 = 0.75$)، متفوقاً على نموذج الانحدار الخطي المتعدد (MLR) الذي كانت قيمة معامل التحديد لديه R^2 مساوية 0.69. تم إجراء التحقق المتبادل (Cross-validation)، واختبار العشوائية (Y-randomization test)، والتحقق الخارجي (External validation) لتقييم دقة وموثوقية النماذج الناتجة. أظهر نموذج SVM أداءً فائقاً مقارنةً بنموذج MLR، محققاً قيم R^2_{cv} و R^2_{test} بلغت 0.63 و 0.72 على التوالي، مقارنةً بـ 0.53 و 0.64 لنموذج MLR. تعتبر هذه النماذج أداة هامة لتقييم السمية الحادة لهذا النوع من المركبات، مع وجود حاجة

نمذجة العلاقة الكمية بين التركيب والسمية المدفوعة بالتعلم الآلي لتقييم السمية الحادة لبعض مبيدات الفطريات المثبطة للتنفس وتخليق الإرغوستيرول

فعلية لتطوير هذه النماذج مستقبلاً باستخدام خوارزميات جديدة ومجموعة بيانات أكثر شموليةً و
اتساعاً.

الكلمات المفتاحية: تعلم الآلة، الإنحدار الخطي المتعدد، خوارزمية آلة المتجهات الداعمة،
السمية الحادة.

Quantitative Structure-Toxicity Relationship (QSTR) Modeling Using Machine Learning for Acute Toxicity of Respiration and Ergosterol Biosynthesis Inhibitor Fungicides

Mohsen Abbod *

Department of Plant Protection, Faculty of Agriculture, Homs University,
Homs, Syria.

Abstract

The global reliance on respiratory and ergosterol biosynthesis inhibitor fungicides in agriculture necessitates a thorough understanding of their potential acute toxicity to non-target organisms, including humans. This research focused on building Quantitative Structure-Toxicity Relationship (QSTR) models to predict the acute toxicity of 30 compounds within these critical fungicide groups. We utilized a combined approach of Genetic Algorithm (GA) with both Multiple Linear Regression (MLR) and Support Vector Machine (SVM) techniques for model development. Our findings indicated that the SVM-based model offered superior predictive performance, achieving an $R^2 = 0.75$, compared to the MLR models' $R^2 =$

0.69. To ensure robustness, the models underwent extensive validation, including cross-validation, Y-randomization, and external validation. The SVM model consistently demonstrated stronger predictive accuracy and reliability, evidenced by its R^2_{CV} and R^2_{test} values of 0.63 and 0.72, respectively, which surpassed the MLR model's corresponding values of 0.53 and 0.64. These QSTR models represent a significant stride in evaluating the acute toxicity of these compounds, underscoring the importance of their further refinement with novel algorithms and more comprehensive datasets.

Keywords: Machine Learning, Multiple Linear Regression, Support Vector Machine, Acute Toxicity.

1- مقدمة ودراسة مرجعية

يمثل الإستخدام المكثف للمبيدات الكيميائية التقليدية تحدياً بيئياً وصحياً بالغ الخطورة، حيث تساهم مبيدات الآفات الاصطناعية ومخلفاتها في الإضرار بالتوازن البيئي وتشكيل تهديد مباشر لصحة الإنسان [1]، وقد أدى الاستخدام المفرط والمستمر للمبيدات الكيميائية الفطرية إلى تفاقم هذه المشكلة، مما ساعد على ظهور سلالات ممرضة جديدة متزايدة المقاومة وصعبة المكافحة. وانطلاقاً من هذه التحديات، اتجهت الجهود البحثية نحو تطوير استراتيجيات بديلة وأكثر أماناً لمكافحة الآفات، مع الحفاظ على الفعالية في حماية المحاصيل [2].

يحظى تطوير مبيدات فطرية جديدة بأولوية بحثية عالية في برامج الشركات الزراعية، لا سيما مع تزايد القيود المفروضة على استخدام المبيدات عالية الخطورة نظراً لتأثيراتها السمية الحادة والمزمنة والمخاطر المصاحبة لها على الإنسان والبيئة [1، 2]. لذا فإن البحث عن بدائل فعالة وآمنة أصبح ضرورة ملحة. أظهرت دراسة وجود مستويات عالية من السمية الخلوية في

الخلايا العصبية القشرية لأجنة الفئران المعرضة لمبيد الأزوكسي ستروبين (Azoxystrobin) المثبط لعملية التنفس في الخلايا الفطرية، ما يعكس احتمالية تعرض النساء الحوامل والأطفال لمستويات سمية بسبب التعرض المتكرر للمبيد حسب الدراسة [3]. كما أظهرت دراسة أخرى إلى حدوث تلف في الأمعاء الوسطى لنحل العسل (*A. mellifera*) المعرض لمبيد الأزوكسي ستروبين مما يستدعي دراسة تأثيره على الأعضاء الأخرى للنحل ومدى تأثيراته السمية على الكائنات غير المستهدفة [4]. على الرغم من الانتشار الواسع لاستخدامها، كشفت العديد من الدراسات عن الأخطار الصحية للتعرض لمبيدات الفطريات المثبطة لتخليق الإرغوستيرول الفطري، حيث يتسبب التعرض لمركبات التريازول Triazoles تأثيرات سامة على الجهاز التناسلي، و حدوث تلفات في الكبد [5، 6]. أظهر مبيد بروبيكونازول (Propiconazole) سمية على أجنة *Daphnia magna* [7] بالإضافة إلى تسببه في ظهور أعراض السمية الـ كـ دـ يـ ة [8].

برزت حاجة ملحة لتطوير منهجيات علمية متقدمة تمكن من التنبؤ بسمية هذه المركبات قبل تطبيقها الميداني وخاصة فيما يتعلق بتقييم سميتها على الكائنات غير المستهدفة، وتأتي منهجية العلاقة الكمية بين البنية والسمية (Quantitative Structure–Toxicity Relationship, QSTR) في طليعة هذه المنهجيات، حيث تقدم أدوات تحليلية دقيقة للربط بين الخصائص الجزيئية والفيزيوكيميائية للمركبات المدروسة وسميتها [9]. وتساعد في بناء نماذج رياضية متقدمة تمكن من التنبؤ بسمية هذه المركبات عبر دراسة الخواص الهيكلية و الكيميائية للمركبات و التي تسمى بالمؤشرات الجزيئية أو Molecular descriptors [10].

