وزارة التعليم العالي والبحث العالي

جامعة الشهيد حمه لخضر – الوادي

كلية التكنولوجيا

مذكرة نهاية الدراسة لنيل شهادة

ماستر أكاديمي

ميدان : التكنولوجيا

شعبة: هندسة الطرائق

تخصص: هندسة كيميائية

إعداد الطلبة

جاب الله اسماعيل

زيد البخاري

فرحات عبد المالك

الموضوع

دراسة تطبيق طريقة QSPR للتنبؤ بالمحتوى الحراري القياسي لتكوين المركبات العضوية في الطور الغازي باستخدام الواصفات الجزيئية

تاريخ المناقشة: 2024/09/29

لجنة المناقشة:

د. عبد السلام بوغزال	رئيسا
د. بلال خالد	مناقشا
د. عبد الغني سروطي	مؤطر

السنة الجامعية: 2024/2023

الإهـــــــداء

الحمد لله رب العالمين الذي هدانا لنعمة العلم، وما كنا لنهتدي لولا أن هدانا الله، ثم الصلاة والسلام على سيدنا محمد صلى الله عليه وسلم

أهدي هذا العمل إلى:

من أوصانا بهما القرآن الكريم" **وقضى ربك ألا تعبدوا إلا إياه وبالوالدين إحسانا."**

إلى من أنجبتني وربتني، إلى من سهرت الليالي، إلى من كانت تسقيني الدعاء حتى وصلت إلى أسمى المراتب، إلى من ضمتني إلى أحضانها، إلى أمي الغالية أطال الله في عمرها.

إلى معلمي الأول، أبي، الذي علمني العطاء دون انتظار، أطال الله في عمره ورزقه الصحة والعافية.

إلى مصدر فخري واعتزازي: اخوتي واخواتي، حفظهم الله جميعاً.

إلى زوجتي عزيزة شريكة دربي وابنائي حفظهم الله جميعا

إلى من أضاء لي الدرب في سبيل تحصيل العلم، ولو بقدر بسيط من المعرفة، معلمي وأساتذتي الكرام.

إلى رفقاء الدرب الذين كانوا بمثابة إخوة، زملائي وأصدقائي الأعزاء.

إلى أستاذي الفاضل **سروطي عبد الغني** على مجهوداته وكل أساتذتي بقسم هندسة الطرائق.

إلى كل هؤلاء، وبأسمى معاني الوفاء، أهدي هذا العمل

شكـــر وتقديـــر

الحمد لله رب العالمين، والصلاة والسلام على المصطفى صلى الله عليه وسلم، وعلى صحبه ومن تبعهم بإحسان إلى يوم الدين، أما بعد:

أولاً وقبل كل شيء، نحمد الله، خالقنا، على منحه لنا المعرفة والقوة والعزيمة والشجاعة لإتمام هذا العمل المتواضع. ونحن نختتم هذا العمل، نعبر عن خالص شكرنا وامتناننا العميق لكل من دعمنا وساهم، بشكل مباشر أو غير مباشر، في إعداد هذا البحث.

أتوجه بالشكر أولاً إلى الدكتور سروطي عبد الغني، المشرف على هذا البحث، على كل الدعم والمساعدة والتوجيه والنصائح القيمة والتشجيع خلال سنوات دراستي وحتى إتمام هذه المذكرة.

كما أعرب عن امتناني العميق لأعضاء لجنة المناقشة، الدكتور عبد السلام بوغزال، والدكتور خالد بلال، لما قدموه من تعليقات بناءة ومقترحات مفيدة وآراء نقدية أسهمت بشكل كبير في تحسين جودة هذا العمل.

إضافةً إلى ذلك، نقدم أسمى آيات الشكر لجميع أعضاء هيئة التدريس على جهودهم خلال سنوات دراستنا، ولكل أعضاء فريق قسم هندسة الطرائق والبتروكيمياء الذين ساهموا بطريقة أو بأخرى، سواء بشكل مباشر أو من خلال دعمهم المعنوي.

ولا ننسى أن نقدم جزيل الشكر لكل من قدم لنا يد العون في أوقاتنا الصعبة على إتمام هذا العمل.

الملخص

يعد التقييم الدقيق للمحتوى الحراري القياسي لتكوين المركبات واعتماده على درجة الحرارة أمرًا بالغ الأهمية في دراسة الديناميكا الحرارية وسلوك المواد. يعتمد الحساب النظري للإنثالبي القياسي الحراري للتكوين على معادلة معقدة تعتمد على العلاقة بين المتغيرات الحرارية ومعادلة الحالة، حيث لا تتوفر صيغة صريحة لحساب هذه القيمة. لذلك، كان تطوير نماذج QSPR (العلاقة الكمية بين التركيب الكيميائي والخصائص الفيزيائية) ضروريًا للتنبؤ بالخصائص الفيزيوكيميائية للمركبات العضوية، مما يسمح بتقدير دقيق للمحتوى الحراري القياسي لتكوين هذه المركبات.

في هذه الدراسة، تم تطبيق تقنيات الانحدار الخطي المتعدد (MLR) والانحدار غير الخطي باستخدام الشبكات العصبية الاصطناعية (ANN) كأدوات تسهيلية للتنبؤ بالمحتوى الحراري القياسي. اعتمد النموذج على ثلاث مجموعات من الواصفات الجزيئية تم استخراجها من قاعدة بيانات تحتوي على 339 مركبًا عضويًا. بعد الحسابات والتحليل، أظهرت النتائج أن معاملات التوافق الخاصة بالنموذج الخطي المتعدد كانت مرتفعة لكل من مجموعة التدريب ومجموعة التحقق، حيث بلغت $R^2 = 0.99$ لمجموعة التدريب و$R^2= 0.96$ لمجموعة التحقق. أما بالنسبة للانحدار غير الخطي باستخدام الشبكات العصبية، فكانت قيم معامل التحديد $R^2 = 0.985$ لكل من مجموعة التدريب والتحقق، مما يعكس دقة عالية في التنبؤ بالخصائص المستهدفة.

الكلمات المفتاحية: نموذج QSPR، الواصفات الجزيئية، المحتوى الحراري القياسي للتكوين، التنبؤ، درجة الارتباط.

Abstract

The accurate evaluation of the standard enthalpy of formation and its dependence on temperature is critical in the study of thermodynamics and material behavior. The theoretical calculation of the standard thermal enthalpy of formation is based on a complex equation that relates thermal variables and the equation of state, with no explicit formula available. Therefore, the development of QSPR (Quantitative Structure-Property Relationship) models has been essential for predicting certain physicochemical properties, enabling the estimation of the standard enthalpy of formation for organic compounds.

In this study, multiple linear regression (MLR) and nonlinear regression using artificial neural networks (ANN) were applied as tools to facilitate the prediction of the standard enthalpy of formation using three sets of molecular descriptors. The database included 339 organic compounds. After calculations and analysis, the agreement coefficients (for training and validation) for the MLR model indicated R^2 = 0.99 for training and R^2 = 0.96 for validation. As for the nonlinear regression using ANN, the results showed R^2 = 0.985 for both training and validation sets, reflecting high accuracy in predicting the targeted properties.

Keywords: QSPR model, molecular descriptors, standard enthalpy of formation, prediction correlation degree.

فهرس المحتويات

المحتويات	الصفحة
شكر وعرفان	I
ملخص الدراسة بالعربية	II
ملخص الدراسة بالإنجليزية	III
مقدمة	1
الفصل الأول: مراجعة أدبيات المركبات العضوية، الواصفات الجزيئية والخواص الكيميائية	
1.1. المركبات العضوية	3
2.1. أنواع الروابط في المركبات العضويّة:	3
3.1. طبيعة الروابط في المركبات العضوي ة	4
1.3.1. الرابطة التساهمية الأحادية (C-C):	4
2.3.1. الرابطة التساهمية الثنائية (C=C):	4
3.3.1. الرابطة التساهمية الثلاثية (C≡C):	5
4.1 الخصائص الفيزيوكيميائية والديناميكية الحرارية:	5
1.4.1 الخواص الفيزيائية:	5
2.4.1 الخواص الكيميائية:	5
3.4.1 المحتوى الحراري للتفاعل:	6
4.4.1 المحتوى الحراري القياسي للتكوين (Standard Enthalpy of Formation):	7
4.1. العوامل المؤثرة على المحتوى الحراري القياسي للتكوين	8
5.1. طرق تحديد المحتوى الحراري القياسي للتكوين	9
6.1 دراسات سابقة حول تطبيق QSPR لتقدير الإنثالبي القياسي للتكوين في المركبات العضوية	14
الخاتمة	16
الفصل الثاني: النظريات والأساليب المستخدمة	
2- تعريف QSPR	20
1.2. المفاهيم الأساسية في QSPR:	20
1.1.2. العلاقة بين البنية والخواص الكيميائية	21

3.1.2. بناء نماذج QSPR 23
2.2. الواصفات الجزيئية 23
1.2.2. تعريف وتحديد الواصفات الجزيئية 25
2.2.2. اختيار الواصفات الجزيئية المناسبة للمركبات العضوية: 26
3.2. الأساليب الحسابية في QSPR 28
1.3.2. الانحدار الخطي المتعدد 28
2.3.2. الانحدار غير الخطي (MLR) 30
3.3.2. تقنيات تحسين اختيار الواصفات 33
الخاتمة: 34

الفصل الثالث: منهجية العمل

1.3 إنشاء قاعدة البيانات وتحديد الأوصاف الجزيئية 37
2.3 منهجية إعداد قاعدة البيانات 37
1.2.3 تصميم قاعدة البيانات 37
2.2.3 إعداد قاعدة البيانات: 39
1.2.3 طرق تحليل البيانات 39
2.2.3 تحليل الواصفات الجزيئية: 39
3.3.3 منهجية الانحدار الخطي المتعدد باستخدام XLSTAT: 41
الخاتمة: 48

الفصل الرابع: النتائج والمناقشة

1.4. نتائج الانحدار الخطي المتعدد: 49
1.1.4. تحليل ومناقشة نتائج الانحدار الخطي المتعدد: 54
2.4 نتائج الانحدار غير الخطي باستعمال الشبكات العصبونية الاصطناعية (NRN): 55
1.2.4 تحليل ومناقشة نتائج الانحدار غير الخطي باستعمال الشبكات العصبونية الاصطناعية (NRN): 58
خاتمة 59
الخاتمة العامة 61

قائمة المختصرات

QSAR	Relation structure-activité quantitative.
QSPR	Relation structure-propriété quantitative.
EQMA	Erreur quadratique moyenne commise sur l'ensemble d'apprentissage.
EQMT	Erreur quadratique moyenne commise sur l'ensemble du test.
MCE	Moyenne carrés des erreurs.
RMCE	Racine carrée de la moyenne carrés des erreurs.
MAPE	Pourcentage d'erreur absolue moyenne.
DW	Coefficient de Durbin-Watson.
C_p	Coefficient de Mallows.
AIC	Critère d'information d'Akaike (Akaike's Information Criterion.
SBC	Critère Bayésien de Schwarz (Schwarz's Bayesian Criterion).
PC	Critère de prédiction d'Amemiya (Amemiya's Prediction Criterion).
RLM	Régression linéaire multiple.
F	Coefficient de Fisher.
nSK	Nombre de non-H atomes.
SCBO	Somme des ordres de liaison classique (H appauvri).
n_O	Nombre d'atomes d'oxygène.
n_F	Nombre d'atomes de fluor.
n_{HM}	Nombre d'atomes lourds.
R_2	le coefficient de correlation.
MW	Molecular Weight
AMW	Average Molecular Weight
Sv	Sum of Atomic van der Waals Volumes (measured on carbon atom)
Se	Sum of Atomic Electronegativity by Sanderson's Scale (measured on carbon atom)
Sp	Sum of Atomic Polarizabilities (measured on carbon atom)
Si	Sum of First Ionization Potentials (measured on carbon atom)
Mv	Average Atomic van der Waals Volume (measured on carbon atom)
Me	Average Atomic Electronegativity by Sanderson's Scale (measured on carbon atom)
Mp	Average Atomic Polarizability (measured on carbon atom)

Mi	Average First Ionization Potential (measured on carbon atom)
nAT	Number of Atoms
nSK	Number of Non-Hydrogen Atoms
nBT	Number of Bonds
nBO	Number of Non-Hydrogen Bonds
nBM	Number of Multiple Bonds

قائمة الجداول

العنوان	الصفحة
جدول 1.1: يوضح الطرق الحسابية والأساليب التجريبية التقليدية.	12
الجداول 2.1: مقارنة بين الدراسات السابقة لاستخدام نماذج QSPR.	18
الجدول 1.2: بعض الواصفات التي تنتمي مجموعة الواصفات الدستورية حسب برنامج Dragon.	27
جدول 1.4: يبين العلاقة بين الواصفات البنيوية والمعاملات الحرارية.	51
الجدول 2.4: معاملات التوافق للنموذج الخاصة بالانحدار الخطي.	52
الجدول 3.4: المتغيرات الخاصة بنموذج الشبكة العصبونية الاصطناعية.	57
الجدول 4.4: الاوزان الخاصة بنموذج الشبكة العصبونية الاصطناعية.	58
الجدول 5.4: معاملات التوافق للنموذج الخاصة بالانحدار للشبكة العصبونية الاصطناعية.	59

قائمة الأشكال

العنوان	الصفحة
الشكل 1.2: مخطط توضيحي لنموذج العلاقة الكمية بين التركيب والخصائص الكيميائية لـ QSAR/QSPR.	20
الشكل .2.2: يمثل مخطط إجراء استراتيجية تطوير نموذج QSPR/QSAR والتحقق من صحة النماذج.	24
الشكل 3.2: بعض المتغيرات الجزيئية (كتلة من المجموعات الوظيفية).	28
الشكل 1.3: تمثيل تخطيطي لهيكل قاعدة البيانات العلائقية.	37
الشكل 2.3: صورة توضح طريقة البحث عن بيانات المركبات باستخدام رقم سجل CAS على موقع الويب الخاص بـ NIST.	39
الشكل 3.3: صورة توضح نافذة برنامج XLSTAT (اختيار الانحدار الخطي المتعدد).	41
الشكل 3.3: يوضح الرسم البياني بشكل عام كيفية إدخال البيانات وخطوات المرور المؤدية إلى المخرج للشبكة العصبونية.	47
الشكل 1.4: يوضح العلاقة الخطية بين القيم التجريبية والتنبؤية للمحتوى الحراري القياسي لتكوين.	53
الشكل 2.4: توضيح كيف تتوزع القيم التجريبية مع البواقي (Residuals).	54
الشكل 3.4: توضيح كيف تتوزع القيم المتوقعة مع البواقي (Residuals).	54
الشكل 4.4: العلاقة غير الخطية بين القيم التجريبية والتنبؤية لتدريب (Apprentissage).	59
الشكل 5.4: العلاقة غير الخطية بين القيم التجريبية والتنبؤية لتحقق (Validation).	60

المقدمـــة عـامـــة

تعتبر الخصائص الفيزيوكيميائية والديناميكية الحرارية من الجوانب الأساسية في دراسة المواد والأنظمة الكيميائية. تتناول الديناميكا الحرارية التنبؤ بالانتظاميات العامة للخصائص الفيزيوكيميائية للمادة، مما يساعد في فهم العمليات التي تحدث في المواد الصلبة مثل الذوبانية والتآكل [1]. على سبيل المثال، تم تطوير نماذج ديناميكية حرارية لتقدير الخصائص الفيزيوكيميائية للمركبات العضوية، مما يسهل تصميم المركبات بخصائص مثلى [2]. بالإضافة إلى ذلك، تلعب الديناميكا الحرارية دورًا حيويًا في فهم تفاعلات النقل في المواد القائمة على الأسمنت، حيث تؤثر تركيزات الكلوريد ودرجة الحرارة ودرجة التشبع على التفاعلات الفيزيوكيميائية [3]. كما أن النماذج الديناميكية الحرارية تساعد في تحسين تصميم المنتجات في صناعات الأغذية والطاقة الحيوية من خلال التنبؤ بسلوك التوازن الطوري للدهون والزيوت والوقود الحيوي [4]. باختصار، تساهم الخصائص الفيزيوكيميائية والديناميكية الحرارية في تحسين فهمنا للمواد وتطوير نماذج دقيقة لتقدير خصائصها، مما يعزز من كفاءة التصميم والتطبيق في مختلف المجالات الصناعية والعلمية.

شهدت العقود الأخيرة تطورًا ملحوظًا في مجال النمذجة الجزيئية، التي تعتمد على برامج حاسوبية متقدمة لتحليل الخصائص الإلكترونية والطيفية والمغناطيسية للجزيئات، مما أتاح للعلماء إمكانية معالجة المشكلات الكيميائية دون الحاجة إلى إجراء التجارب المعملية. وقد أدى استخدام تقنيات بديلة للتجريب، مثل علاقات الخصائص الكمية (QSPR/QSAR)، إلى ظهور أساليب أكثر كفاءة ودقة في تصميم المواد الكيميائية والتنبؤ بخصائصها، حيث أصبحت هذه الطرق موصى بها في العديد من اللوائح التنظيمية الحديثة [5].

تعد نماذج QSPR/QSAR أدوات مهمة للغاية في تصميم المواد الكيميائية المستخدمة في التطبيقات الصناعية والدوائية، حيث توفر إمكانية التنبؤ بالنشاط البيولوجي والخصائص الفيزيائية والكيميائية والسمية للمواد الكيميائية غير المختبرة [6]، [7]. يتمحور هدف هذا البحث حول تطوير نماذج QSPR جديدة تسهم في التنبؤ بالخصائص الفيزيائية والكيميائية للمركبات المدروسة، وذلك من خلال دمج البيانات التجريبية المتاحة لتطوير نماذج موثوقة والتحقق من صحتها. كما سيتم التركيز في هذه الدراسة على المحتوى الحراري القياسي كخاصية أساسية للتنبؤ بها.

تهدف هذه الدراسة إلى تطوير نماذج QSPR جديدة لتنبؤ الخصائص الفيزيائية والكيميائية للمركبات العضوية. استنادًا إلى الدراسات السابقة، سيتم دمج البيانات التجريبية المتاحة لتطوير نماذج تنبؤية دقيقة والتحقق

من صحتها. يركز هذا البحث على المحتوى الحراري القياسي للمركبات العضوية، مع تقديم منهجية شاملة لتطوير النماذج التنبؤية وتحليل النتائج.

تتكون هذه الدراسة من أربعة فصول، حيث يتناول الفصل الأول شرحًا تفصيليًا للمركبات العضوية وأهمية الخصائص الفيزيائية والديناميكية الحرارية، بالإضافة إلى مراجعة للمناهج التنبؤية. يتطرق الفصل الثاني إلى تعريف النموذج المدروس (QSPR) وشرح الواصفات الجزيئية والتقنيات المستخدمة في النمذجة. في الفصل الثالث، يتم توضيح منهجية إعداد قاعدة البيانات ومراحل الحساب باستخدام البرمجيات المختلفة. أما الفصل الرابع، فيخصص لعرض النتائج التي تم الحصول عليها ومناقشتها بالتفصيل، مع تطبيق التقنيات المطورة على العديد من المركبات الكيميائية.

الفصل الاول

مراجعة الأدبيات: المركبات العضوية، الواصفات الجزيئية والخواص الكيميائية

مراجعة الأدبيات: المركبات العضوية، الواصفات الجزيئية والخواص الكيميائية

يتناول هذا الفصل دراسة المركبات العضوية وخصائصها الفيزيائية والديناميكية الحرارية، والتي تشكل الأساس لفهم سلوك هذه المركبات في التفاعلات الكيميائية المختلفة. تبرز أهمية هذه الخصائص في التطبيقات الصناعية والبحثية المتعلقة بمجالات مثل الكيمياء العضوية والهندسة الكيميائية. يشمل الفصل أيضًا مناقشة تفصيلية للمحتوى الحراري للتفاعل، والذي يعتبر أحد العوامل الحاسمة في تقييم التفاعلات الكيميائية. سيتم تعريف المحتوى الحراري وتوضيح الطرق المختلفة لحسابه، مثل حساب المحتوى الحراري القياسي للاحتراق والتكوين، بالإضافة إلى التحديد التجريبي لهذه القيم. كما يتم تسليط الضوء على الطرق الحسابية المستخدمة لتقدير هذه الخصائص، بما في ذلك طرق مساهمة المجموعات وطريقة بينسون، مع التركيز على توسيع هذه الطرق لتشمل الحالات السائلة والصلبة [8]، [9].

1.1. المركبات العضوية

المركبات العضوية تُعرّف بأنها مركبات الكربون، وغالبًا ما تشمل الهيدروجين، وقد تحتوي أيضًا على عناصر أخرى مثل الأكسجين والنيتروجين والكبريت والفوسفور والهالوجينات [10]، [11]. تاريخيًا، كان مصطلح "عضوي" يشير إلى المركبات التي تُستخرج من الكائنات الحية، لكن هذا المفهوم تغير بعد تخليق اليوريا من سيانات الأمونيوم (NH_4CNO). في وقت لاحق، نجح كولبي (1845) في تحضير حمض الأسيتيك، أول مركب عضوي في المختبر [10]. مما أثبت أن المركبات العضوية يمكن تخليقها من مواد غير عضوية [10]. تتميز المركبات العضوية بروابط تساهمية، حيث يمكن لذرات الكربون أن تشكل سلاسل طويلة أو حلقات من خلال عملية تُعرف بالتكتل، وهي خاصية فريدة للكربون وتعد سمة مميزة للمركبات العضوية [10]. يمكن تصنيف المركبات العضوية بناءً على بنيتها، مثل الهيدروكربونات (الألكانات، الألكينات، الألكاينات، والهيدروكربونات العطرية) [12]. ومع ذلك، لا تُصنّف جميع المركبات التي تحتوي على الكربون كمركبات عضوية؛ فعلى سبيل المثال، تُعد الكربونات والبيكربونات والسيانيدات وبعض أكاسيد الكربون مركبات غير عضوية رغم احتوائها على الكربون [11]، [13]. تتنوع المركبات العضوية بشكل كبير في تعقيدها وترتبط غالبًا بالعمليات البيولوجية، ويمكن تمثيلها بنواة كربونية مع مجموعات تفاعلية أو سلاسل جانبية تمنحها خصائص معينة [14].

1.2. أنواع الروابط في المركبات العضويّة:

تتميز المركبات العضوية بتنوع كبير في الروابط الكيميائية التي تحدد بنيتها وخصائصها وتفاعليتها. تشمل الأنواع الرئيسية للروابط في الكيمياء العضوية روابط **سيجما** (σ) وروابط **باي** (π) والروابط **الهيدروجينية**. فهم هذه الروابط أساسي لفهم سلوك الجزيئات العضوية وتفاعلاتها.

تتشكل روابط سيجما من تداخل مباشر للمدارات الذرية الهجينة، بينما تتشكل روابط باي من تداخل جانبي للمدارات الذرية p. تلعب هاتان النوعان من الروابط دورًا حاسمًا في تحديد الشكل ثلاثي الأبعاد والخصائص الكيميائية للمركبات العضوية [15]. وقد اقترحت دراسات حديثة إمكانية وجود روابط π أحادية (C–π–C)في بعض الجزيئات، رغم ندرتها مقارنة بالروابط سيجما وباي التقليدية [16].

يساهم مفهوم التهجين في تقديم صورة أكثر دقة لبنية الجزيئات من خلال تفسير تكوين الروابط سيجما عبر تداخل المدارات الهجينة [15]. تلعب الروابط الهيدروجينية دورًا مهمًا في تنظيم البنية الفراغية للعديد من المركبات العضوية، كما تساهم في خصائص مثل التوصيل الكهربائي والأيوني في المواد العضوية [17]. بالإضافة إلى ذلك، تلعب الروابط الهيدروجينية الضعيفة، مثل تفاعلات CH/π، دورًا محوريًا في استقرار الجزيئات وتحديد تفاعلاتها [18]، [19].

توفر نظرية رابطة الزوج الإلكتروني، مدعومة بحسابات ميكانيكا الكم، قواعد أساسية لفهم قوة وطول وزوايا الروابط في الجزيئات العضوية [20]. وفي بعض الحالات، قد تظهر روابط غير تقليدية، مثل روابط مانح-مستقبل، والتي يمكن أن تؤثر على فهمنا للروابط C-C وC-X في بعض المركبات العضوية [21].

1.3. طبيعة الروابط في المركبات العضوية

بما ان الكربون هو العنصر الأساسي في المركبات العضوية، فانه يشكل روابط تساهمية مع ذرات أخرى لتكوين مركبات عضوية متنوعة. يمكن أن تكون هذه الروابط تساهمية أحادية، ثنائية، أو ثلاثية، وكل نوع من هذه الروابط له خصائصه وتطبيقاته في الكيمياء العضوية.

1.3.1. الرابطة التساهمية الأحادية (C-C):

- الروابط التساهمية الأحادية بين ذرات الكربون هي الأساس في بناء هياكل المركبات العضوية، وتستخدم في تفاعلات مثل تفاعل غرينارد وتفاعلات الألدول [22] [23].
- يمكن تفعيل وكسر الروابط التساهمية الأحادية باستخدام محفزات معدنية، مما يتيح مسارات جديدة في التخليق العضوي [24].

1.3.2. الرابطة التساهمية الثنائية (C=C):

- الروابط التساهمية الثنائية بين ذرات الكربون تلعب دورًا مهمًا في تفاعلات الإضافة والتفاعلات الحلقية، وتستخدم في تخليق مركبات معقدة [13]، [15].
- يمكن استخدام مركبات غير مشبعة لتكوين روابط C-C و C-Xبطرق فعالة [23].

1.3.3. الرابطة التساهمية الثلاثية (C≡C):

- الروابط التساهمية الثلاثية بين ذرات الكربون تستخدم في تخليق مركبات تحتوي على روابط ثلاثية مثل الألكاينات، وتعتبر مهمة في تفاعلات الكربوكسلة [26].
- تفاعلات الكربوكسلة باستخدام ثاني أكسيد الكربون تمثل طريقة فعالة لتكوين روابط C-C جديدة [26].

1.4. الخصائص الفيزيوكيميائية والديناميكية الحرارية:

الخصائص الفيزيوكيميائية والديناميكية الحرارية تشكل جزءًا أساسيًا من دراسة سلوك المواد وتفاعلاتها في مختلف الظروف. تشير الخصائص الفيزيوكيميائية إلى الصفات الفيزيائية والكيميائية للمادة، مثل الكثافة واللزوجة والقدرة التفاعلية، وهي تُظهر كيفية تصرف المواد دون تغيير في تركيبها الكيميائي أو عند تعرضها لتفاعلات كيميائية. من ناحية أخرى، تركز الخصائص الديناميكية الحرارية على تحليل تحويل الطاقة والحرارة بين الأنظمة، ودراسة كيفية تأثير العوامل مثل الحرارة والضغط على المواد وفقًا لقوانين الديناميكا الحرارية. هذه الخصائص مهمة في مجالات الهندسة والكيمياء لفهم وتحسين العمليات الصناعية والطبيعية.

1.4.1 الخواص الفيزيائية :

تلعب الخصائص الفيزيائية للمركبات العضوية دورًا أساسيًا في فهم سلوكها في البيئات المختلفة والتطبيقات الصناعية والعلمية. تشمل هذه الخصائص درجات الغليان والانصهار، الكثافة، الضغوط البخارية، وقابلية الذوبان في الماء. يمكن تقدير هذه الخصائص باستخدام البيانات الهيكلية الجزيئية، حيث تساعد التفاعلات السطحية الجزيئية (MSI) في حساب درجات الحرارة الحرجة، الحجم المولي الحرجي، الضغط الحرجي، ودرجة الغليان للمركبات العضوية النقية [27]. بالإضافة إلى ذلك، يؤثر التماكب (isomerism) بشكل كبير على الخصائص الفيزيائية، حيث يؤدي تغيير مواضع الروابط داخل الجزيء إلى اختلافات في هذه الخصائص [28].

2.4.1 الخواص الكيميائية :

المركبات العضوية تتميز بعدة خواص كيميائية تجعلها فريدة في الكيمياء. من أبرز هذه الخواص اعتمادها على الروابط التساهمية، حيث تتشكل معظم المركبات العضوية من روابط قوية بين ذرات الكربون وعناصر أخرى مثل الهيدروجين والأكسجين والنيتروجين، مما يعطيها استقرارًا وتنوعًا في التفاعلات الكيميائية [29]. تتميز المركبات العضوية أيضًا بخاصية الأيزومرية، حيث يمكن لجزيئاتها أن تمتلك نفس الصيغة الجزيئية، ولكن بتراكيب مختلفة، مما يؤدي إلى اختلاف في خواصها الكيميائية والفيزيائية [30]. بالإضافة إلى ذلك، فإن لهذه المركبات خواص حرارية مميزة تتأثر بالتغيرات في درجات الحرارة، مثل نقاط الغليان والانصهار، حيث تعتمد هذه الخواص على التركيب الجزيئي [31]. كما تتميز المركبات العضوية بتنوع في التفاعلات الكيميائية مثل تفاعلات الإضافة والاستبدال، التي تلعب دورًا كبيرًا في الكيمياء العضوية الصناعية والطبيعية [32].

3.4.1 المحتوى الحراري للتفاعل:

المحتوى الحراري للتفاعل هو كمية الحرارة التي تُمتص أو تُطلق أثناء تفاعل كيميائي. يُعبّر عنه بوحدة الكيلوجول لكل مول(kJ/mol) ، وهو يُعدّ من أهم الخصائص الديناميكية الحرارية للمركبات العضوية، حيث يؤثر في استقرار المركبات وتوازن التفاعلات الكيميائية. يتم حساب المحتوى الحراري للتفاعل من خلال فروقات الطاقة بين المتفاعلات والنواتج، ويعتمد ذلك على القيم القياسية للمحتويات الحرارية للتكوين أو الاحتراق. إحدى الطرق الشائعة لحساب هذه القيم هي استخدام "طريقة الإضافة للمجموعات" التي تستند إلى مساهمة المجموعات الوظيفية في المركب [33]. بالإضافة إلى ذلك، تم تطوير طرق تحليلية تعتمد على تفاعلات متجانسة لزيادة دقة التنبؤ بالقيم القياسية للمحتوى الحراري [34].

لإيضاح مفهوم المحتوى الحراري للتفاعل، يمكن تقديم عدة أمثلة:

1. **التفاعلات الهيدروجينية** :في حالة تفاعل الهيدروجين مع المركبات العضوية مثل الألكينات أو الألكاينات، يتطلب هذا التفاعل طاقة حرارية تتعلق بتكسير الروابط الثنائية أو الثلاثية وتكوين روابط جديدة. على سبيل المثال، تفاعل هدرجة الإثيلين إلى إيثان يعتبر تفاعلًا ماصًا للطاقة في البداية بسبب كسر الروابط، ولكنه يُطلق كمية طاقة عند تكوين الروابط الجديدة [35].

2. **التكوينات المتجانسة** :الطريقة المتجانسة تُستخدم في حساب المحتوى الحراري للتكوينات العضوية مثل الأوكسيتانات، حيث يتم تقسيم المركب إلى مجموعات داخلية ويتم حساب المحتوى الحراري بناءً على

تفاعلات بين هذه المجموعات. مثال على ذلك هو حساب المحتوى الحراري لتكوين 3،3-ديميثيل أوكسيتان، حيث يتم تطبيق عدة تفاعلات متجانسة [36].

3. **تفاعل احتراق الهيدروكربونات** :يُعد احتراق الهيدروكربونات مثالًا شائعًا على تفاعل إطلاق الطاقة. يتم حساب المحتوى الحراري لهذا التفاعل باستخدام البيانات القياسية، والتي تتضمن تفاعلات مثل احتراق الميثان (الشكل 1.1) لإنتاج ثاني أكسيد الكربون والماء. هذه التفاعلات تُطلق كمية كبيرة من الطاقة الحرارية، مما يجعلها أساسية في استخدامات الوقود [37].

$CH_4(g)+2O_2(g)\rightarrow CO_2(g)+2H_2O\ (g)+2H_2O(l)\ \Delta H = -890.3\ kJ/mol$ [38].