في هذه الدراسة، تم استخدام طريقة الإنحدار الخطي المتعدد مدعومة بخوارزميات تعلم الآلة (Machine learning) لبناء نماذج QSTR موثوقة لتقييم السمية الحادة لبعض مبيدات الفطريات المثبطة للتنفس وتخليق الإرغوستيرول. و قد تم التحقق من قوة هذه النماذج عبر

المعايير المتبعة في هذا النوع من الدراسات لتشكل لجنة للتحقق من سمية المركبات المنتمة لهذه المجموعات مستقبلاً.

2- مبررات البحث و أهدافه

نظراً لأن الإستخدام المفرط لمبيدات الفطريات، خاصةً مثبطات التنفس وتخليق السيتيرول (الإرغوستيرول)، يرتبط بسمية حادة للكائنات غير المستهدفة ويثير مخاوف بيئية، فإن هناك حاجة ملحة لتطوير منهجيات علمية متقدمة للتنبؤ بسمية هذه المركبات قبل تطبيقها الحقلية. يتطلب ذلك إرساء أطر علمية قوية ومتكاملة، تشمل نماذج تنبؤية تعتمد على البيانات الكيميائية والبيولوجية، وتقنيات السموميات الحاسوبية لفهم آليات السمية.

تهدف هذه الدراسة بشكل أساسي إلى تطوير نماذج رياضية، مدعومة بخوارزميات تعلم الآلة، للتنبؤ بالسمية الحادة لعدد من مبيدات الفطريات. ستركز الدراسة على المبيدات المنتمة لمجموعتي مثبطات التنفس ومثبطات التخليق الحيوي للإرغوستيرول، وذلك باستخدام منهجية العلاقة الكمية بين البنية و السمية (QSTR).

3- مواد وطرق العمل

3-1-البيانات المستخدمة في الدراسة و تهيئة هياكل المركبات ثلاثية الأبعاد

تتألف مجموعة البيانات المستخدمة في هذه الدراسة من 30 جزيئاً من مبيدات الفطريات، والتي تعمل كمثبطات للتنفس وتخليق الإرغوستيرول [11]. تم جمع قيم الجرعة المميطة النصفية عن طريق الفم (LD_{50}) لهذه المركبات، والتي تُقاس بـ (ملغ/كغ) وتُحدد ضد الفئران. تُعتبر هذه القيم معياراً عالمياً معتمداً لقياس السمية الحادة للمركبات الكيميائية. وقد تم الحصول على هذه

البيانات من مصادر موثوقة تشمل: منظمة الصحة العالمية [12]، وقاعدة بيانات PubChem [13] ، والمكتبة الوطنية للأدوية في الولايات المتحدة الأمريكية [14].

تم بناء كل مركب ورسم هيكله، ثم تحسينه باستخدام حقول القوة (+MM) المتاحة في برنامج HyperChem. بعد هذه الخطوة الأولية، تم إجراء تحسين للهندسة الجزيئية ثلاثية الأبعاد للمركب باستخدام طريقة AM1 الكيمائية-الكمومية، مع استهداف جذر متوسط مربع التدرج بقيمة 0.01 كيلو كالوري لكل مول [15].

3-2- حساب المؤشرات الجزيئية (Molecular descriptors) للمركبات المدروسة

لأغراض النمذجة، استُخدمت مجموعة واسعة من المؤشرات الجزيئية النظرية، شملت المؤشرات البنائية (0D) ، وأحادية الأبعاد (1D) ، وثنائية الأبعاد (2D) ، وثلاثية الأبعاد (3D) ، بهدف التقاط وتوضيح جوانب مختلفة من التركيب الكيميائي والخواص الفيزيوكيميائية لكل مركب من المركبات المدروسة. جرى حساب المؤشرات المختارة في كل نموذج من إجمالي 1497 مؤشراً باستخدام برنامج E-Dragon 3.0 [16] (يوفر هذا البرنامج إمكانية حساب 1497 مؤشراً لكل مركب كيميائي)، تم استبعاد الوصفات التي يقل انحرافها المعياري عن 0.001، وكذلك تلك التي تحتوي على قيمة مفقودة واحدة على الأقل. كما أُزيلت الوصفات غير المتغيرة والمترابطة (معامل الارتباط $R > 0.9$) وفي النهاية، استُخدمت المؤشرات الباقية كمدخلات لنمذجة QSTR. تم توفير قائمة بهذه المؤشرات ومعانيها، وإجراءات حسابها، بالإضافة إلى المراجع ذات الصلة، من خلال برنامج E-Dragon 3.0 [16].

3-3- بناء النماذج التنبؤية

3-3-1- الخوارزمية الجينية (Genetic algorithm-GA)

تُعد الخوارزمية الجينية (GA) تقنية بحث مستوحاة من نظرية داروين للتطور والانتخاب الطبيعي. نظرًا لكفاءتها وبساطتها، فقد استُخدمت الخوارزمية الجينية على نطاق واسع كاستراتيجية واحدة لاختيار المتغيرات الأكثر تأثيراً على المخرجات المدروسة، كقيمة السمية في الدراسات المشابهة لدراستنا [17]. في هذا البحث، استُخدمت الخوارزمية الجينية (GA) لتحديد المؤشرات الجزيئية الرئيسية التي تؤثر في السمية الحادة للمركبات المدروسة. بحيث تم اختيار أربعة مؤشرات من بين المؤشرات المحسوبة سابقاً لبناء المعادلة الرياضية النهائية. تم اختيار أربعة مؤشرات فقط لتجنب مشكلة Overfitting في النموذج الناتج بما يتوافق مع عدد المركبات المدروسة، وقد تم تنفيذ الخوارزمية الجينية باستخدام برنامج MATLAB إصدار 2019 [18].

3-3-2- تقسيم البيانات وبناء نموذج الانحدار الخطي المتعدد (Multiple linear regression-MLR)

بعد تحديد أهم الواصفات الجزيئية باستخدام الخوارزمية الجينية (GA)، تم تقسيم مجموعة البيانات إلى مجموعة تدريب (Train set) ومجموعة اختبار (Test set)، حيث تم بناء النماذج باستخدام 25 مركب من المركبات، في حين استخدمت المركبات الخمسة الباقية، والتي اختيرت بطريقة عشوائية لاختبار قدرة النماذج التنبؤية على المركبات الخارجية التي لم تدخل في بناء النموذج الأصلي.

تستخدم تقنية الانحدار الخطي المتعدد لنمذجة العلاقات الخطية بين المتغيرات التابعة والمستقلة [19]. في دراستنا، تمثل قيم الفعالية السمية الحادة (LD_{50}) المتغير التابع و تشير قيم

المؤشرات الجزيئية إلى المتغيرات المستقلة. يمكن تمثيل معادلة الانحدار كما يلي (المعادلة رقم 1):

$$Y = a_0 + \sum_{i=1}^n a_i X_i \quad (1)$$

حيث Y هو المتغير التابع (LD_{50})، و X_i هي المتغيرات المستقلة (المؤشرات الجزيئية)، n هو عدد المركبات، a_0 هو ثابت المعادلة، و a_i تمثل coefficient المتغيرات المستقلة. تم تحويل قيم التركيزات الفعالة (LD_{50}) للمركبات إلى القيم اللوغاريتمية أو $pLD_{50} = -\log LD_{50}$ ، وتم استخدامها كمتغيرات تابعة، بينما تم استخدام المؤشرات الجزيئية المحدد بواسطة GA كمتغيرات مستقلة في حساب معادلة الانحدار.

3-3-3- التنبؤ باستخدام خوارزميات تعليم الآلة (Machine learning-QSTR models)

تشكل خوارزميات التعلم الآلي - وعلى رأسها آلة المتجهات الداعمة (Support vector machine-SVM) بديلاً فعالاً لدراسة العلاقات غير الخطية المعقدة في نمذجة QSTR، خاصة عندما تعجز أساليب الانحدار الخطي الكلاسيكية عن تمثيل العلاقة بين النشاط البيولوجي والخصائص الجزيئية [20، 21]. تمتاز خوارزمية SVM بقدرتها على التعامل مع البيانات غير الخطية وقوة تعميمها، مما جعلها أداة قيمة في هذا المجال. في هذه الدراسة، جرى تطوير نموذج ML-QSTR باستخدام SVM بالاعتماد على نفس المؤشرات الجزيئية الأربعة المستخدمة في نموذج الانحدار الخطي المتعدد (MLR)، مع تحسين معايير النموذج (Kernel function و Box constraint و Epsilon) باستخدام التحسين البايزي (Bayesian Optimization). وتقييم أداء النموذج الناتج عبر معامل التحديد (R^2) ومتوسط مربع الخطأ (MSE) ضمن بيئة برنامج MATLAB [18].

3-4-4- التحقق من صحة النماذج (Validation)

3-4-1- التحقق الداخلي (Internal validation) بطريقة Cross Validation

تهدف عملية التحقق إلى تقييم مدى موثوقية النماذج الإحصائية. لتحقيق ذلك، استُخدمت طريقة التحقق المتقاطع بترك مجموعة عناصر (LMO-CV)، وهي أسلوب إحصائي شائع للتحقق الداخلي. تتضمن هذه الطريقة استبعاد جزء من مجموعة البيانات بشكل متكرر لاستخدامه كمجموعة تحقق منفصلة، بهدف تقييم القدرة التنبؤية لنماذج QSTR الناتجة.

من بين تقنيات التحقق المتقاطع (Cross-validation) الشائعة في دراسات QSTR، يُعد التحقق بترك عنصر واحد (Leave-One-Out Cross-Validation, LOO-CV) والتحقق بترك مجموعة عناصر (Leave-Many-Out Cross-Validation, LMO-CV) الأكثر استخداماً. في طريقة LOO-CV، يُخصص مركب واحد ليكون مجموعة التحقق في كل خطوة تكرارية، بينما تُستخدم المركبات المتبقية كمجموعة تدريب. تتكرر هذه العملية لكل مركب، لينتج عنها n تكراراً، حيث يمثل n العدد الإجمالي للمركبات، وهي الطريقة المستخدمة في الدراسة. يعتبر النموذج قوياً داخلياً في تنبؤاته إذا حقق معامل ارتباط (R^2_{cv}) أكبر من 0.5 [22].

3-4-2- اختبار Y-Randomization

يُعد اختبار Y-randomization أسلوباً مهماً يُستخدم لتقييم مدى تأثر النماذج الإحصائية بالصدفة في ملاءمة البيانات. الهدف الرئيسي من هذا الاختبار هو التحقق من وجود ترابطات عشوائية بين المؤشرات الجزيئية والفعالية البيولوجية (السمية) ضمن النموذج الأصلي الذي تم بناؤه باستخدام طريقة الانحدار الخطي المتعدد MLR. تتم عملية اختبار Y-randomization عن طريق خلط المتغير التابع عشوائياً، بينما تظل المتغيرات المستقلة (المؤشرات الجزيئية) في

مواقعها الأصلية دون تغيير. بعد ذلك، تُبنى نماذج QSTR جديدة باستخدام هذه البيانات المختلطة.

يُعتبر نموذج QSAR صالحًا ويعكس علاقة حقيقية، وليس مجرد نتيجة للصدفة للبحث، عندما تكون قيم متوسط معامل التحديد العشوائي (R^2) ومعامل التحديد المتقاطع (R^2_{cv}) للنماذج المولدة عشوائياً أقل من قيمة معامل التحديد (R^2) للنموذج الأصلي [23].

3-4-3- التحقق الخارجي من قوة النماذج (External validation)

تم تقييم أداء النماذج المطورة أيضاً على مجموعة اختبار خارجية (External test set) مكونة من خمسة مركبات تم اختيارها عشوائياً كما ذكر سابقاً. تتكون مجموعة الاختبار هذه من 5 مركبات، تمثل ما يزيد عن 15% من عدد المركبات الكلي بحيث لم يتم استخدامها في تطوير وتدريب النماذج. تم تقييم قدرة النموذج على التنبؤ بالسمية لمجموعة الاختبار الخارجية باستخدام معيارين: R^2_{test} و MSE_{test} . إذا كانت قيمة R^2_{test} أكبر من 0.5، يُعتبر النموذج مقبولاً من حيث أدائه التنبؤي [22].