الشكل 1.1: تفاعل احتراق الميثان لإنتاج ثاني أكسيد الكربون والماء

4.4.1 المحتوى الحراري القياسي للتكوين (Standard Enthalpy of Formation):

هو التغير في الطاقة الذي يحدث عند تكوين 1 مول من مادة معينة من عناصرها الأساسية في حالتها القياسية عند ضغط 1 بار ودرجة حرارة 298.15 كلفن. هذا المفهوم يعتبر أساسيًا في الديناميكا الحرارية لأنه يسمح بحساب الطاقة المطلوبة أو المحررة أثناء التفاعلات الكيميائية. يتم حساب المحتوى الحراري القياسي للتكوين عادة باستخدام طرق مثل **طريقة الإضافة للمجموعات** التي تعتمد على مساهمة المجموعات الوظيفية في المركب. على سبيل المثال، يتم استخدام هذه الطريقة لحساب المحتوى الحراري القياسي لمركبات عضوية معقدة بناءً على المساهمة الترموديناميكية لكل مجموعة وظيفية داخل المركب [39]. كما يمكن استخدام الطرق التجريبية مثل **التفاعلات المتجانسة** للحصول على قيم دقيقة للمركبات العضوية المعقدة [33].

فيما يلي امثلة لحساب المحتوى الحراري القياسي لتكوين:

1. **حساب المحتوى الحراري القياسي لتكوين مركب عضوي باستخدام طريقة الإضافة للمجموعات**

لنأخذ **الإيثان** كمثال. يتم حساب المحتوى الحراري القياسي للتكوين من خلال معادلة تكوين الإيثان من عناصره الأساسية:

$$2C\ (s)+3H_2(g)\rightarrow C_2H_6(g)$$

في هذا المثال، يُحسب المحتوى الحراري القياسي للتكوين عن طريق جمع مساهمات المجموعات الوظيفية، حيث إن الكربون والهيدروجين يشكلان مركب الإيثان. تعتمد هذه الطريقة على القيم المحسوبة لمساهمات المجموعات [39].

2. التفاعلات المتجانسة لحساب المحتوى الحراري القياسي:

لنأخذ **3،-3 ديميثيل أوكسيتان** كمثال. التفاعل المتجانس المستخدم لحساب المحتوى الحراري القياسي لهذا المركب يمكن أن يُمثل كالتالي:

$$C_3H_6O\ C_3H_6+O_2\rightarrow$$

في هذا التفاعل، يتم تقسيم المركب إلى مجموعاته الوظيفية الأساسية (مثل مجموعات الألكانات والأكسجين)، ثم يتم حساب المحتوى الحراري لكل منها وتجميعها للحصول على المحتوى الحراري الكلي. يتم استخدام هذه الطريقة للحصول على تقديرات دقيقة للطاقة المطلوبة [36].

3. مثال على التنبؤ بالمحتوى الحراري القياسي باستخدام Gaussian-4:

لنأخذ تفاعل **احتراق الميثان** كمثال على حساب المحتوى الحراري القياسي باستخدام طريقة Gaussian-4. المعادلة الكيميائية لتكوين ثاني أكسيد الكربون والماء هي:

$$CO_2(g)+2H_2O(l) \rightarrow CH_4(g)+2O_2(g)$$

يتم حساب المحتوى الحراري القياسي للتكوين باستخدام تقنيات المحاكاة الكمية مثل Gaussian-4 للحصول على تقديرات دقيقة للطاقة المتحررة أو المستهلكة خلال التفاعل [40].

4 العوامل المؤثرة على المحتوى الحراري القياسي للتكوين

المحتوى الحراري القياسي للتكوين يتأثر بعدة عوامل تؤثر بشكل مباشر أو غير مباشر على الطاقة التي يتم إطلاقها أو امتصاصها أثناء تكوين المركبات من عناصرها الأساسية. من بين هذه العوامل:

a) **المجموعات الوظيفية**:

وجود مجموعات وظيفية مثل الهيدروكسيل (-OH) أو الأمينو (-NH2) يؤثر على المحتوى الحراري للتكوين، حيث تعتمد الطاقة الكلية للمركب على مساهمات هذه المجموعات الفردية. وقد تم تطوير نظام يعتمد على زيادة المجموعات لتقدير المحتوى الحراري القياسي للمركبات العضوية [39].

b) **الهيكل الجزيئي**:

يعتبر هيكل المركب وتوزيع الروابط بين الذرات عاملاً رئيسيًا. المركبات ذات الروابط الأقوى، مثل الروابط المزدوجة أو الثلاثية، تتطلب طاقة أكبر لتكوينها أو كسرها مقارنة بالروابط الأحادية، مما يؤدي إلى اختلافات في المحتوى الحراري [41].

c) **التفاعلات بين المجموعات الوظيفية**:

تتأثر دقة حساب المحتوى الحراري القياسي أيضًا بالتفاعلات بين المجموعات الوظيفية داخل المركب. عند حساب المحتوى الحراري لمركبات متعددة الوظائف، يجب أخذ هذه التفاعلات في الاعتبار للحصول على نتائج دقيقة [33].

d) **الحالة الفيزيائية**:

الحالة الفيزيائية (غاز، سائل، أو صلب) تؤثر على المحتوى الحراري القياسي. تختلف القيم بناءً على الطاقة المطلوبة لتغيير الحالة الفيزيائية للمادة، حيث يتم تعديل المحتوى الحراري ليعكس الحالة الفيزيائية عند درجة حرارة وضغط معيّنين [42].

e) الواصفات الجزيئية:

استنادًا إلى قائمة الواصفات الجزيئية المحسوبة باستخدام برنامج Dragon، هناك عدة عوامل تؤثر على المحتوى الحراري القياسي للتكوين. هذه العوامل تشمل الواصفات الجزيئية التالية:

أ. **الوزن الجزيئي** (MW):

الوزن الجزيئي للمركب يمكن أن يؤثر بشكل مباشر على المحتوى الحراري، حيث تزيد كتلة الجزيئات الكبيرة من الطاقة المطلوبة لتكوينها [39].

ب. **مجموع أحجام فان دير فالس الذرية** (Sv):

أحجام الذرات وقطر الجزيء يؤثران على كيفية تكوين الروابط، مما يؤثر على الطاقة الإجمالية المطلوبة لتكوين المركب.

ج. **عدد الروابط المتعددة** (nBM) و**عدد الروابط الثنائية والثلاثية** (nDB) **و**(nTB):

كلما زاد عدد الروابط المتعددة في المركب، زادت الطاقة المطلوبة أو المنبعثة أثناء تكوين هذه الروابط، مما يؤثر على المحتوى الحراري للتكوين [41].

د. **عدد الذرات غير الهيدروجينية** (nSK) و**عدد الروابط غير الهيدروجينية** (nBO):

هذه الواصفات ترتبط مباشرة بتعقيد الجزيء وتأثير ذلك على الطاقة المطلوبة لتكوين المركب [33].

باختصار، المحتوى الحراري القياسي للتكوين هو خاصية معقدة تتأثر بعدة عوامل متداخلة. فهم هذه العوامل يساعد في تفسير سبب اختلاف المحتوى الحراري للتكوين بين المركبات المختلفة، وفي التنبؤ بخصائص المركبات الجديدة.

5.1. طرق تحديد المحتوى الحراري القياسي للتكوين

1.5.1. الأساليب الحسابية

- **طريقة إضافة المجموعات**:

تعتمد هذه الطريقة على تقسيم الجزيء إلى مجموعات وظيفية مختلفة، ثم جمع مساهمات كل مجموعة في تحديد المحتوى الحراري. تُعتبر طريقة بينسون من أشهر الأساليب المستخدمة، حيث توفر نتائج دقيقة للمركبات الهيدروكربونية والهيدروكربونات المؤكسجة [43].

- **طريقة Gaussian-4**:

تُستخدم هذه الطريقة في حساب المحتوى الحراري القياسي للمركبات العضوية ذات الغلاف الإلكتروني المغلق، وهي تعتمد على أساليب الكيمياء الكمومية، مثل CCSD(T) الذي يقلل من التكلفة الحسابية مع دقة جيدة [44].

- **استخدام الشبكات العصبية الاصطناعية (ANN):**

تعتمد هذه الأساليب على إدخال الواصفات الجزيئية في نموذج التعلم الآلي للتنبؤ بالمحتوى الحراري القياسي. يُستخدم هذا الأسلوب لتحسين دقة الحسابات الكمومية التي تعتمد على مستوى نظري منخفض [45].

2.5.1. الأساليب التجريبية

- **التحليل الحراري (Calorimetry)**:

هو الطريقة التقليدية لحساب المحتوى الحراري، حيث تُقاس كمية الحرارة التي تُمتص أو تُطلق أثناء التفاعل. يتم استخدام هذه الطريقة للحصول على القيم التجريبية الدقيقة.

- **قياس الانحلال التفاضلي**:

يستخدم لتحديد المحتوى الحراري القياسي للتكوين في المواد الصلبة والسوائل، ويعتبر إحدى التقنيات الحديثة التي توفر دقة عالية في قياس الحرارة الناتجة عن التفاعلات الكيميائية [37].

جدول 1.1: يوضح الطرق الحسابية والأساليب التجريبية التقليدية.

الأسلوب	دقة الحساب	سرعة الأداء	المراجع
طريقة إضافة المجموعات	عالية	متوسطة	[43]
Gaussian-4	عالية	منخفضة	[44]
الشبكات العصبية الاصطناعية	متوسطة	عالية	[45]
التحليل الحراري	عالية	منخفضة	[37]

هذه الأساليب تتيح تقدير المحتوى الحراري القياسي للتكوين بدقة وكفاءة، سواء من خلال الطرق الحسابية التي تعتمد على الكيمياء الكمومية أو من خلال الأساليب التجريبية التقليدية.

3.5.1. التطورات في طرق QSPR (Quantitative Structure-Property Relationship) لحساب المحتوى الحراري القياسي للتكوين:

تُعدّ نماذج **QSPR** من أهم الأساليب في مجال الكيمياء التنبؤية، حيث تُستخدم لحساب الخصائص الفيزيائية والكيميائية للجزيئات بناءً على بنية الجزيء. على مر السنين، شهدت هذه النماذج تطورًا كبيرًا، خاصة في حساب المحتوى الحراري القياسي للتكوين (Standard Enthalpy of Formation) للمركبات العضوية. تعتمد هذه الطرق على استخدام واصفات جزيئية مختلفة (Descriptors) لتحسين دقة الحسابات مع تقليل الجهد والتكلفة.

أ. منهجيات QSPR التقليدية:

كانت الطرق الأولى لنماذج QSPR تعتمد على واصفات بسيطة مثل الطوبولوجيا الجزيئية والخواص الهندسية. وقد تم تطوير نماذج مثل **النماذج متعددة المتغيرات** (Multivariate Models) التي تستند إلى العلاقة بين الخصائص الكيميائية والبنية الجزيئية باستخدام واصفات هندسية أو طوبولوجية.

نذكر على سبيل المثال، قدمت دراسة في عام 2009 نموذج QSPR لحساب المحتوى الحراري للمركبات العضوية بناءً على واصفات طوبولوجية وحسابات بسيطة [46]. تم تطوير هذه النماذج لتتوافق مع المركبات الهيدروكربونية باستخدام مجموعة من المؤشرات الطوبولوجية لتعزيز دقة التنبؤ.

ب. التطورات الحديثة:

مع تقدم الحوسبة، ظهرت طرق أكثر تطورًا تعتمد على الذكاء الاصطناعي والخوارزميات المتقدمة مثل **الخوارزميات الجينية** (GA) و**الشبكات العصبية الاصطناعية** (ANN). هذه الأساليب تقدم تحسينات كبيرة في الدقة والقدرة على التعامل مع مجموعات بيانات كبيرة ومعقدة.

- **النماذج الهجينة** :يعتمد نموذج حديث على الدمج بين **الخوارزميات الجينية** و**الشبكات العصبية الاصطناعية** لتحسين دقة التنبؤ بالمحتوى الحراري للمركبات الهيدروكربونية. وقد أظهر هذا النموذج قدرة عالية على تحسين الدقة مع استخدام مجموعة من المتغيرات البنيوية [26]. **طريقة الأوزان التكيفية** :في هذا التطوير، تُستخدم **الأوزان التكيفية** مع الخوارزميات لتعديل مساهمة كل واصف بناءً على أداء النموذج. تُظهر هذه الطريقة تحسينًا في دقة الحسابات للمركبات المعقدة مثل المركبات العضوية المعدنية [47].

ج. تطبيقات حديثة:

- **الشبكات العصبية:** يتم حاليًا استخدام **الشبكات العصبية** (ANN). بشكل واسع في نماذج QSPR لتحسين القدرة على التنبؤ بالخصائص الكيميائية. على سبيل المثال، قدّمت دراسة في عام 2013 نموذجًا يستخدم ANNمع **التحسين التطوري** لحساب المحتوى الحراري القياسي للتكوين [48]. تم تحسين هذه النماذج باستخدام خوارزميات استكشاف متقدمة مثل خوارزمية السرب الجيني (PSO) لانتقاء أفضل المتغيرات المؤثرة.
- **التحليل الميداني الجزيئي المقارن :(CoMFA)** هذه التقنية تعتمد على الحقول الإلكترونية والستيريكية المحيطة بالجزيئات لتحسين حساب المحتوى الحراري. تم استخدامها مؤخرًا لحساب محتوى التبخر والحرارة للمركبات العطرية متعددة الحلقات (PAHs) [49].

لقد أدت التطورات الأخيرة في نماذج QSPR إلى تحسين كبير في القدرة على التنبؤ بالمحتوى الحراري القياسي للتكوين. هذه النماذج أصبحت أكثر دقة مع الاستخدام المتزايد للتعلم الآلي والخوارزميات الجينية. الاعتماد على الذكاء الاصطناعي والتقنيات الحديثة مثل الشبكات العصبية الاصطناعية والخوارزميات التطورية فتح المجال لتحسين النماذج التقليدية وتوسيع نطاق تطبيقها في الكيمياء العضوية.

4.5.1. دور الواصفات الجزيئية في التنبؤ بالخواص الكيميائية

الواصفات الجزيئية (Molecular Descriptors) تلعب دورًا أساسيًا في التنبؤ بالخواص الكيميائية، حيث تحوّل المعلومات الكيميائية المتعلقة ببنية الجزيء إلى أرقام يمكن استخدامها في النماذج الرياضية والتحليلات الكمية. هذا التمثيل العددي للمركبات يسهل التنبؤ بخصائصها الفيزيائية والكيميائية، مثل النشاط الحيوي والخواص الديناميكية الحرارية.

1.4.5.1. واصفات مؤشرات الطوبولوجيا (Topological Indices):

تُستخدم مؤشرات الطوبولوجيا بشكل واسع في الكيمياء للتعبير عن البنية الجزيئية للمركبات. تعتمد هذه المؤشرات على تمثيل الجزيئات على هيئة رسوم بيانية حيث تمثل الذرات نقاطًا، وتمثل الروابط الخطوط التي تصل بين هذه النقاط. تُستخدم مؤشرات الطوبولوجيا في فهم العلاقة بين بنية الجزيء وخصائصه الكيميائية. فيما يلي نذكر **أهم مؤشرات الطوبولوجيا:**

- **مؤشر Wiener:** يعبر عن المسافة الإجمالية بين جميع أزواج الذرات في الجزيء. يُستخدم هذا المؤشر في التنبؤ بخصائص الجزيء مثل نقطة الغليان والاستقرار الكيميائي.
- **مؤشر Balaban:** يُعتبر مؤشرًا متقدمًا يعتمد على العلاقة بين المسافة وعدد الروابط بين الذرات. يُستخدم هذا المؤشر في نماذج التنبؤ بالنشاط البيولوجي للمركبات، خاصة في تصميم الأدوية [50].

التطبيقات:

- **النشاط البيولوجي** :يتم استخدام مؤشرات الطوبولوجيا لتنبؤ النشاط البيولوجي للمركبات الكيميائية، خاصةً في تصميم الأدوية والتنبؤ بالتفاعلات بين الجزيئات والبروتينات.
- **الخواص الفيزيائية** :تُستخدم مؤشرات الطوبولوجيا أيضًا في التنبؤ بالخواص الفيزيائية مثل نقطة الغليان، نقطة الانصهار، والكثافة.