4- النتائج و المناقشة

4-1- تهيئة البيانات الخاصة بالمركبات و اللازمة لعملية النمذجة

تم جمع بيانات السمية الحادة لثلاثين مركب من المبيدات الفطرية و المنتمية لمجموعتي مثبطات التنفس (16 مركب)، وأربع عشر مركب من مثبطات التخليق الحيوي للستيرول الفطري وذلك من قواعد البيانات الموضحة في مواد و طرق العمل. اختيرت قيمة الجرعة القاتلة النصفية

للجرذ الأبيض (LD₅₀) كنقطة السمية المرجعية للمركبات، وتم تحويل قيم السمية إلى القيم اللوغاريتمية السالبة (pLD₅₀) لزيادة كفاءة النموذج الرياضي الناتج.

تم حساب 1497 من المؤشرات الجزيئية النظرية لكل مركب باستخدام برنامج E-Dragon 3.0 و تم تقليل عدد المؤشرات إلى 1232 حسب المعايير التي تم إيضاحها في مواد وطرق العمل. بعد ذلك استخدمت الخوارزمية الجينية (GA) لتحديد المؤشرات الأربعة الأكثر تأثيراً على سمية المركبات و التي نتج عنها اختيار المؤشرات الأربعة الموضحة في الجدول 1، والذي يوضح اسم المؤشر و المعنى الكيميائي الخاص [16، 24]. و يوضح الجدول 2 قيمة هذه المؤشرات المحسوبة للمركبات الثلاثين موضع الدراسة.

جدول 1. المؤشرات الجزيئية المستخدمة في بناء النماذج ومعانيها الكيميائية

المؤشر الجزيئي	المعنى الكيميائي للمؤشر [16، 24]	المجموعة التي ينتمي لها المؤشر
BEHm7	أعلى قيمة ذاتية رقم 7 لمصفوفة Burden / مرجحة بالكتل الذرية	Burden Eigenvalue
Mor25m	3DMoRSE - إشارة 25 / مرجحة بالكتلة الذرية	3DMoRSE
G3e	مؤشر WHIM الاتجاهي التناظري للمكون الثالث / مرجح بالكهربية الذرية لسانديرسون	WHIM
G3s	مؤشر WHIM الاتجاهي لتمائل المكون الثالث / المرجح بالحالات الكهروطوبولوجية الذرية	WHIM

نمذجة العلاقة الكمية بين التركيب والسمية المدفوعة بالتعلم الآلي لتقييم السمية الحادة لبعض مبيدات الفطريات المثبطة للتنفس وتخليق الإرغوستيرول

جدول 2. قيم المؤشرات الجزيئية و قيم التركيز القاتل النصفى للمركبات المدروسة

قيمة المؤشرات الجزيئية				pLD ₅₀	LD ₅₀ ملغ/كغ	المجموعة الكيميائية [11]	اسم المبيد
G3s	G3e	Mor25m	BEHm7				
مجموعة التدریب							
0.191	0.169	0.689	2.331	- 2.4886	308	G1	Azaconazole
0.207	0.193	0.823	2.658	- 3.6998	5010	C2	Boscalid
0.2	0.171	0.349	2.539	- 3.4130	2588	C2	Carboxin
0.177	0.16	1.058	2.983	- 3.6994	5005	C2	Cyclobutrifluram
0.183	0.16	0.703	2.691	- 2.6513	448	C1	Diflumetorim
0.185	0.16	0.481	2.458	- 2.6758	474	G1	Diniconazole
0.181	0.177	0.759	2.624	- 3.4997	3160	G1	Epoxiconazole
0.185	0.162	0.767	2.68	- 3.1281	1343	G1	Etaconazole
0.183	0.159	0.466	2.624	- 3.3071	2028	C3	Fenamidone
0.204	0.198	0.36	1.983	- 4.1106	12900	C2	Fenfuram
0.183	0.161	0.594	2.631	- 2.8287	674	G1	Flusilazole
0.188	0.161	0.397	2.554	- 3.3402	2189	G1	Hexaconazole
0.191	0.165	0.642	2.451	- 2.3560	227	G1	Imazalil
0.192	0.154	0.56	2.604	- 2.2553	180	C2	Inpyrfluxam
0.181	0.153	0.419	2.712	- 2.9484	888	G1	Iponazole

0.174	0.152	0.401	2.995	- 3.3010	2000	C2	Isoflucypram
0.183	0.164	0.601	2.612	- 3.0969	1250	G1	Nuarimol
0.193	0.168	0.586	2.576	- 3.2127	1632	C2	Oxycarboxin
0.181	0.153	0.243	2.639	- 3.3010	2000	C2	Penflufen
0.181	0.159	0.828	2.952	- 3.0099	1023	G1	Prochloraz
0.175	0.156	0.538	3.027	- 3.7003	5015	C2	Pydiflumetofen
0.197	0.174	0.006	2.348	- 3.6990	5000	C7	Silthiofam
0.181	0.164	0.758	2.67	- 3.0133	1031	G1	Tetraconazole
0.175	0.167	1.015	3.134	- 3.8129	6500	C2	Thifluzamide
0.174	0.151	0.525	2.976	- 2.5866	386	C1	Tolfenpyrad
مجموعة الاختبار							
0.174	0.166	1.214	2.983	- 3.1623	1453	G1	Difenoconazole
0.171	0.165	0.971	3.069	- 3.6128	4100	C5	Fluazinam
0.171	0.151	0.832	2.985	- 3.0000	1000	C3	Pyraoxystrobin
0.185	0.156	0.315	2.598	- 3.2304	1700	G1	Tebuconazole
0.188	0.167	0.733	2.408	- 2.8579	721	G1	Triadimenol

حيث **C** هي مجموعة المبيدات الفطرية المثبطة للتنفس و **G** تشير للمبيدات الفطرية من مثبطات التخليق الحيوي للإرغوستيرول ، بينما يشير الرقم المرافق للرمز إلى المجموعة الفرعية ضمن المجموعة الرئيسية حسب المنظمة الدولية لمقاومة المبيدات الفطرية [11].