2.4.5.1. مؤشرات الاتصال (Connectivity Indices):

مؤشرات الاتصال تعتمد على هيكل الجزيء وكيفية ارتباط الذرات ببعضها البعض. تُستخدم هذه المؤشرات لتقدير الخصائص الفيزيائية والكيميائية للمركبات من خلال تقييم مدى ترابط الذرات داخل الجزيء.

- **أهم مؤشرات الاتصال:**
- **مؤشر** :**Randic** يعتمد على درجة الاتصال بين الذرات في الجزيء، ويُستخدم بشكل رئيسي لتنبؤ الخواص الفيزيائية مثل الذوبانية والنشاط الحيوي للمركبات.
- **مؤشر** :**Hosoya** يعبر عن عدد الطرق المختلفة التي يمكن فيها تقسيم الجزيء إلى أجزاء أصغر، وهو مؤشر مرتبط بتعقيد بنية الجزيء، ويُستخدم في النماذج التنبؤية للنشاط الكيميائي والاستقرار الحراري.

التطبيقات:

- **التنبؤ بالخواص الديناميكية الحرارية** :تُستخدم مؤشرات الاتصال في نماذج التنبؤ بالخواص الديناميكية الحرارية للمركبات مثل التفاعل الكيميائي والذوبانية [51].
- **التصميم الجزيئي** :تُستخدم هذه المؤشرات بشكل كبير في تصميم المركبات الدوائية لتقدير النشاط الحيوي للمركب قبل إجراء الاختبارات التجريبية.

3.4.5.1. واصفات المؤشرات الدستورية (Constitutional indices):

المؤشرات الدستورية هي مجموعة من الواصفات الجزيئية التي تعتمد على البنية الجزيئية الأساسية للمركب، بما في ذلك عدد الذرات، الروابط، المجموعات الوظيفية، والكتلة الجزيئية. هذه الواصفات لا تعتمد على التوجيهات المكانية للجزيء(3D) ، بل توفر وصفًا بسيطًا وسهل الفهم للخصائص الكيميائية والفيزيائية للمركب.

5.5.1. دور المؤشرات الدستورية في التنبؤ بالخواص الكيميائية:

1. **الوزن الجزيئي** :يُعتبر الوزن الجزيئي (Molecular Weight, MW) من أهم المؤشرات الدستورية، ويُستخدم على نطاق واسع للتنبؤ بالخواص الكيميائية مثل الذوبانية، درجة الغليان، والنشاط الحيوي للمركبات. كلما زاد الوزن الجزيئي، زاد تأثيره على الخصائص الفيزيائية.
2. **عدد الروابط** :عدد الروابط الأحادية (Single Bonds) والثنائية (Double Bonds) يؤثر بشكل كبير على ثباتية المركب وخواصه التفاعلية. يُعتبر هذا النوع من المؤشرات الدستورية مهمًا في نماذج التنبؤ بالتفاعلات الكيميائية [52].
3. **عدد الذرات الثقيلة** :تمثل الذرات الثقيلة (Heavy Atoms) تلك الذرات التي ليست هيدروجينًا، ويستخدم هذا المؤشر لتحديد حجم الجزيء وتأثيره على الخصائص الكيميائية مثل النشاط الحيوي

والتفاعل الكيميائي. هذا الواصف يمكن أن يكون جزءًا مهمًا في نماذج QSAR التي تركز على التفاعلات البيولوجية [53].

4. **عدد المجموعات الوظيفية**: وجود مجموعات وظيفية محددة مثل الكحولات (OH-)، الأحماض الكربوكسيلية (COOH-)، والأمينات (NH2-) يؤثر على النشاط الكيميائي للمركب وقدرته على التفاعل مع جزيئات أخرى. تُستخدم هذه المؤشرات الدستورية في نماذج التنبؤ بالتفاعلات الكيميائية والنشاط الحيوي.

6.5.1. أهمية المؤشرات الدستورية في النماذج التنبؤية:

تُعد المؤشرات الدستورية بسيطة، ولكنها فعالة في إنشاء نماذج QSPR (العلاقة الكمية بين البنية والخصائص) وQSAR (العلاقة الكمية بين البنية والنشاط). تُستخدم هذه الواصفات بشكل خاص في الحالات التي تكون فيها الحسابات الكمومية المعقدة غير مطلوبة، مما يجعلها أداة عملية وسريعة للتنبؤ بالخواص الفيزيائية والكيميائية للمركبات.

A. مثال على تطبيق المؤشرات الدستورية:

في إحدى الدراسات التي استخدمت المؤشرات الدستورية لتنبؤ الذوبانية والهيدروفيلية، تم الاعتماد على الواصفات مثل الوزن الجزيئي وعدد الروابط الثنائية وعدد الذرات الثقيلة. هذه المؤشرات قدمت نتائج دقيقة وقابلة للتطبيق في مجموعة كبيرة من المركبات العضوية، بما في ذلك المركبات الدوائية [54].

B. التطبيقات الحديثة للواصفات الدستورية:

تم استخدام الواصفات الدستورية بشكل فعال في خوارزميات التعلم الآلي التي تعتمد على قواعد بيانات كبيرة من المركبات الكيميائية. على سبيل المثال، يمكن دمج الواصفات الدستورية مع خوارزميات مثل الغابات العشوائية (Random Forest) لتحسين التنبؤ بالنشاط الحيوي للمركبات الكبيرة [26].

تمثل المؤشرات الدستورية أساسًا قويًا في التنبؤ بالخواص الكيميائية للمركبات نظرًا لبساطتها وكفاءتها في وصف البنية الجزيئية الأساسية. هذه الواصفات توفر وسيلة سريعة لتحليل المركبات وتنبؤ خواصها الفيزيائية والكيميائية، ويمكن استخدامها بشكل مستقل أو مدمجة مع نماذج أخرى لزيادة الدقة والتفصيل.

1.5.5.1. دراسات سابقة حول تطبيق QSPR لتقدير الإنثالبي القياسي للتكوين في المركبات العضوية

أظهرت العديد من الدراسات تطبيقات مختلفة لنماذج **QSPR** (العلاقة الكمية بين البنية والخصائص) في تقدير الإنثالبي القياسي للتكوين في المركبات العضوية. تم استخدام تقنيات متنوعة مثل الانحدار المتعدد، الشبكات العصبية الاصطناعية، وخوارزميات الغابات العشوائية لتحسين دقة التنبؤ. سنستعرض في هذا التحليل بعض الدراسات البارزة مع عرض النتائج والطرق المستخدمة.

A. دراسة باستخدام الانحدار المتعدد الخطيات المدعوم بالخوارزميات الجينية

في دراسة **(Vatani et al., 2007)**، تم استخدام **الانحدار المتعدد الخطيات المدعوم بالخوارزميات الجينية (GA-MLR)** لتقدير الإنثالبي القياسي لتكوين 1115 مركبًا من مجموعات كيميائية مختلفة. اعتمدت الدراسة على خمس واصفات جزيئية أظهرت تأثيرًا كبيرًا في النموذج مع معامل ارتباط (R^2) يصل إلى 0.983، مما أظهر دقة عالية في التنبؤ [55].

B. تطبيق النماذج غير الخطية باستخدام الشبكات العصبية الاصطناعية

قدمت دراسة (**Bagheri et al., 2013**) والتي تتناول بناء نماذج **QSPR غير خطية** باستخدام الشبكات العصبية الاصطناعية (ANN) لتقدير الإنثالبي القياسي للتكوين لمجموعة متنوعة من المركبات. استند النموذج على مجموعة بيانات شملت 1765 مركبًا من 82 فئة كيميائية مختلفة. أظهرت النتائج دقة عالية مع **خطأ الجذر التربيعي** (RMSE)بلغ 97 كيلوجول/مول في النموذج غير الخطي، مما يوضح تفوق الشبكات العصبية في تحسين التنبؤات مقارنة بالنماذج الخطية التقليدية [48].

C. نموذج باستخدام خوارزميات الغابات العشوائية(Random Forest)

في دراسة أجرتها **(Teixeira et al., 2013)**، تم استخدام خوارزميات الغابات العشوائية لاختيار الواصفات الجزيئية المهمة وتطوير نموذج QSPR لتقدير الإنثالبي القياسي للتكوين في الهيدروكربونات. أظهرت الدراسة أن استخدام الغابات العشوائية أدى إلى تحسين دقة التنبؤات بنسبة 23% مقارنةً بالنماذج التي لا تعتمد على اختيار الواصفات، مع تقليل عدد الواصفات المطلوبة إلى 6% من العدد الإجمالي [56].

D. التقدير باستخدام طريقة المؤشرات المعيارية(Norm Indexes)

في دراسة حديثة من قبل (Xuefeng Yan et al., 2020)، تم تطوير نموذج QSPR يعتمد على مؤشرات معيارية لتقدير الإنثالبي القياسي للتكوين في الحالتين الغازية والسائلة. أظهر النموذج معامل ارتباط

(R^2)بلغ 0.990 في التنبؤ بالحالة الغازية و0.987 في الحالة السائلة، مما يدل على دقة عالية وثبات ممتاز [6].

E. تطبيق على المركبات العضوية المعدنية(Organometallic Compounds)

- في دراسة أخرى أجرتها (Jover et al., 2008)، تم تطوير نموذج QSPR لتقدير الإنثالبي القياسي للتكوين للمركبات العضوية المعدنية. استخدم النموذج تسع واصفات من الأنواع الدستورية والطوبولوجية والهندسية، وحقق دقة ممتازة مع معامل ارتباط (R^2) بلغ 0.988 في مجموعة التدريب و0.990 في مجموعة التنبؤ [57].

الجداول 2.1: مقارنة بين الدراسات السابقة لاستخدام نماذج QSPR.

الدراسة	النموذج المستخدم	حجم مجموعة البيانات	R^2	دقة التنبؤ (RMSE)	المرجع
Vatani et al., 2007	GA-MLR	1115	0.983	20.5 kJ/mol	[55]
Bagheri et al., 2013	ANN	1765	0.990	97 kJ/mol	[48]
Teixeira et al., 2013	Random Forest	800	0.960	23% تحسين في الدقة	[56]
Jover et al., 2008	MLR	132	0.988	29.1 kJ/mol	[57]

يوضح الجدول اعلاه الدراسات السابقة لاستخدام نماذج QSPR بشكل فعال لتحسين دقة التنبؤ بالإنثالبي القياسي للتكوين في المركبات العضوية. تتفاوت النماذج في تعقيدها ودقتها، حيث تقدم النماذج غير الخطية مثل الشبكات العصبية الاصطناعية تفوقًا في الأداء مقارنة بالنماذج الخطية التقليدية.

الخاتمة

في هذا الفصل، تم تناول الأساسيات المتعلقة بالمركبات العضوية وخصائصها الفيزيائية والديناميكية الحرارية، مع التركيز على أهمية هذه الخصائص في السياق العلمي والتطبيقي. تم شرح مفهوم المحتوى الحراري للتفاعل وتوضيح طرق حساب المحتوى الحراري القياسي للاحتراق والتكوين، بما في ذلك الطرق المختلفة لتحديد هذه القيم. كما تم استعراض عدة أساليب حسابية مثل طرق مساهمة المجموعات وطريقة بينسون، بالإضافة إلى توسيع طريقة بينسون لتشمل الحالة السائلة والصلبة. وأخيرًا، تم التطرق إلى استخدام الواصفات الجزيئية كأداة حديثة في حساب المحتوى الحراري، مما يسهم في تحسين دقة التنبؤات الحرارية للمركبات العضوية في مختلف الحالات الفيزيائية.

الفصل الثانـي:

النظريات والأساليب المستخدمة

النظريات والأساليب المستخدمة

تهدف دراسة العلاقة الكمية بين البنية والخاصية (QSPR) إلى فهم كيفية تأثير بنية الجزيء على خواصه الفيزيائية والكيميائية والبيولوجية. يعتمد هذا الفهم على بناء نماذج رياضية تربط بين الواصفات الجزيئية، التي تصف البنية الجزيئية، والخواص المختلفة. يُمكن استخدام هذه النماذج للتنبؤ بخواص مركبات جديدة دون الحاجة إلى إجراء تجارب مكلفة، مما يُسهل عمليات التصميم والاكتشاف في الكيمياء والصيدلة وعلوم المواد.

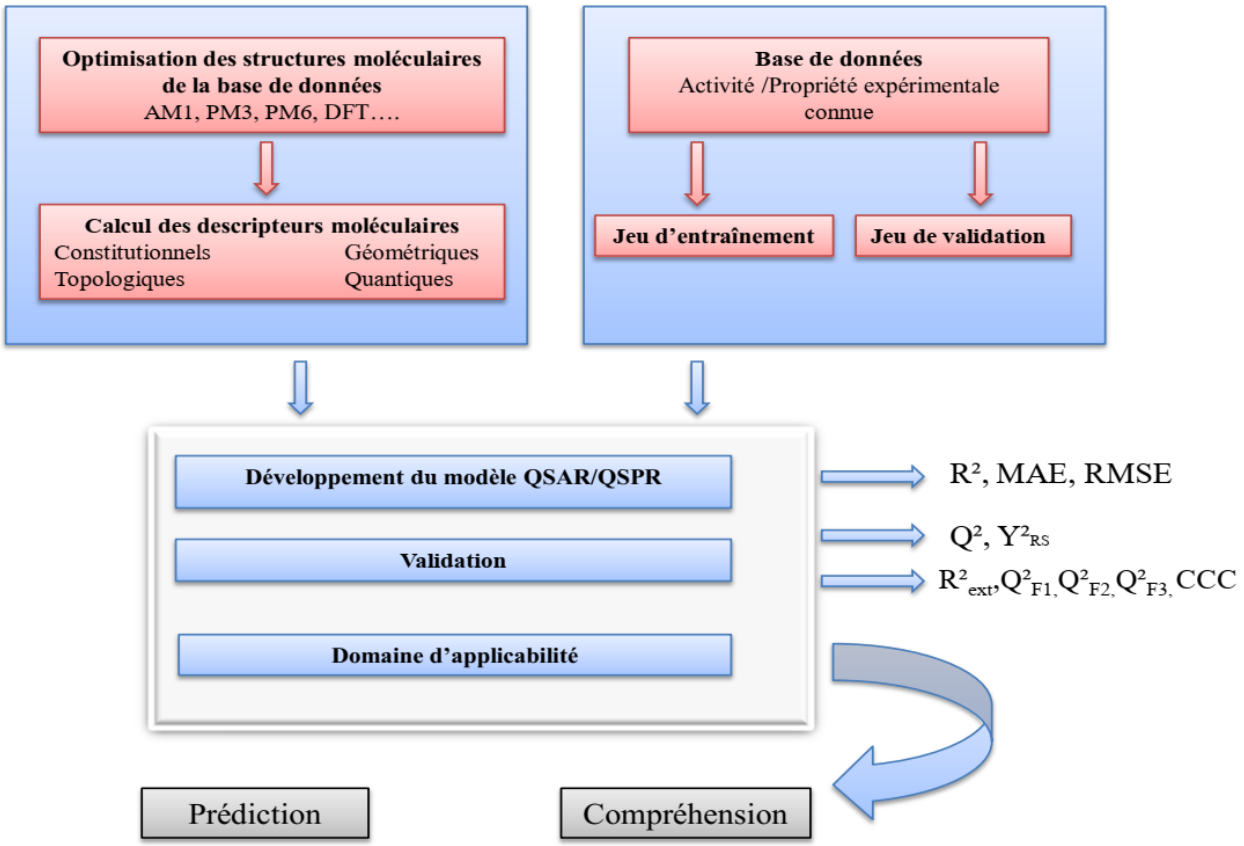
يتناول هذا الفصل بالتفصيل المفاهيم والأساليب المستخدمة في دراسات .QSPR بدايةً، يُقدم الفصل تعريفًا شاملاً لـQSPR ، ثم يناقش المفاهيم الأساسية المتعلقة بالعلاقة بين البنية والخواص، بما في ذلك تأثير البنية على النشاط الكيميائي، والذوبانية، والاستقرار الحراري، والنشاط البيولوجي، والخصائص الإلكترونية. بعد ذلك، يُوضح الفصل خطوات بناء نماذجQSPR ، بدءًا من جمع البيانات الكيميائية وحساب الواصفات الجزيئية، وصولًا إلى اختيار النموذج الرياضي المناسب، والتحقق من صحته، وتحليل النتائج وتطبيق النموذج.

يُركز الفصل أيضًا على الواصفات الجزيئية، مُقدمًا تعريفًا لها وكيفية تحديدها، بالإضافة إلى مناقشة كيفية اختيار الواصفات المناسبة للمركبات العضوية، مع التركيز على المؤشرات الدستورية وواصفات المجموعات الوظيفية. أما بالنسبة للأساليب الحسابية المستخدمة في QSPR، فيُغطي الفصل كلاً من الانحدار الخطي المتعدد (MLR) والانحدار غير الخطي (MnLR)، مُوضحًا مفاهيمها ومزاياها وعيوبها وتطبيقاتها العملية في الكيمياء. كما يتطرق الفصل إلى تقنيات تحسين اختيار الواصفات، مع التركيز على الخصائص الفيزيائية والكيميائية والتفاعلات الجزيئية. في الختام، يُلخص الفصل أهم النقاط التي تم تناولها ويُقدم قائمة بالمراجع المستخدمة.

تعريف QSPR (Quantitative Structure-Property Relationship):

تعتمد نماذج العلاقة الكمية بين الخصائص والتركيب الكيميائي (QSPR) على الفرضية الأساسية بأن هناك ارتباطًا وثيقًا بين بنية الجزيء الكيميائي وخصائصه الفيزيائية أو الكيميائية. وفقًا لهذا النهج، يُعتقد أن الخصائص المميزة للمركبات الكيميائية تتأثر بالبنية الجزيئية، حيث يتم التعبير عن هذه العلاقة من خلال خوارزميات رياضية تستند إلى خصائص جزيئية قابلة للقياس. يعتمد نموذج QSPR على استخدام متغيرات كمية مستمدة من التركيب الكيميائي لتحديد العلاقات الرياضية التي تربط بين التركيب الكيميائي وخصائص المركب، مما يسمح بالتنبؤ بالخصائص قبل تصنيع المركبات الجديدة.

بشكل عام، تعتمد هذه النماذج على استخدام مؤشرات أو واصفات جزيئية (descriptors) تُعبّر عن البنية الكيميائية للمركب، سواء كانت مستمدة من الحسابات الكمية أو التجريبية، ويتم تطبيق تقنيات رياضية متقدمة لتحليل البيانات، مثل الانحدار الخطي المتعدد أو الشبكات العصبية، لإنشاء هذه النماذج [7].



الشكل 1.2: مخطط توضيحي لنموذج العلاقة الكمية بين التركيب والخصائص الكيميائية لـ QSAR/QSPR.

3.1. المفاهيم الأساسية في QSPR:

يستند مبدأ نموذج العلاقة الكمية كما هو موضح في الشكل 1.2 بين التركيب والخصائص الكيميائية (QSPR) إلى صياغة علاقة رياضية تربط الخصائص الجزيئية، سواء كانت إلكترونية أو هندسية، والتي يتم التعبير عنها من خلال الواصفات (Descriptors) بخصائص قابلة للملاحظة مثل النشاط البيولوجي أو السمية أو الخصائص الفيزيائية والكيميائية. يعتمد هذا النموذج على تحليل بيانات مجموعة من المركبات الكيميائية المتشابهة، حيث يتم استخدام أدوات تحليل بيانات إحصائية لاكتشاف المتغيرات التي تؤثر على هذه الخصائص. ويأخذ الشكل العام لهذا النموذج الصيغة التالية:

$$\text{Property Value} = F\ (\text{Descriptors})$$

الهدف من هذا النهج هو تحليل البنية الجزيئية واستخدام الواصفات المناسبة لبناء نموذج قادر على التنبؤ بالخصائص المختلفة للمركبات الجديدة، وذلك باستخدام أدوات مثل الانحدار الخطي المتعدد أو الشبكات العصبية [58]، [59].

3.1.1. العلاقة بين البنية والخواص الكيميائية

تُعتبر العلاقة بين **البنية الجزيئية** و**الخواص الكيميائية** من المفاهيم الأساسية في الكيمياء، حيث تعكس البنية الجزيئية كيفية ترتيب الذرات والروابط داخل الجزيء، وهو ما يؤثر بشكل مباشر على سلوك الجزيء وخصائصه الكيميائية والفيزيائية. يمكن القول إن بنية الجزيء هي المفتاح لفهم خواصه مثل الذوبانية، النشاط البيولوجي، والاستقرار الحراري.

- **البنية الجزيئية وتأثيرها على النشاط الكيميائي**

تؤثر البنية الجزيئية بشكل مباشر على قدرة الجزيء على التفاعل مع الجزيئات الأخرى. فمثلاً، **الشكل الفراغي** للجزيء يحدد كيفية اقتراب الجزيء من المواد الأخرى والتفاعل معها. الجزيئات التي تحتوي على روابط مزدوجة أو ثلاثية تكون أكثر نشاطًا من حيث التفاعل مقارنة بالجزيئات ذات الروابط الأحادية، بسبب الطاقة الأعلى في هذه الروابط. على سبيل مثال الإيثين (C_2H_4) يحتوي على رابطة مزدوجة بين ذرتي الكربون، مما يجعله أكثر تفاعلًا مقارنةً مع الإيثان (C_2H_6) الذي يحتوي على روابط أحادية فقط. يتفاعل الإيثين بسهولة في تفاعلات الإضافة بسبب الرابطة المزدوجة، بينما يحتاج الإيثان إلى ظروف أكثر حدة لإحداث التفاعل [41].

- **البنية والذوبانية**

الذوبانية تتأثر بشكل مباشر بالبنية الجزيئية. الجزيئات ذات الروابط القطبية أو المجموعات الوظيفية القطبية مثل مجموعة الهيدروكسيل (OH-) أو مجموعة الكربوكسيل (COOH-) تكون أكثر قابلية للذوبان في المذيبات القطبية مثل الماء. على العكس من ذلك، المركبات غير القطبية مثل الهيدروكربونات غير قابلة للذوبان في الماء، لكنها تذوب في المذيبات العضوية غير القطبية. مثال على ذلك الإيثانول (C_2H_5OH) يذوب بسهولة في الماء بفضل وجود مجموعة الهيدروكسيل القطبية التي يمكنها تكوين روابط هيدروجينية مع جزيئات الماء.

في المقابل، الهكسان (C_6H_{14})، وهو مركب هيدروكربوني غير قطبي، لا يذوب في الماء، ولكنه يذوب في مذيبات غير قطبية مثل البنزين [60].

- **البنية والاستقرار الحراري**

الاستقرار الحراري للجزيئات يعتمد على البنية الجزيئية والطاقة المطلوبة لتكسير الروابط الكيميائية. الجزيئات ذات الروابط القوية مثل الروابط التساهمية المزدوجة أو المتناسقة تكون أكثر استقرارًا من الجزيئات التي تحتوي على روابط ضعيفة مثل الروابط الهيدروجينية أو الفان دير فالس. مثل البنزين (C_6H_6) الذي يحتوي على بنية حلقية مع روابط مزدوجة متوزعة بشكل متساوٍ (رنين)، مما يعطيه استقرارًا حراريًا عاليًا. بالمقابل، البيوتادايين (C_4H_6) الذي يحتوي على روابط مزدوجة في سلسلة خطية يكون أقل استقرارًا ويحتاج إلى طاقة أقل لتحطيمه في التفاعلات الكيميائية [41].

- **البنية والنشاط البيولوجي**

النشاط البيولوجي للجزيئات يعتمد أيضًا على بنيتها الجزيئية، حيث تلعب الشكل الهندسي والقدرة على تكوين روابط مع الأنزيمات والمستقبلات البيولوجية دورًا حاسمًا في تفاعل المركبات مع الأنظمة الحيوية. يمكن للبنية الجزيئية أن تؤثر على كيفية تفاعل الدواء مع مستقبلات الجسم. مثلما يتم في تصميم الأدوية، والتي تعتمد على تعديل البنية الجزيئية للدواء لتناسب الموقع النشط في الإنزيمات. على سبيل المثال، بعض الأدوية تعتمد على البنية الفراغية للجزيء لتكوين روابط قوية مع مستقبلات معينة في الجسم، مثل الأسبرين الذي يتفاعل مع إنزيم كوكس المسؤول عن الألم والالتهاب [48].

- **البنية والخصائص الإلكترونية**

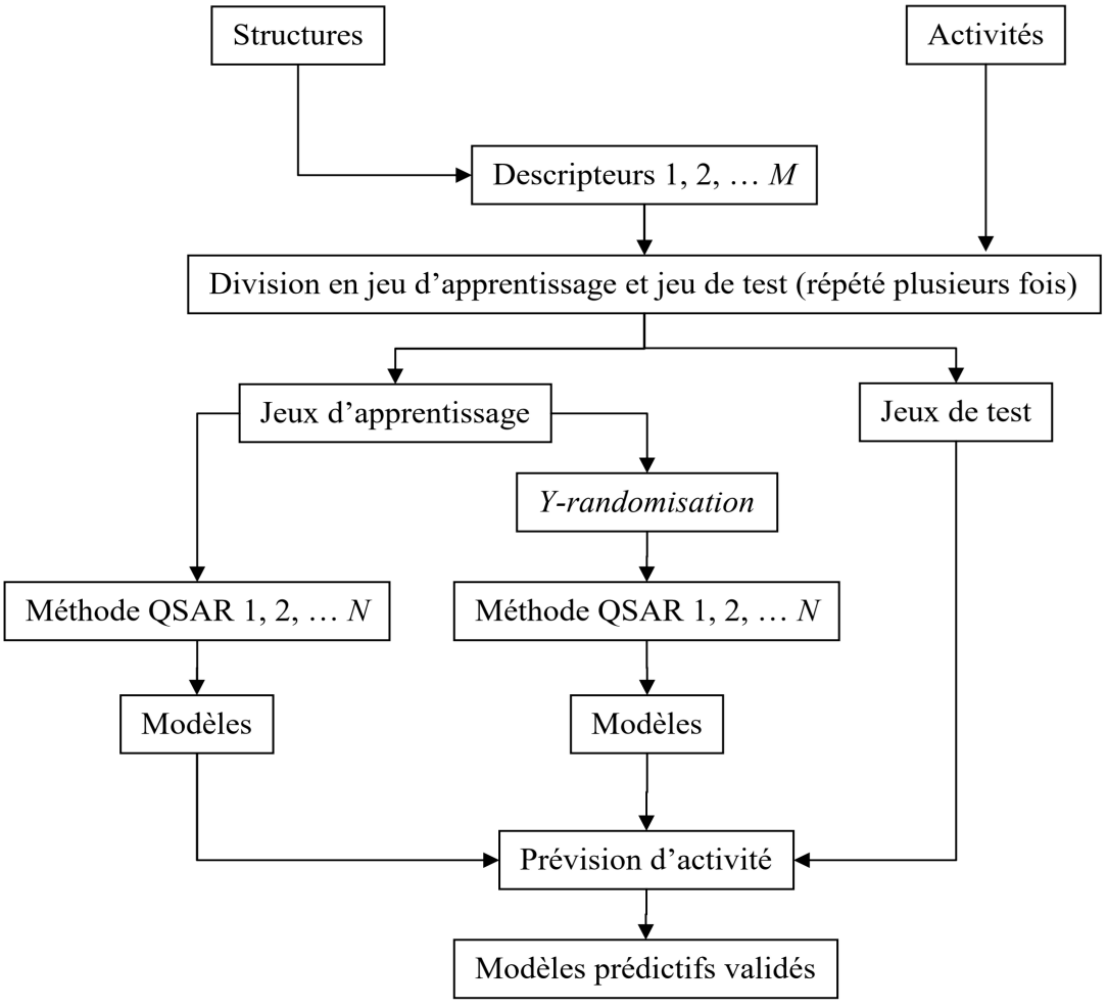
البنية الجزيئية تؤثر على الخصائص الإلكترونية للجزيء، مثل توزيع الشحنة والتفاعلية. المركبات ذات الروابط المزدوجة أو المجموعات الوظيفية القطبية تمتلك توزيعًا إلكترونيًا غير متساوٍ يجعلها قادرة على التفاعل مع الجزيئات الأخرى عن طريق تفاعلات مثل الإلكتروفيلية أو النيوكليوفيلية. كالألدهيدات التي تحتوي على مجموعة الكربونيل (-C=O) القطبية، مما يجعلها عرضة للتفاعل مع النيوكليوفيلات مثل الأنيونات أو الجزيئات المحتوية على أزواج إلكترونات غير مرتبطة. هذه الخاصية تجعل الألدهيدات مركبات أساسية في تفاعلات الإضافة النيوكليوفيلية [41].

3.1.2. بناء نماذج QSPR

إن بناء نموذج QSAR/QSPR يعتمد بشكل كبير على البيانات التجريبية المرجعية. لذلك، يعتبر اختيار قاعدة البيانات خطوة حاسمة في تطويره. في معظم الحالات، تستند البيانات التجريبية إلى الدراسات الأدبية السابقة. عادةً ما تُعتبر البيانات المستخدمة في النمذجة إما كأثر بيولوجي أو كخاصية فيزيائية-كيميائية. تشير نماذج العلاقة الكمية بين البنية والنشاط أو الخاصية (QSAR/QSPR) إلى تطوير نماذج تتوقع الأنشطة البيولوجية المختلفة، بناءً على المعلومات المتعلقة بالتركيبة الجزيئية ومجموعة من المركبات. تهدف هذه النماذج إلى إيجاد علاقة رياضية تربط بشكل كمي بين النشاط البيولوجي أو الخاصية المُقاسة لمجموعة من المركبات المشابهة في ظل نفس الظروف التجريبية، باستخدام واصفات جزيئية وتوظيف الأساليب الإحصائية [61].

تعتبر نماذج العلاقات الكمية بين البنية والنشاط/الخاصية (QSPR/QSAR) أدوات إحصائية تهدف إلى ربط المعلومات الكيميائية الكمية بالنشاط البيولوجي أو الخصائص الأخرى. وفقًا لـ (Fortuné, 2006)، تتكون استراتيجية تطوير نموذج QSPR/QSAR من خمسة نقاط رئيسية. فيما يلي تحليل للأبحاث المتعلقة بهذه النقاط:

1) **تنظيم البيانات وتنقيحها:** تنقيح البيانات وتحديد المتغيرات المناسبة يعتبران من الخطوات الأساسية لتطوير نماذج QSPR/QSAR [62].
2) **اختيار المتغيرات:** استخدام طرق اختيار المتغيرات المختلفة وأدوات التعلم الآلي للتعامل مع تعقيد مجموعات البيانات واختيار الميزات المناسبة للنماذج [62] [63].
3) **تقسيم البيانات:** تقسيم البيانات إلى مجموعات تدريب واختبار لضمان صحة النموذج وتجنب التحيز [62].
4) **اختيار خوارزميات النمذجة:** استخدام استراتيجيات الانحدار والتصنيف لتطوير نماذج استجابة كمية ومصنفة، بالإضافة إلى استخدام أدوات التعلم الآلي لتحليل البيانات المعقدة [63].
5) **التحقق من صحة النموذج:** استخدام تقنيات مثل التحقق المتقاطع المزدوج والنمذجة التوافقية لإزالة التحيز وضمان صحة النموذج [62].



الشكل .2.2: يمثل مخطط إجراء استراتيجية تطوير نموذج QSPR/QSAR والتحقق من صحة النماذج.

تتضمن استراتيجية تطوير نموذج QSPR/QSAR خمس نقاط رئيسية: تنظيم وتنقيح البيانات، اختيار المتغيرات، تقسيم البيانات، اختيار خوارزميات النمذجة، والتحقق من صحة النموذج. هذه الخطوات تضمن تطوير نماذج دقيقة وموثوقة يمكن استخدامها في مختلف التطبيقات الكيميائية والبيولوجية.

أ. جمع البيانات الكيميائية

تبدأ عملية بناء نموذج QSPR بتجميع بيانات الخصائص الكيميائية أو الفيزيائية المراد التنبؤ بها. يجب أن تكون البيانات متاحة بشكل موثوق لمجموعة واسعة من المركبات الكيميائية، سواء كانت متاحة من تجارب سابقة أو من قواعد بيانات. مثلما أجريت دراسة على 645 مركبًا عضويًا لتقدير ضغط البخار عند درجة حرارة معينة، تم تجميع البيانات من قواعد بيانات تحتوي على هذه الخصائص [53].