4-2- نموذج الانحدار الخطي المتعدد MLR:

نمذجة العلاقة الكمية بين التركيب والسمية المدفوعة بالتعلم الآلي لتقييم السمية الحادة لبعض مبيدات الفطريات المثبطة للتنفس وتخليق الإرغوستيرول

بعد تهيئة البيانات بالشكل المناسب لعملية بناء النموذج الرياضي تم تقسيم العدد الكلي للمركبات لمجموعتين، الأولى مكونة من 25 مركب استخدمت لبناء النماذج الرياضية و تسمى مجموعة التدريب (Train set)، والثانية مكونة من خمسة مركبا اختيرت بشكل عشوائي لتشكّل مجموعة الاختبار (Test set).

استخدمت المركبات الخمس وعشرين (مجموعة التدريب) لإنشاء نموذج انحدار خطي متعدد MLR، يربط بين المؤشرات الأربعة المختارة لمركبات مع قيمة الجرعة القاتلة النصفية المحددة للسمية الحادة لهذه المركبات ممثلة بالقيم اللوغاريتمية (pLD_{50})، كما في المعادلة التالية:

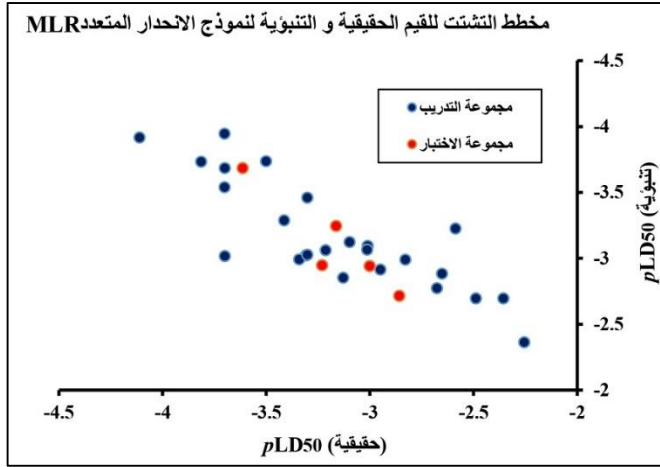
$$pLD_{50} = 4.43 - 1.414(BEHm7) + 1.221(Mor25m) - 56.74(G3e) + 25.8(G3s) \quad (2)$$

حيث:

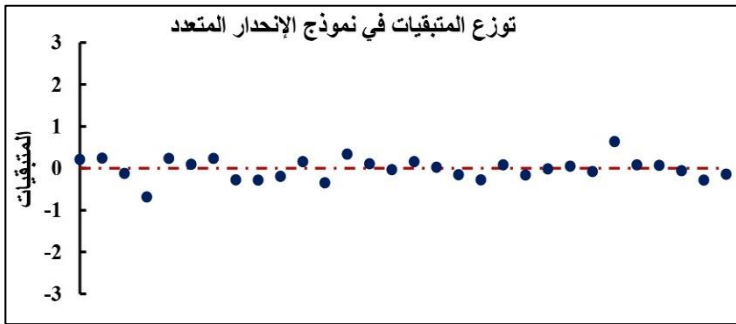
$$n = 25, R^2_{(train)} = 0.69, R^2_{(adj.)} = 0.63, MSE_{(train)} = 0.07, R^2_{(test)} = 0.64, R^2_{cv} = 0.53, MSE_{(test)} = 0.023, P < 0.001$$

تبين المعادلة الناتجة (المعادلة 2) علاقة خطية جيدة بين الواصفات الأربعة المختارة بواسطة الخوارزمية الجينية وسمية المركبات (pLD_{50}). تُظهر المعلمات الإحصائية $R^2_{(train)} = 0.69$ ، وقيمة p أقل من 0.001 أن نموذج الارتباط الخطي المتعدد (MLR) جيد إحصائياً، و ذو قدرة تنبؤية مقبولة لسمية المركبات المدروسة. يشير معامل ارتباط (Cross validation) $R^2_{cv} = 0.53$ (< 0.5) إلى أن النموذج دقيق في توقع السمية الحادة للبيانات المستبعدة من مجموعة التدريب خلال عملية Cross validation، ما يؤكد موثوقية النموذج الناتج في التنبؤ بسمية هذا النوع من المركبات.

يُظهر الشكل 1 توزيع القيم التجريبية و المتوقعة لـ pEC_{50} عبر نموذج MLR، ويمكن الكشف عن ارتباط جيد بين القيم التجريبية و القيم المتوقعة. كما نلاحظ التوزيع الطبيعي للمتبقيات (Residuals) على طرفي الحد (0) وعدم وجود أي نمط معين لهذا التوزيع ما يجعل النتائج المتحصل عليها مقبولة إحصائياً (شكل 2).



شكل 1. مخطط التشتت للقسم الحقيقية و التنبؤية لنموذج الانحدار المتعدد



شكل 2. التوزيع الطبيعي للمتبقيات (Residuals) لنموذج الانحدار المتعدد

3-4- اختبار العشوائية أو Y-Randomization الخاص بنموذج الانحدار المتعدد MLR

لضمان متانة نموذج الانحدار المتعدد، تم إجراء اختبار يسمى "Y-randomization" أو "اختبار العشوائية". تم فيه إعادة توزيع قيم المتغير التابع pLD_{50} عشوائياً على المتغيرات المستقلة الأربعة الثابتة التي لم تتغير. كررت هذه العملية 100 مرة، لينتج عنها 100 نموذج عشوائي جديد.

أظهرت النتائج أن متوسط قيم معاملات التحديد (R^2) والتحديد المتقاطع (R^2_{cv}) لهذه النماذج العشوائية المئة كان أقل بشكل واضح من القيم الأصلية التي سجلها النموذج الحقيقي. كما أن قيمة معامل (cRp^2) المحسوب للاختبار كانت أكبر من 0.5، وهو ما يؤكد قوة وصدق النموذج الأصلي ويدل على أن نتائجه ليست وليدة الصدفة. وتوضح هذه النتائج بشكل جماعي في الجدول (3).