ب. حساب الواصفات الجزيئية

الواصفات الجزيئية هي الخصائص العددية التي تعبر عن البنية الكيميائية للمركب. يمكن حساب هذه الواصفات باستخدام برامج متخصصة مثل Dragon وCODESSA. وتشمل هذه الواصفات مؤشرات طوبولوجية، هندسية، إلكترونية، ودستورية. في دراسة للتنبؤ بمعامل هنري لمجموعة متنوعة من المركبات، تم استخدام أربعة أنواع من الواصفات الجزيئية التي تشمل الخصائص الطوبولوجية والإلكترونية والهندسية [64].

ج. اختيار النموذج الرياضي

يتم بعد ذلك استخدام النماذج الرياضية لربط الواصفات الجزيئية بالخصائص التي تم جمعها. تُستخدم عدة تقنيات لبناء هذه النماذج، مثل الانحدار الخطي المتعدد (MLR)، الشبكات العصبية الاصطناعية (ANN)، والغابات العشوائية (Random Forest). كما تم في دراسة أجريت لتنبؤ الكثافة في السوائل الأيونية، تم استخدام ثلاث نماذج هندسية مختلفةHF) ، PM7، و(B3LYP لتحسين الأداء التنبؤي للنموذج، وحقق نموذج HFأفضل النتائج [65].

د. التحقق من صحة النموذج

تُعد عملية التحقق من صحة النموذج خطوة أساسية للتأكد من أن النموذج قادر على التنبؤ بدقة بخصائص المركبات غير المعروفة. تتضمن طرق التحقق الشائعة التقسيم العشوائي لمجموعات التدريب والاختبار، والتحقق المتقاطع. كما قام تروبشا وآخرون، 2003 في دراسة حول بناء نموذج QSPR لتنبؤ بالخصائص الحرارية، قاموا باستخدام التحقق الخارجي لتقييم دقة النموذج، مما ساعدهم في تحسين دقة التنبؤ وجعل النموذج أكثر موثوقية [61].

ه. تحليل النتائج وتطبيق النموذج

بعد بناء النموذج والتحقق منه، يمكن استخدامه للتنبؤ بخصائص مركبات جديدة. تُستخدم هذه النماذج بشكل واسع في الصناعات الكيميائية والدوائية لتوفير الوقت والموارد التي قد تتطلبها التجارب المعملية.

3.2. الواصفات الجزيئية

3.2.1. تعريف وتحديد الواصفات الجزيئية

هي عبارة عن معلمة رقمية تصف بنية كيميائية معينة، مما يتيح إمكانية قياس بعض خصائص الجزيء. يمكن الحصول على هذه المؤشرات إما من خلال التجارب العملية أو من خلال حسابات مستندة إلى بنية

الجزيء. وتعد المؤشرات المحسوبة ذات أهمية خاصة في النمذجة الجزيئية، حيث تُمكِّن من التنبؤ بالخصائص أو الأنشطة دون الحاجة إلى تصنيع الجزيئات فعليًا، مما يسهم في تقليل الوقت والتكلفة التجريبية [62].

تلعب المؤشرات الجزيئية دورًا حاسمًا في دراسات العلاقات الكمية بين البنية والفعالية (QSAR) والعلاقات الكمية بين البنية والخواص .(QSPR) حيث تُستخدم كمتغيرات مستقلة في النماذج التنبؤية، بينما تكون الفعالية أو الخاصية المراد التنبؤ بها هي المتغير التابع. ومع ذلك، فإن دمج المؤشرات في نماذج QSAR/QSPRليس مهمة سهلة نظرًا للعدد الكبير من المؤشرات المتاحة وتنوعها. ففي السنوات الأخيرة، تم تطوير العديد من المؤشرات الجزيئية ذات التعقيدات المختلفة، إلا أنه لا توجد قاعدة عامة لاختيار المؤشرات الأكثر ملاءمة. يعتمد هذا الاختيار غالبًا على الخبرة والحدس العلمي للباحثين أو على التقاليد المعمول بها في هذا المجال [62]، [66].

3.2.2. اختيار الواصفات الجزيئية المناسبة للمركبات العضوية:

فيما يلي، سنستعرض أهم المؤشرات الجزيئية المستخدمة وكذلك تلك التي تم تطبيقها في مجمل أعمالنا. سنبدأ بالمؤشرات الأبسط التي تتطلب معرفة محدودة حول بنية الجزيء. ثم سنوضح كيف أتاحت التطورات في مجال النمذجة الجزيئية الوصول إلى البنية ثلاثية الأبعاد للجزيء وحساب المؤشرات استنادًا إلى هذه البنية. تاريخيًا، تم وضع مخططان رئيسيان لتصنيف المؤشرات الجزيئية: الأول يعتمد على أصلها (دستوري، طوبولوجي، هندسي، كمي، حراري ديناميكي...)، والثاني يعتمد على أبعادها (D0، D1، D2 أوD3).

A. المؤشرات الدستورية (Constitutional descriptors):

تُعد المؤشرات الدستورية من أكثر المؤشرات استخدامًا بسبب بساطتها الشديدة، ليس فقط من الناحية المفاهيمية، ولكن أيضًا من حيث سهولة الحساب. هذه المؤشرات تُستخدم بشكل رئيسي في تحليل مجموعات المساهمة، مثل طريقة بينسون، حيث يتم إدخال تأثير البدائل في شكل قيمة محددة في جداول. أبسط طريقة لتمثيل النظام الجزيئي هي أخذ التركيب الكيميائي فقط في الاعتبار، دون النظر في أي خصائص هندسية أو إلكترونية.

أمثلة على المؤشرات الدستورية تشمل:

- **الأعداد المطلقة والنسبية للذرات** (مثل الكربون، الهيدروجين، النيتروجين، ...).
- **الأعداد المطلقة والنسبية للمجموعات الوظيفية** (مثل NO2، COOH، OH، ...).
- **الأعداد المطلقة والنسبية للروابط** (مثل الروابط الأحادية، الثنائية، العطرية، ...).

- **الأعداد المطلقة والنسبية للحلقات** (سواء كانت عطرية أو غير عطرية، ...).
- **الكتلة الجزيئية**.

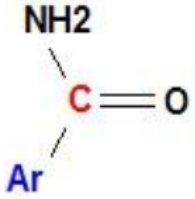
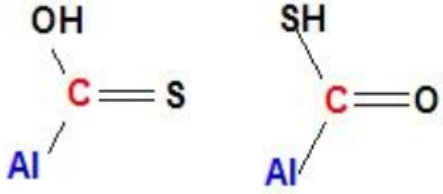
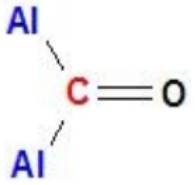
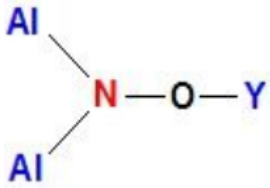
كما نذكر بعض الواصفات الدستورية التي تم إنشاؤها بواسطة برنامج Dragon في الجدول التالي:

الجدول 1.2: بعض الواصفات التي تنتمي مجموعة الواصفات الدستورية حسب برنامج Dragon.

الرقم	الاسم	الوصف
1	MW	الوزن الجزيئي
2	AMW	متوسط الوزن الجزيئي
3	Sv	مجموع أحجام فان دير فالز الذرية (مقاسة على ذرة الكربون)
4	Se	مجموع الكهرسلبية الذرية حسب مقياس ساندیرسون (مقاسة على ذرة الكربون)
5	Sp	مجموع الاستقطابات الذرية (مقاسة على ذرة الكربون)
6	Si	مجموع إمكانات التأين الأولى (مقاسة على ذرة الكربون)
7	Mv	متوسط حجم فان دير فالز الذري (مقاسة على ذرة الكربون)
8	Me	متوسط الكهرسلبية الذرية حسب مقياس ساندیرسون (مقاسة على ذرة الكربون)
9	Mp	متوسط الاستقطابية الذرية (مقاسة على ذرة الكربون)
10	Mi	متوسط إمكان التأين الأول (مقاسة على ذرة الكربون)
11	nAT	عدد الذرات
12	nSK	عدد الذرات غير الهيدروجينية
13	nBT	عدد الروابط
14	nBO	عدد الروابط غير الهيدروجينية
15	nBM	عدد الروابط المتعددة

B. واصفات المجموعات الوظيفية

واصفات المجموعات الوظيفية هي واصفات جزيئية بسيطة، يتم تعريفها بعدد المجموعات الوظيفية المحددة الموجودة في الجزيء. يتم حسابها بناءً على التركيب الجزيئي والاتصالات الذرية [67].

N°	Symbole	Structure	Définition
01	nArOCN		Nombre de cyanates (aromatique)
02	nArNCO		Nombre d'isocyanates (aromatique)
03	nArCONH2		Nombre d'amides primaires (aromatique)
04	nRCSOH		Nombre de thioacides (aliphatique)
05	nRCO		Nombre de cétones (aliphatique)
06	nRNHO		Nombre d'hydroxylamines (aliphatique) **Y** = H, Al ou Ar

الشكل 3.2: بعض المتغيرات الجزيئية (كتلة من المجموعات الوظيفية).

3.3. الأساليب الحسابية في QSPR:

تلعب الأساليب الحسابية دورًا محوريًا في بناء نماذج QSPR. تستخدم هذه الأساليب لربط الواصفات الجزيئية بالخواص الفيزيائية والكيميائية للمركبات، مما يسمح بتنبؤ سلوك المركبات العضوية بدون الحاجة إلى التجارب المعملية. يتم تطبيق مجموعة من التقنيات الرياضية والإحصائية لحل هذه المشكلة.

1.3.3. الانحدار الخطي المتعدد (Multiple Linear Regression, MLR)

يُعتبر MLR من أبسط وأشهر الطرق الحسابية المستخدمة في بناء نماذج .QSPR يعتمد على افتراض وجود علاقة خطية بين الواصفات الجزيئية والخواص المراد التنبؤ بها. يتم حساب الواصفات الجزيئية وتكوين معادلة خطية تتناسب مع البيانات المتاحة.

- مثال :في دراسة (Karelson et al., 1996)، تم استخدام MLR لتحليل العلاقة بين الواصفات الجزيئية والمعاملات الفيزيائية مثل الذوبانية في الماء، وتمكن النموذج من تحقيق نتائج دقيقة لعدد كبير من المركبات العضوية [68].

1. مفهوم MLR في QSPR:

الهدف من MLR هو تطوير نموذج يصف الخاصية الكيميائية أو الفيزيائية المراد التنبؤ بها على شكل معادلة خطية تجمع بين مجموعة من الواصفات الجزيئية. يتم ذلك من خلال إيجاد القيم المثلى للمعاملات المرتبطة بكل واصف بحيث تكون المعادلة الناتجة هي الأكثر دقة في التنبؤ.

المعادلة العامة للانحدار الخطي المتعدد تكون على الشكل التالي:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n$$

حيث:

- Y هو المتغير التابع (الخاصية المراد التنبؤ بها).
- X، X_1، X_2، ...، X_n هي الواصفات الجزيئية (المتغيرات المستقلة).
- b، b_0، b_1، ...، b_n هي المعاملات التي تحدد الأهمية النسبية لكل واصف.

2. استخدام MLR في نماذج QSPR

تُستخدم MLR بشكل واسع في نماذج QSPR بسبب قدرتها على التنبؤ بخصائص متنوعة مثل الذوبانية، النشاط البيولوجي، والاستقرار الحراري. يتيح MLR بناء نماذج بسيطة وسهلة الفهم، مما يساعد في تفسير العلاقة بين البنية الجزيئية والخصائص المراد تنبؤها.

- **مثال:** في دراسة (Karelson et al., 1996)، تم استخدام MLR لتقدير الخواص الفيزيائية مثل الذوبانية ومعامل تقسيم الأوكتانول/الماء .(LogP) أظهرت الدراسة أن MLR كان قادرًا على بناء نموذج خطي دقيق يعتمد على مجموعة من الواصفات الطوبولوجية والكمومية، مما يعزز من موثوقية التنبؤات [53].

3. مزايا وعيوب MLR:

أولا المزايا:

- البساطة :يُعد MLR من أبسط النماذج الرياضية التي يسهل تطبيقها وفهمها.
- السرعة والكفاءة: يمكن بناء النماذج بسرعة باستخدام بيانات تدريب كبيرة نسبيًا.
- التفسير الواضح: المعادلة الخطية الناتجة تتيح تفسير واضح لتأثير كل واصف على الخاصية المستهدفة.

ثانيا العيوب:

- **الافتراض الخطي** :يفترض MLR أن العلاقة بين الواصفات والخاصية المستهدفة هي علاقة خطية، وهذا قد لا يكون دقيقًا في الحالات التي تتطلب نماذج غير خطية.
- **الاعتماد على الاستقلالية** :يجب أن تكون المتغيرات المستقلة (الواصفات الجزيئية) غير مرتبطة بشكل كبير ببعضها البعض، وإلا فإن أداء النموذج قد يتأثر سلبًا.

4. التطبيقات العملية لـ MLR في الكيمياء

تُستخدم MLR في مجموعة متنوعة من التطبيقات في الكيمياء، خاصة في مجال تصميم الأدوية وتطوير المواد الكيميائية. على سبيل المثال، يتم استخدام MLR لتنبؤ خواص الجزيئات مثل:

- تنبؤ النشاط البيولوجي :حيث يُستخدم MLR لنمذجة النشاط الدوائي للمركبات العضوية.
- تقدير الخصائص الحرارية :مثل الذوبانية أو نقطة الغليان.

يُعتبر الانحدار الخطي المتعدد (MLR) من أكثر الأساليب شيوعًا في بناء نماذج QSPR نظرًا لبساطته وقدرته على تقديم تنبؤات دقيقة. على الرغم من القيود التي تفرضها العلاقة الخطية، إلا أن MLR يظل خيارًا فعالًا خاصة في الحالات التي تكون فيها العلاقات بين المتغيرات خطية أو شبه خطية.

2.3.3. الانحدار غير الخطي (MnLR)

الانحدار غير الخطي هو شكل من أشكال تحليل الانحدار حيث يتم نمذجة البيانات الرصدية باستخدام دالة غير خطية تجمع بين معاملات النموذج وتعتمد على متغير مستقل أو أكثر. يُعتبر هذا النهج مفيدًا بشكل خاص عندما تكون العلاقة بين المتغيرات التابعة والمستقلة معقدة ولا يمكن تمثيلها بشكل كافٍ من خلال نموذج خطي. فيما يلي نذكر بعض الأمثلة على الانحدار غير الخطي:

❖ **الانحدار متعدد الحدود:(Polynomial Regression)**

في هذا النوع من الانحدار، تكون العلاقة بين المتغيرات مستقلة (X) والتابع (Y) تمثل معادلة من الدرجة الثانية أو أعلى. على سبيل المثال: $Y=b_0+b_1X+b^2_2X+\cdots+b^n_nX$ هذا النوع من الانحدار يمكن أن يلتقط أنماطًا منحنية في البيانات.

❖ **الانحدار اللوغاريتمي:(Logarithmic Regression)**

العلاقة بين المتغير المستقل والتابع تكون على شكل دالة لوغاريتمية. على سبيل المثال:

$Y=a+b*\ln(X)$

يستخدم هذا النوع من الانحدار عندما يتباطأ معدل التغير في المتغير التابع مع زيادة المتغير المستقل.

❖ **الانحدار الأسي:(Exponential Regression)**

العلاقة تكون على شكل دالة أسية، ويستخدم لتمثيل البيانات التي تنمو أو تتناقص بمعدل أسي. الصيغة العامة هي: $Y=ae^{bX}$.

3.3.3. تطبيقات الانحدار غير الخطي:

كذلك هنا يمثل الانحدار غير الخطي أداة قوية للتنبؤ بالعديد من الخصائص الفيزيائية الكيميائية والديناميكية الحرارية. فيما يلي بعض الدراسات ذات الصلة التي استخدمت تقنيات الانحدار غير الخطي في هذا السياق:

1. **التحديد المتزامن للثوابت التوازنية والدوال الديناميكية الحرارية:**

تطبق هذه الدراسة الانحدار غير الخطي لتحديد المعلمات الديناميكية الحرارية لتوازنات الخطوة الواحدة من سعات الاسترخاء الكيميائي. تركز الدراسة على التقييم المتزامن للثوابت التوازنية (K) والتغيرات في الإنتالبي (ΔH) لتكوين معقد بنسبة 1:1، وتمدد النهج ليشمل التفاعلات السريعة المرتبطة بتأخر الإيزوميرات البطيء [69].

2. **طرق الانحدار غير الخطي للقياسات المتزامنة للخصائص:**

تستعرض هذه الورقة إجراءً عامًا لتقدير الخصائص الحرارية من البيانات التجريبية باستخدام الانحدار غير الخطي. تناقش إزالة القيود المحددة التي تفرض غالبًا على تجارب قياس الخصائص وتقدم معيار تقارب جديد للحسابات التكرارية [70].

3. **جزيئات الكروموفور غير الخطية من المرتبة الثانية:**

تطور الدراسة نموذج علاقة بنية-خاصية كمية (QSPR) باستخدام الانحدار غير الخطي للتنبؤ بالثبات الحراري لجزيئات الكروموفور غير الخطية. يتفوق النموذج غير الخطي الذي تم تطويره باستخدام آلات دعم المتجهات (SVM) على النموذج الخطي من حيث دقة التنبؤ [71].

4. **النمذجة الحاسوبية غير الخطية للبيانات الكيميائية والبيوكيميائية:**

يغطي هذا الكتاب الشامل تطبيقات مختلفة للانحدار غير الخطي في نمذجة البيانات الكيميائية والبيوكيميائية. يشمل أمثلة من عمليات المعايرة، التوازنات الجزيئية الكبيرة، والاسترخاء بالرنين المغناطيسي النووي، ويوفر دليلًا مفصلاً حول بناء وتحسين النماذج غير الخطية [72].

5. **التقييم المتزامن والارتباط للبيانات الديناميكية الحرارية:**

تقدم الورقة استراتيجية موجهة بالحاسوب للتقييم المتزامن والارتباط للبيانات الديناميكية الحرارية باستخدام الانحدار غير الخطي. تصف معادلة تجريبية عامة للتبعية الحرارية وتطبق الطريقة على أنظمة مثل Fe-C-H-O وN-H-O [73].

تسلط هذه الدراسات مجتمعة الضوء على تنوع وفعالية الانحدار غير الخطي في التنبؤ بمجموعة واسعة من الخصائص الفيزيائية الكيميائية والديناميكية الحرارية.

طرق التحقق من صحة نماذج الانحدار الخطي وغير الخطي:

طرق التحقق من صحة نماذج الانحدار الخطي وغير الخطي تتنوع بين عدة تقنيات، ومنها:

1. **تقنية تقسيم البيانات (Data Splitting):**

حيث يتم تقسيم البيانات إلى مجموعتي تدريب واختبار، حيث يتم تدريب النموذج على مجموعة التدريب وتقييم أدائه باستخدام مجموعة الاختبار [74].

2. **التحقق المتقاطع (Cross-validation):**

يتم تقسيم البيانات إلى مجموعات فرعية واستخدام كل مجموعة كبيانات اختبار مرة واحدة، بينما يتم تدريب النموذج على باقي المجموعات. هذه الطريقة تعتبر من أكثر الطرق شيوعاً لاختبار أداء النماذج [75].

3. **تحسين المعلمات باستخدام التحقق المتقاطع (Parameterized Cross-validation):**

تم اقتراح هذه الطريقة للتحقق من أداء النماذج غير الخطية باستخدام معايير إضافية، مثل تقدير المخاطر باستخدام نظرية أبعاد VC [76].

4. **مؤشرات الأداء التنبؤية (Predictive Performance Criteria)** [77]**:**

وتشمل المعايير التالية:

- **معامل التحديد (R^2 The correlation coefficient):**

معامل التحديد R^2 هو مقياس إحصائي يستخدم لتقييم جودة نموذج الانحدار وتحديد مدى تفسير المتغير المستقل أو المتغيرات المستقلة للمتغير التابع. يتم حسابه باستخدام الصيغة:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y - \bar{y}_i)^2} \quad \text{..... (2.5) .}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} w_i y_i \quad \text{..... (2.6) .}$$

حيث:

- y_i هي القيم الفعلية للمتغير التابع.
- $\hat{y}_i$ هي القيم المتوقعة من النموذج.
- $\bar{y}$ هي متوسط القيم الفعلية لـ y.

قيمة R^2 تتراوح بين 0 و1، حيث تشير القيمة 1 إلى أن النموذج يفسر 100% من التباين في البيانات، بينما تشير القيمة 0 إلى أن النموذج لا يفسر أي تباين.

- **MCE (Mean squares of errors):**

متوسط مربعات الأخطاء هو مقياس يُستخدم في الإحصاء لتقييم جودة نموذج التنبؤ. بحيث يقيس الفرق بين القيم المتوقعة بواسطة النموذج والقيم الفعلية المرصودة. فيما يلي الصيغة العامة:

$$\text{MCE} = \frac{1}{n} \sum_{i=1}^{n} (\text{y}_i - \hat{\text{y}_i})^2$$

حيث:

n: هو عدد النقاط في البيانات.

y_i : هو القيمة الفعلية.

$\hat{y_i}$: هو القيمة المتوقعة.

كلما كان متوسط مربعات الأخطاء أقل، كان النموذج أفضل في التنبؤ بالقيم الفعلية.

- **(Square root of MCE) RMCE:**

جذر متوسط مربعات الأخطاء او (RMSE) هو مصطلح يُستخدم لقياس دقة النماذج التنبؤية. يُعبر عن مدى تباين الأخطاء في التنبؤ حول القيم الحقيقية. كما تكتب الصيغة الرياضية لجذر متوسط مربعات الأخطاء على النحو التالي:

$$RMSE = \sqrt{\text{MCE}}$$

- **(Mean Absolute Percentage Error) MAPE:**

نسبة الخطأ المطلق التي تعني **"متوسط نسبة الخطأ المطلق"**، هي مقياس يُستخدم لتقييم دقة النماذج التنبؤية. يتم حساب MAPE عن طريق قياس الخطأ في التنبؤ كنسبة مئوية من القيم الحقيقية.

الصيغة لحساب MAPE هي:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y_i}}{y_i}\right| \times 100$$

n: هو عدد النقاط في البيانات.

y_i : هو القيمة الفعلية.

$\hat{y_i}$: هو القيمة المتوقعة.

يقيس MAPE الخطأ المتوسط كنسبة مئوية من القيم الفعلية. كلما كان MAPE أقل، كانت دقة التنبؤات أفضل.

4.3.3. تقنيات تحسين اختيار الواصفات

أولا الخصائص الفيزيائية والكيميائية:

تعتبر الخصائص الفيزيائية والكيميائية من العوامل الأساسية في اختيار الواصفات الجزيئية المناسبة. تتضمن هذه الخصائص التأثيرات الإلكترونية والتأثيرات الفراغية التي يمكن أن تؤثر على تفاعلية وانتقائية

الجزيئات. على سبيل المثال، يمكن استخدام الواصفات الجزيئية الكلاسيكية مثل معلمات Hammett وTaft وCharton لفحص التأثيرات الإلكترونية والفراغية بشكل مستقل، ولكنها قد تكون محدودة في معالجة التغيرات المتزامنة وغير الإضافية لأكثر من خاصية جزيئية واحدة [78].

بالإضافة إلى ذلك، يمكن استخدام الحسابات الكيميائية الكمومية لتوفير وصف دقيق ومفصل للتأثيرات الإلكترونية. توفر هذه الحسابات واصفات جزيئية يمكن أن تعبر عن جميع الخصائص الإلكترونية والهندسية للجزيئات وتفاعلاتها [79]. كما أن التقدم في الأجهزة الحاسوبية وتطوير الخوارزميات الفعالة قد ساعد في تطوير حسابات كيميائية كمومية شبه تجريبية توفر كميات جزيئية واقعية في إطار زمني قصير نسبيًا [79].

ثانيا التفاعلات الجزيئية:

تلعب التفاعلات الجزيئية دورًا حاسمًا في اختيار الواصفات الجزيئية المناسبة. يمكن استخدام الواصفات التي تعتمد على التفاعلات الثنائية والثلاثية بين الذرات لتوقع خصائص الجزيئات بدقة باستخدام تقنيات التعلم الآلي [80]. هذه الواصفات تكون غير متغيرة بالنسبة للترجمة والدوران وفهرسة الذرات، مما يجعلها مفيدة في التنبؤ بخصائص الجزيئات العضوية الصغيرة [80].

كما أن استخدام التمثيلات الجزيئية المستندة إلى الرسوم البيانية العصبية يمكن أن يتفوق على الطرق التقليدية في التنبؤ بخصائص الجزيئات، خاصة عند توفر مجموعات بيانات كبيرة [81]، [82]. هذه التمثيلات يمكن أن تتعلم العلاقة المعقدة بين البنية الجزيئية والهدف التنبؤي، مما يجعلها أدوات قوية في التنبؤ بالخصائص الكيميائية والفيزيائية [81]، [82].

الخاتمة:

في الختام، قدمنا في هذا العمل نظرة شاملة على مفهوم QSPR وآليات بناء النماذج المرتبطة به .لقد تبين أن QSPR أداة قوية للتنبؤ بالخواص الكيميائية للمركبات العضوية، مما يساهم في تسريع عملية اكتشاف المركبات الجديدة وتصميمها .ومع ذلك، لا تزال هناك تحديات تواجه هذا المجال، مثل اختيار أفضل مجموعة من الواصفات الجزيئية وتطوير أساليب أكثر دقة لتقييم أداء النماذج .في المستقبل، من المتوقع أن يشهد مجال QSPRتطورات كبيرة بفضل التقدم في مجال الحوسبة وتوفر قواعد بيانات ضخمة من البيانات الجزيئية . يمكن أن تساهم هذه التطورات في بناء نماذج أكثر دقة وشمولية، مما يفتح آفاقًا جديدة في مجالات مثل تصميم الأدوية، المواد المتقدمة، والبيئة.

الفصـــل الثالـــث:

منـهجيـــة العمـــل

الفصل الثالث: منهجية العمل

يهدف هذا الفصل إلى شرح المنهجية العملية المتبعة في تطوير نموذج تنبؤي يعتمد على العلاقة الكمية بين التركيب الكيميائي والخصائص الفيزيائية (QSPR) للمركبات العضوية. تبدأ هذه المنهجية بتجميع البيانات المتعلقة بالمركبات الكيميائية، حيث تم تحديد مصادر البيانات المستخدمة واختيار مجموعة متنوعة من المركبات العضوية ذات الأهمية لتحليل الخصائص الفيزيائية والكيميائية. يعكس هذا الجزء أهمية الحصول على بيانات موثوقة وشاملة، حيث تعد خطوة تجميع البيانات أساساً في بناء النموذج.

ينتقل الفصل بعد ذلك إلى عملية حساب الواصفات الجزيئية، وهي خطوة حيوية تهدف إلى تحويل الهياكل الكيميائية إلى متغيرات كمية يمكن استخدامها في بناء النماذج التنبؤية. يتم في هذا الجزء توضيح الأدوات والبرامج التي تم استخدامها لحساب هذه الواصفات، بالإضافة إلى إجراء تحليل للواصفات من أجل تحديد المتغيرات الأكثر تأثيراً على الخصائص المستهدفة.

يتناول الفصل أيضاً عملية بناء النموذج التنبؤي، بدءاً من اختيار النموذج الأنسب لتحليل QSPR بناءً على خصائص البيانات المتاحة. يشمل هذا الجزء خطوات تدريب النموذج واختباره باستخدام طرق إحصائية مختلفة لضمان دقته وكفاءته في التنبؤ بالخصائص الفيزيائية للمركبات العضوية (4.3). يتم كذلك تحليل النتائج لتقييم أداء النموذج ومقارنته مع القيم التجريبية.

في القسم الأخير من هذا الفصل، يتم التركيز على تحسين النموذج من خلال تطبيق خوارزميات متقدمة تهدف إلى تعزيز دقة التنبؤات وتقليل الأخطاء المحتملة. يتم تقييم أداء النموذج المحسن باستخدام مجموعة من المقاييس الإحصائية للتأكد من موثوقية النتائج وقدرة النموذج على تقديم توقعات دقيقة ومفيدة. يعكس هذا القسم أهمية تحسين وتقييم النماذج بشكل مستمر لضمان تطوير نموذج تنبؤي فعال وقابل للتطبيق في الدراسات الكيميائية المختلفة.

1.3 إنشاء قاعدة البيانات وتحديد الأوصاف الجزيئية

قاعدة البيانات هي الأداة التي تساعد في تنظيم المعلومات وتحليلها بفعالية. إنشاء قاعدة بيانات حول $\Delta H_f^\circ{}_{gaz}$ يتطلب جمع وتنظيم البيانات حول المواد الكيميائية المختلفة وخصائصها المتعلقة بالاحتراق.

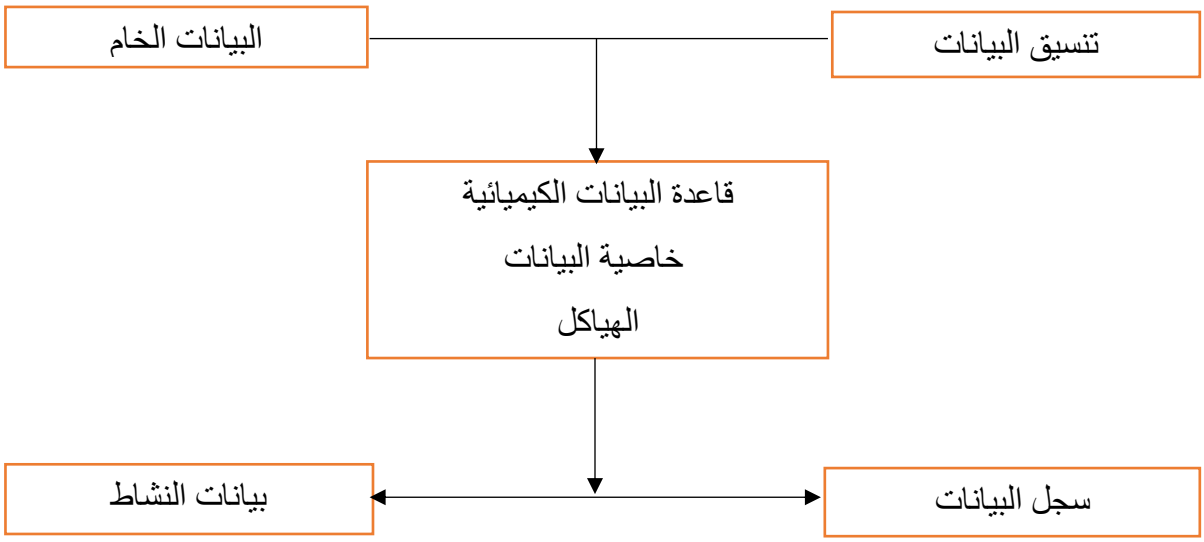
2.3 منهجية إعداد قاعدة البيانات

1.2.3 تصميم قاعدة البيانات

تُعد الحاجة إلى التحليل السريع وإدارة الكميات المتزايدة من المواد الكيميائية، بالإضافة إلى أرشفتها، من الأسس الرئيسية لتطوير البحث والتطوير الكيميائي. لتحقيق هذا، يعتبر تصميم وتنفيذ أنواع مختلفة من قواعد البيانات أمرًا بالغ الأهمية.

في هذا السياق، يُعتبر مفهوم "قاعدة البيانات العلائقية" أمرًا أساسيًا، حيث يتم ربط المكونات وأنواع المعلومات المختلفة وتوفيرها للمستخدمين ضمن سياق محدد. على سبيل المثال، في مجال الكيمياء، يمكن أن توفر قاعدة بيانات علائقية بسيطة تخزينًا للهياكل الكيميائية، وتوافر المواد، وبيانات تحليلية. كما يمكن أن تتضمن قاعدة بيانات أكثر تقدمًا بيانات متعلقة بالنشاط البيولوجي للمركبات.