جدول 3. نتائج اختبار Y-Randomization الخاص بنموذج الانحدار المتعدد

متوسط القيمة لمئة تكرار	معيار اختبار Y-Randomization
0.32	R
0.117	R^2
-0.28	R^2_{cv}
0.68	cRp^2

4-4- نماذج التنبؤ الناتجة باستخدام خوارزميات تعليم الآلة (Machine learning- QSTR models)

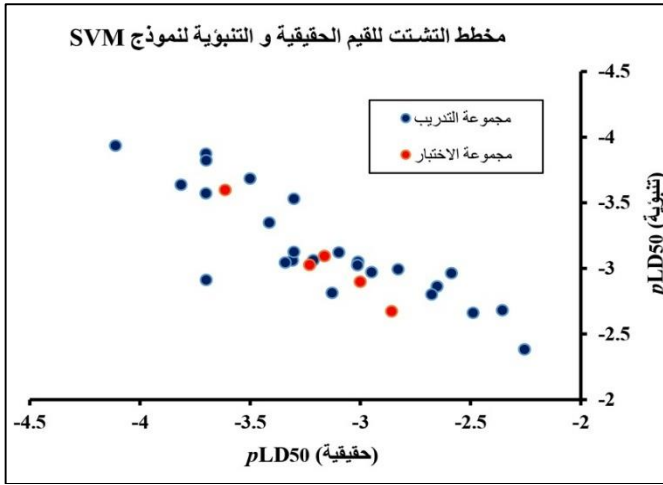
لتحسين العلاقة الرياضية بين سمية المركبات المدروسة والمؤشرات الجزيئية المختارة التي تم الحصول عليها من خلال الخوارزمية الجينية، تم استخدام هذه المؤشرات الأربعة كمدخلات (Inputs) لخوارزمية آلة المتجهات الداعمة (Support vector machine-SVM).

تم اختيار دالة خطية (Linear kernel function) ومستوى box constraint يساوي 8.449 ومقياس Epsilon يساوي 0.1731 كمعاملات مثلى للخوارزمية عبر التحسين البايزي (Bayesian Optimization).

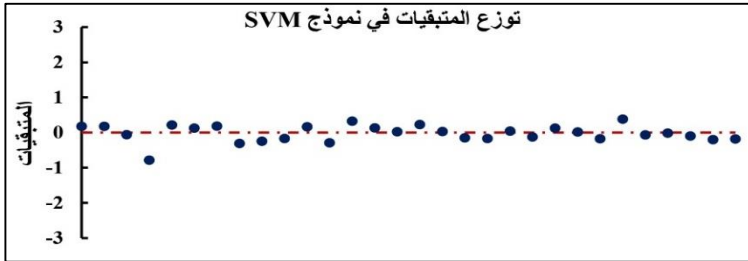
كانت المعلمات الإحصائية الخاصة بالنموذج الناتج كالتالي: متوسط مربع الخطأ $MSE_{(train)}$ هو 0.058 ، مع معامل التحديد $R^2_{(train)}$ و R^2_{cv} معادل 0.75 و 0.63 على التوالي (الجدول 4) ، مما يشير إلى تفوق لنموذج QSTR القائم على SVM في التنبؤ بسمية المركبات المدروسة مقارنة بنموذج MLR.

نلاحظ من الشكلين (3 و 4) أن القيم المتوقعة تتوافق بشكل ممتاز مع القيم التجريبية سواء في مجموعة التدريب أو مجموعة الاختبار. كما أن البقايا الخاصة بكلا المجموعتين لا تظهر أي نمط معين و متقاربة بشكل كبير من خط القيمة 0، ما يشير إلى قوة النموذج الناتج من خوارزمية SVM.

نمذجة العلاقة الكمية بين التركيب والسمية المدفوعة بالتعلم الآلي لتقييم السمية الحادة لبعض مبيدات الفطريات المثبطة للتنفس وتخليق الإرغوستيرول



شكل 3. مخطط التشتت للقسم الحقيقية و التنبؤية لنموذج SVM



شكل 4. التوزع الطبيعي للمتبقيات (Residuals) لنموذج SVM

4-5- التحقق من النماذج باستخدام مجموعة الاختبار External validation

يغية تقييم قدرة النماذج الرياضية على التنبؤ بسمية مركبات جديدة لم تُستخدم في بنائها، تم إخضاع نموذجي الانحدار الخطي المتعدد (MLR) و نموذج آلة المتجهات الداعمة (SVM) لاختبار خارجي باستخدام مجموعة مستقلة من خمسة مركبات. وأظهرت نتائج التقييم المعتمدة على معياري R^2_{test} و MSE_{test} تفوقاً واضحاً لنموذج SVM من حيث دقة التنبؤ والقدرة على

التعميم. حيث سجّل النموذج قيماً بلغت 0.72 و 0.018 للمعيارين على التوالي، مقارنةً بقيمتي 0.64 و 0.023 لنموذج الانحدار المتعدد الخطي (الجدول 4). وعموماً، تُظهر المقارنة بناءً على معايير التحقق الداخلي والخارجي أن نموذج SVM قد تفوّق على نموذج MLR، حيث حقق قيمة أقل لـ MSE وقيماً أعلى لمعامل التحديد R^2 ، مما يعكس قدرة تنبؤية واستقرائية أفضل للنموذج.

جدول 4. مقارنة المعايير الإحصائية للنماذج الناتجة عن الدراسة

النموذج	المعيار الإحصائي				
	R^2_{train}	MSE_{train}	R^2_{cv}	R^2_{test}	MSE_{test}
MLR	0.69	0.07	0.53	0.64	0.023
SVM	0.75	0.058	0.63	0.72	0.018

ويوضح الجدول التالي (جدول 5) القيم الحقيقية و القيم المتنبأ بها للمركبات المدروسة باستخدام النماذج الرياضية الناتجة:

جدول 5. القيم الحقيقية و المتنبأ بها لسمية للمركبات المدروسة (pLD_{50}) عبر النماذج

الناتجة

pLD_{50} التنبؤية		pLD_{50} الحقيقية	LD_{50} (ملغ/كغ)	اسم المبيد
MLR	SVM			
-	-	-	308	Azaconazole
2.69779	-2.6628	2.4886		
-	-	-	5010	Boscalid
3.94626	-3.8743	3.6998		
-	-	-	2588	Carboxin
3.28899	-3.3502	3.4130		

نمذجة العلاقة الكمية بين التركيب والسمية المدفوعة بالتعلم الآلي لتقييم السمية الحادة لبعض مبيدات الفطريات المثبطة للتنفس وتخليق الإرغوستيرول

-	-	-	5005	Cyclobutrifluram
3.01915	-2.9136	3.6994		
-	-	-	448	Diflumetorim
2.88526	-2.8633	2.6513		
-	-	-	474	Diniconazole
2.77536	-2.8033	2.6758		
-	-	-	3160	Epoxiconazole
-3.7381	-3.6849	3.4997		
-	-	-	1343	Etaconazole
-2.8535	-2.8138	3.1281		
-	-	-	2028	Fenamidone
3.02324	-3.0608	3.3071		
-	-	-	12900	Fenfuram
-3.918	-3.9370	4.1106		
-	-	-	674	Flusilazole
2.99027	-2.9935	2.8287		
-	-	-	2189	Hexaconazole
2.99323	-3.0463	3.3402		
-	-	-	227	Imazalil
2.69798	-2.6825	2.3560		
-	-	-	180	Inpyrfluxam
-2.3647	-2.3848	2.2553		
-	-	-	888	Ipconazole
2.91621	-2.9730	2.9484		
-	-	-	2000	Isoflucypram
3.46199	-3.5312	3.3010		
-	-	-	1250	Nuarimol
3.12505	-3.1234	3.0969		
-	-	-	1632	Oxycarboxin
3.06189	-3.0633	3.2127		

-	-	-	2000	Penflufen
3.02793	-3.1275	3.3010		
-	-	-	1023	Prochloraz
-3.0965	-3.0514	3.0099		
-	-	-	5015	Pydiflumetofen
-3.5411	-3.5733	3.7003		
-	-	-	5000	Silthiofam
3.68526	-3.8246	3.6990		
-	-	-	1031	Tetraconazole
3.06682	-3.0260	3.0133		
-	-	-	6500	Thifluzamide
-3.7339	-3.6381	3.8129		
-	-	-	386	Tolfenpyrad
3.22692	-2.8644	2.5866		
-	-	-	1453	Difenoconazole
3.24627	-3.0947	3.1623		
-	-	-	4100	Fluazinam
3.68524	-3.5987	3.6128		
-	-	-	1000	Pyraoxystrobin
2.94192	-2.8997	3.0000		
-	-	-	1700	Tebuconazole
-2.9492	-3.0278	3.2304		
-	-	-	721	Triadimenol
2.71673	-2.6736	2.8579		

تُستخدم المبيدات الفطرية، ولا سيما تلك التي تنتمي إلى فئة مثبطات التنفس ومثبطات تخليق الإرجوستيرول، على نطاق واسع في المجالين الطبي والزراعي لمكافحة العوامل الممرضة الفطرية [25، 26]. وقد أدى الاستخدام المكثف لهذه المركبات إلى ظهور سلالات مقاومة على نحو متسارع، مما حدا بالباحثين والشركات إلى استكشاف مركبات وصيغ جديدة منخفضة

المخاطر ضمن هذه الفئة من العوامل المضادة للفطريات [27]. وتُعد السمية البيئية والدوائية المحتملة عاملاً حاسماً في تطوير مبيدات فطرية جديدة تتميز بالفعالية وانخفاض المخاطر على الإنسان والبيئة. وفي إطار هذه الدراسة، تم تقديم نماذج QSTR أظهرت كفاءة عالية في التنبؤ بالآثار الضارة المرتبطة بهذا النوع من المركبات الكيميائية، مما يمهد الطريق أمام الباحثين لتطوير مركبات أكثر أماناً في المستقبل.

أظهر نموذج الانحدار الخطي المتعدد (MLR) قدرة تنبؤية جيدة لسمية المركبات قيد الدراسة، حيث بلغ معامل التحديد (R^2) 0.69 لمجموعة التدريب و 0.64 لمجموعة الاختبار، مع تأكيد موثوقيته التنبؤية إحصائياً من خلال تجاوز اختبار العشوائية (Y-Randomization) ومع ذلك، تفوق النموذج المنشأ باستخدام خوارزمية آلة المتجهات الداعمة (SVM) بشكل ملحوظ خلال مرحلة التدريب والاختبار الخارجي، حيث سجل معامل تحديد قدره 0.72 لمجموعة الاختبار.

أظهرت عدة دراسات فاعلية تقنيات التعلم الآلي مثل SVM في أبحاث QSTR [28]، وتعد خوارزمية آلة المتجهات الداعمة (SVR) أداة واسعة الانتشار في تصميم الأدوية والتنبؤ بالسمية [29]. وفي هذه الدراسة، أظهر نموذج SVM أداءً متميزاً في التنبؤ بسمية المركبات المدروسة، متجاوزاً أداء نموذج الانحدار المتعدد. وعلى الرغم من عدم وجود إجماع حول خوارزمية مثالية موحدة لتطوير النماذج في مجال QSAR [21]. تشير نتائج هذه الدراسة إلى الأهمية البالغة لخوارزميات التعلم الآلي، ولا سيما خوارزمية آلة المتجهات الداعمة (SVM)، في أبحاث QSTR الخاصة بالمبيدات الفطرية.

1. أظهر نموذج الانحدار الخطي المتعدد (MLR) قدرة تنبؤية جيدة لسمية المركبات قيد الدراسة، سواء خلال مرحلة تدريب النموذج أو عند اختباره على مجموعة المركبات الخارجية.
2. تفوق النموذج المنشأ باستخدام خوارزمية آلة المتجهات الداعمة (SVM) تفوقاً ملحوظاً على أداء نموذج الانحدار الخطي المتعدد من حيث دقة التنبؤ.
3. أثبت كلا النموذجين موثوقية إحصائية في القدرة على التنبؤ بالسمية الحادة للمركبات المدروسة.

6- التوصيات

1. توظيف النماذج المطورة في هذه الدراسة كأداة فعالة للتنبؤ بالسمية الحادة للمبيدات الفطرية التابعة لفئتي مثبطات التنفس ومثبطات التخليق الحيوي للستيرول.
2. بناء نماذج تنبؤية جديدة تعتمد على مجموعة أوسع من المركبات، واستكشاف خوارزميات تعلم آلي إضافية إلى جانب تلك المستخدمة في الدراسة الحالية لتعزيز القدرة التنبؤية ودقة النتائج.

المراجع

1. Gupta, R. C., & Gupta, P. K. (2025). Toxicity of fungicides. In *Veterinary toxicology* (pp. 581–593). **Academic Press**.
2. Fest, C., & Schmidt, K. J. (2012). The chemistry of organophosphorus pesticides. **Springer Science and Business Media**.
3. Hu, W., Liu, C. W., Jiménez, J. A., McCoy, E. S., Hsiao, Y. C., Lin, W., Engel, S. M., Lu, K., & Zylka, M. J. (2022). Detection of azoxystrobin fungicide and metabolite azoxystrobin-acid in pregnant women and children, estimation of daily intake, and evaluation of

- placental and lactational transfer in mice. **Environmental Health Perspectives**, 130(2), 027013.
4. Serra, R. S., Martínez, L. C., Cossolin, J. F. S., Resende, M. T. C. S., Carneiro, L. S., Fiaz, M., & Serrão, J. E. (2023). The fungicide azoxystrobin causes histopathological and cytotoxic changes in the midgut of the honey bee *Apis mellifera* (Hymenoptera: Apidae). **Ecotoxicology**, 32(2), 234–242.
 5. Goetz, A. K., Ren, H., Schmid, J. E., Blystone, C. R., Thillainadarajah, I., Best, D. S., Nichols, H. P., Strader, L. F., Wolf, D. C., Narotsky, M. G., & Rockett, J. C. (2007). Disruption of testosterone homeostasis as a mode of action for the reproductive toxicity of triazole fungicides in the male rat. **Toxicological Sciences**, 95(1), 227–239.
 6. Heise, T., Schmidt, F., Knebel, C., Rieke, S., Haider, W., Pfeil, R., Kneuer, C., Niemann, L., & Marx-Stoelting, P. (2015). Hepatotoxic effects of (tri)azole fungicides in a broad dose range. **Archives of Toxicology**, 89(11), 2105–2117.
 7. Kast-Hutcheson, K., Rider, C. V., & LeBlanc, G. A. (2001). The fungicide propiconazole interferes with embryonic development of the crustacean *Daphnia magna*. **Environmental Toxicology and Chemistry**, 20(3), 502–509.
 8. World Health Organization. (2004). *Pesticide residues in food-2004: Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues, Rome, Italy, 20-29 September 2004*. **Food and Agriculture Organization**.
 9. 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.
 10. Liu, Z., Dang, K., Gao, J., Fan, P., Li, C., Wang, H., Li, H., Deng, X., Gao, Y., & Qian, A. (2022). Toxicity prediction of 1, 2, 4-triazoles compounds by QSTR and interspecies QSTTR models. **Ecotoxicology and Environmental Safety**, 242, 113839.

11. Fishel, F. M., & Dewdney, M. M. (2012). Fungicide resistance action committee's (FRAC) classification scheme of fungicides according to mode of action. University of Florida.
12. World Health Organization. (2010). The WHO recommended classification of pesticides by hazard and guidelines to classification 2009. **World Health Organization**.
13. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., & Zaslavsky, L. (2021). PubChem in 2021: New data content and improved web interfaces. **Nucleic Acids Research**, 49(D1), D1388–D1395.
14. National Library of Medicine (NLM). (2011). <https://www.nlm.nih.gov/>
15. Froimowitz, M. (1993). HyperChem: A software package for computational chemistry and molecular modeling. *BioTechniques*, 14(6), 1010–1013.
16. Todeschini, R., Consonni, V., & Pavan, M. (2001). DRAGON–Software for the calculation of molecular descriptors, rel. 1.12 for Windows. <http://www.disat.unimib/chm>
17. Tang, K. S., Man, K. F., Kwong, S., & He, Q. (1996). Genetic algorithms and their applications. **IEEE Signal Processing Magazine**, 13(6), 22–37.
18. The MathWorks Inc. (2019). MATLAB (Version 9.7.0 (R2019b)).
19. Aiken, L. S., West, S. G., & Reno, R. R. (1991). Multiple regression: Testing and interpreting interactions. **Sage**.
20. Yu, X. (2021). Support vector machine-based model for toxicity of organic compounds against fish. **Regulatory Toxicology and Pharmacology**, 123, 104942.
21. Wu, Z., Zhu, M., Kang, Y., Leung, E. L., Lei, T., Shen, C., Jiang, D., Wang, Z., Cao, D., & Hou, T. (2021). Do we need different machine learning algorithms for QSAR modeling? A comprehensive assessment of 16 machine learning algorithms on 14 QSAR data sets. **Briefings in Bioinformatics**, 22(4), bbaa321.
22. Golbraikh, A., & Tropsha, A. (2002). Beware of q²! **Journal of Molecular Graphics and Modelling**, 20(4), 269–276.
23. Rücker, C., Rücker, G., & Meringer, M. (2007). Y-randomization—a useful tool in QSAR validation, or folklore. **Journal of Chemical Information and Modeling**, 47(6), 2345–2357.

24. Mauri, A., Consonni, V., Pavan, M., & Todeschini, R. (2006). Dragon software: An easy approach to molecular descriptor calculations. **MATCH Communications in Mathematical and in Computer Chemistry**, 56(2), 237–248.
25. Hewitt, H. G. (1998). Fungicides in crop protection. **Cabi International**.
26. Feng, Y., Huang, Y., Zhan, H., Bhatt, P., & Chen, S. (2020). An overview of strobilurin fungicide degradation: Current status and future perspective. **Frontiers in Microbiology**, 11, 389.
27. Chowdhary, A., Kathuria, S., Xu, J., & Meis, J. F. (2013). Emergence of azole-resistant *Aspergillus fumigatus* strains due to agricultural azole use creates an increasing threat to human health. **PLOS Pathogens**, 9(10), e1003633.
28. Abbod, M., Safaie, N., & Gholivand, K. (2024). Genetic algorithm multiple linear regression and machine learning-driven QSTR modeling for the acute toxicity of sterol biosynthesis inhibitor fungicides. **Heliyon**, 10(16).
29. Patne, A. Y., Dhulipala, S. M., Lawless, W., Prakash, S., Mohapatra, S. S., & Mohapatra, S. (2024). Drug discovery in the age of artificial intelligence: Transformative target-based approaches. **International Journal of Molecular Sciences**, 25(22), 12233.