يعتبر استخدام قواعد البيانات الكيميائية العلائقية من الجوانب الأساسية التي تُعزز العديد من تطبيقات المعلوماتية الكيميائية. بالإضافة إلى ذلك، توفر قواعد البيانات العلائقية منصة ملائمة لتفاعل العلماء المختصين في الكيمياء النظرية والتجريبية، مما يسهم في تحقيق تقدم كبير في هذا المجال.



الشكل 1.3: تمثيل تخطيطي لهيكل قاعدة البيانات العلائقية.

تعتمد قواعد البيانات العلائقية على تنظيم البيانات ضمن جداول مترابطة لتخزين وتنظيم المعلومات بطريقة منهجية. تساهم لغة الاستعلام الهيكلية (SQL) في تسهيل التفاعل مع هذه الجداول واسترجاع البيانات المطلوبة. عادةً ما يتم تقديم واجهة مستخدم رسومية لتبسيط عملية التفاعل مع المستخدم، حيث يتم إخفاء تعقيدات SQL عن المستخدم النهائي .تتميز قواعد البيانات العلائقية بمرونة عالية تسمح بدمج أنواع مختلفة من البيانات وهياكل البيانات، مما يجعلها مناسبة لتلبية احتياجات تطبيقات متنوعة. في مجال الكيمياء، تلعب قواعد البيانات العلائقية دورًا حاسمًا في دعم الأبحاث وتوثيق النتائج، حيث توفر إطارًا موحدًا لتخزين وتنظيم البيانات المتعلقة بالمركبات الكيميائية والتفاعلات الكيميائية .يوضح الشكل 1.3 العناصر الأساسية التي يجب أن تتضمنها قاعدة بيانات كيميائية فعالة [83] [84] [85].

2.2.3 إعداد قاعدة البيانات :

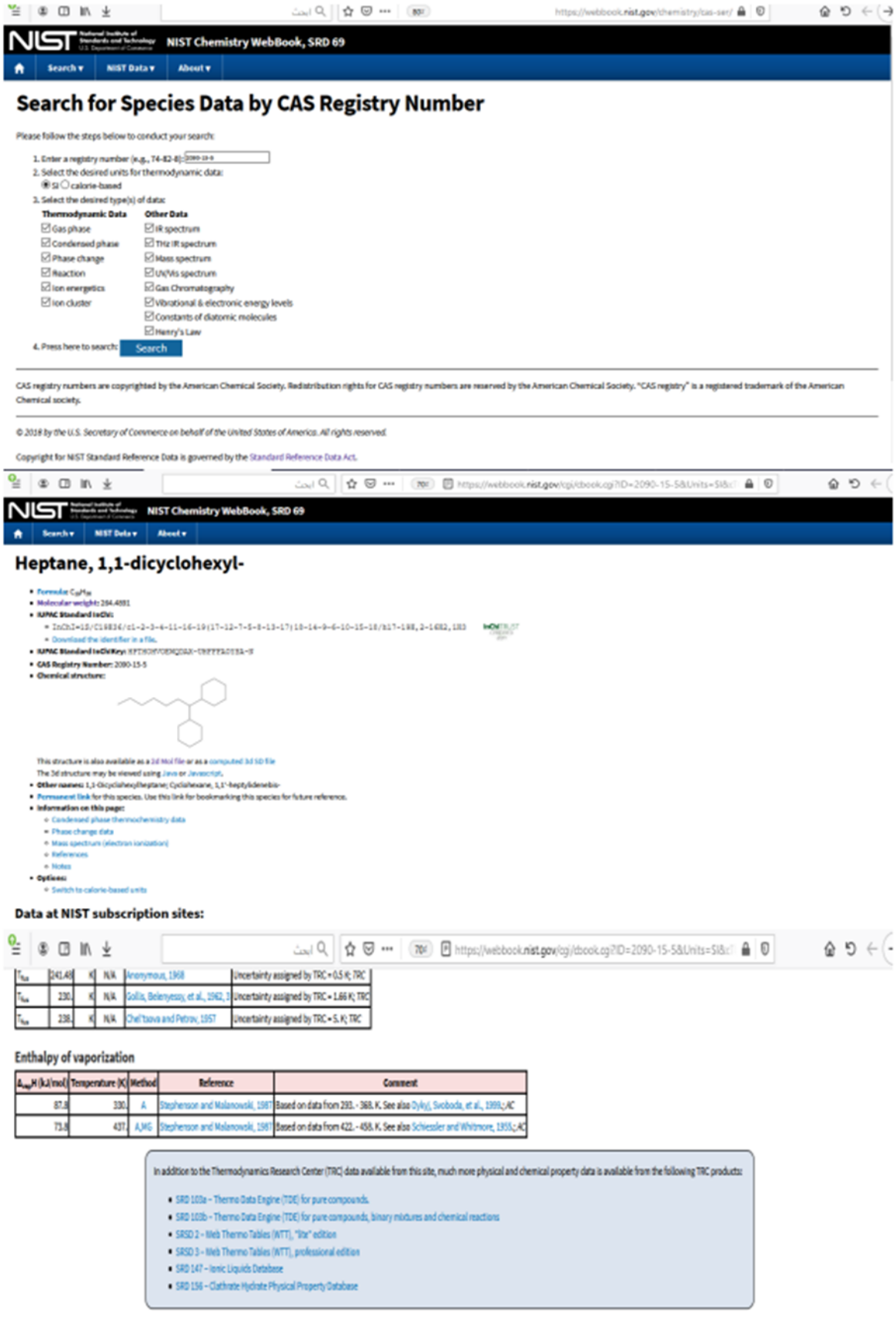
قمنا في هذه الدراسة بمنهجية بحثية دقيقة بدأت بجمع قاعدة بيانات أولية شاملة. اعتمدنا في المرحلة الأولى على رقم CAS (Chemical Abstracts Service Registry Number) كمعرف فريد لكل مركب، مما سمح لنا بتحديد هوية كل مركب بشكل دقيق. ثم تم الحصول على القيم التجريبية للخصائص الفيزيائية والكيميائية لـ 339 مركباً كيميائياً متنوعة من قاعدة بيانات المعايير القومية للتكنولوجيا (NIST) [86]. في المرحلة الثانية، تم توسيع البحث ليشمل قيم المحتوى الحراري للاحتراق لهذه المركبات، والتي تم الحصول عليها أيضاً من قاعدة بيانات NIST (الشكل 2.3). علاوة على ذلك، تم تمثيل الهياكل الجزيئية لجميع المركبات بصيغة الملف الجزيئي(Mol2) لتسهيل التحليل الإضافي باستخدام أدوات المحاكاة الجزيئية.

1.2.3 طرق تحليل البيانات

إعداد نماذج QSPR ليس بالأمر السهل. إذ تكمن الصعوبة الأولى في التباين بين المقاييس المرتبطة بالبيانات المراد ربطها. فالهيكل يتم دراسته على المستوى الجزيئي بينما تكون الخصائص المُراد التنبؤ بها على نطاق مجهري. بالإضافة إلى ذلك، يجب مراعاة مشكلات عدم اليقين على مستوى الهياكل الجزيئية (مثل مستوى الحساب المرتبط) والبيانات التجريبية (مثل بروتوكولات القياس). تتجلى إحدى التحديات المهمة في التعامل مع كميات كبيرة من البيانات، حيث قد يتطلب الأمر تحليل عدد كبير من الواصفات والجزيئات، دون وجود قاعدة صارمة لاختيار الأكثر أهمية من بينها. في الواقع، يتطلب الأمر استخدام العديد من الأدوات للعثور على الطريقة الأنسب لبناء نموذج موثوق من البيانات المتاحة. بناءً على الحالة، هناك طرق مختلفة يمكن اتباعها، ويجب اختيار الطريقة التي تتيح أفضل توصيف للنظام. الطرق المختلفة المعروضة أدناه هي تلك المستخدمة أثناء الدراسة لتطوير النماذج، سواء كانت ذات عدد أكبر أو أقل من المعلمات، خطية أو غير خطية، قابلة للتفسير أو غير قابلة للتفسير. من الضروري اختيار الإعدادات الأكثر صلة والتحقق من صحة هذه النماذج (داخليًا وخارجيًا) وتحديد مجالات تطبيقها [87] [88].

2.2.3 تحليل الواصفات الجزيئية:

تم حساب الواصفات الجزيئية باستخدام برامج متخصصة، وتحليلها لاختيار المتغيرات الأكثر تأثيراً على المحتوى الحراري للتكوين. تم استخدام أدوات مثل Dragon وHyperChem لتحديد القيم العددية للواصفات الجزيئية، والتي تُستخدم كمدخلات للشبكة العصبية.



الشكل 2.3: صورة توضح طريقة البحث عن بيانات المركبات باستخدام رقم سجل CAS على موقع الويب الخاص بـ NIST.

3.3.3 منهجية الانحدار الخطي المتعدد باستخدام XLSTAT:

يُعد الانحدار الخطي المتعدد (Multiple Linear Regression - MLR) أحد الأساليب الإحصائية المستخدمة على نطاق واسع لتحليل العلاقات بين متغير تابع وعدة متغيرات مستقلة. في هذه الدراسة، تم استخدام برنامج XLSTAT لبناء نموذج التنبؤ الذي يعتمد على الواصفات الجزيئية للمركبات العضوية لتقدير قيم المحتوى الحراري للتكوين. وفيما يلي المنهجية العملية المتبعة لتطبيق الانحدار الخطي المتعدد باستخدام برنامجXLSTAT . [89] [90].

1.3.3.3 تحديد مجموعة المركبات العضوية:

تم جمع البيانات اللازمة للدراسة من قاعدة بيانات NIST، وتم اختيار 339 مركبًا عضويًا ذات خصائص جزيئية تمثل القيم التجريبية للمحتوى الحراري للتكوين. تم تقسيم هذه المجموعة إلى مجموعتين:

- مجموعة التدريب (Training Set): تشمل 266 مركبًا، وتستخدم لتدريب النموذج.
- مجموعة التحقق (Validation Set): تضم 72 مركبًا، وتستخدم لاختبار قدرة النموذج على التعميم على بيانات جديدة.

أ. تهيئة البيانات:

قبل بناء النموذج، تم تنفيذ خطوات أساسية لمعالجة البيانات شملت:

- التنظيم والتحقق من البيانات للتأكد من عدم وجود بيانات مفقودة أو قيم متطرفة.
- تطبيع المتغيرات (Normalization) لضمان أن جميع الواصفات تتبع مقياسًا موحدًا لتجنب التأثير غير المتكافئ لمتغيرات محددة.
- التحقق من الخطية بين المتغيرات المستقلة والمتغير التابع لضمان مناسبة النموذج.

ب. تنفيذ الانحدار الخطي المتعدد باستخدام XLSTAT:

يتم تنفيذ الانحدار الخطي المتعدد باستخدام XLSTAT كما هو موضح في الشكل 3.3.

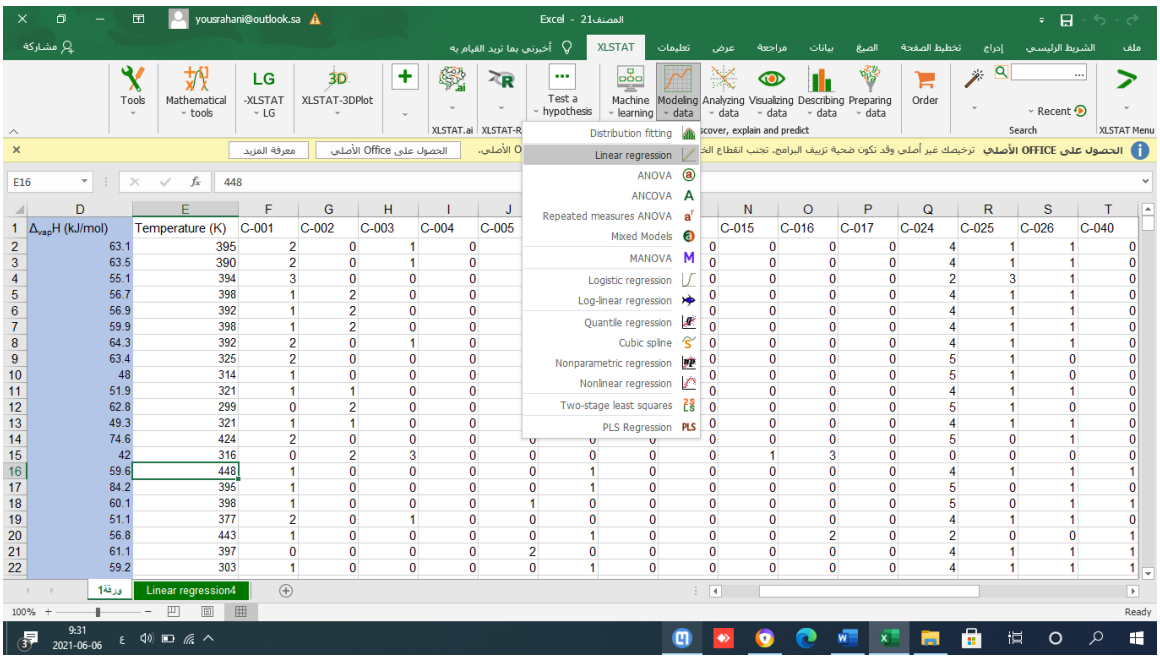
❖ **تحليل الانحدار في XLSTAT:**

تم إجراء تحليل الانحدار الخطي المتعدد باستخدام الأداة الخاصة بالانحدار في برنامج XLSTAT. يتطلب ذلك تحديد المتغير التابع (المحتوى الحراري للتكوين) والمتغيرات المستقلة (الواصفات الجزيئية) التي سيتم استخدامها في النموذج.

❖ إعدادات النموذج:

في إطار XLSTAT ، تم ضبط الإعدادات التالية:

- **طريقة التقدير**: تم استخدام طريقة المربعات الصغرى (Ordinary Least Squares - OLS) لتقدير معاملات الانحدار. تقوم هذه الطريقة بتقليل مجموع مربعات الفروق بين القيم التنبؤية والقيم التجريبية [91].
- **اختيار المتغيرات**: تم استخدام تقنيات مثل التحليل التدريجي (Stepwise Regression) لتحديد المتغيرات الأكثر تأثيراً في النموذج، مع استبعاد المتغيرات التي لا تسهم بشكل كبير في تحسين دقة التنبؤ [91].



الشكل 3.3: صورة توضح نافذة برنامج XLSTAT (اختيار الانحدار الخطي المتعدد).

❖ افتراضات النموذج:

تم التحقق من الافتراضات الأساسية للانحدار الخطي المتعدد قبل بناء النموذج [92]:

- الاستقلالية بين المتغيرات المستقلة للتأكد من عدم وجود تداخل (Multicollinearity) قد يؤثر على دقة تقدير المعاملات.
- الخطية بين المتغيرات المستقلة والمتغير التابع، بحيث يتم تحليل وتحديد العلاقات الخطية بشكل دقيق.
- التوزيع الطبيعي للبواقي (Residuals) للتحقق من توافق توزيع البواقي مع التوزيع الطبيعي.

❖ **تدريب النموذج:**

أولا: تدريب النموذج:

بعد تحديد الواصفات الجزيئية المناسبة، تم تدريب نموذج الانحدار الخطي المتعدد باستخدام بيانات مجموعة التدريب. تم تقدير معاملات الانحدار التي تصف العلاقة بين كل واصف جزيئي والمحتوى الحراري للتكوين.

ثانيا: تحليل الأداء:

تم تحليل أداء النموذج من خلال عدة مؤشرات إحصائية [93]:

- **معامل التحديد** (R^2): يُستخدم لتحديد نسبة التباين في المتغير التابع التي يمكن تفسيرها بواسطة المتغيرات المستقلة. كلما اقتربت قيمة R^2 من 1، كان النموذج أكثر دقة في تفسير التباين.
- **معامل التحديد المعدّل** (Adjusted R^2): يأخذ في الاعتبار عدد المتغيرات المستقلة ويُعدل قيمة R^2 للتغلب على احتمال تضخمها بسبب زيادة عدد المتغيرات.
- **الخطأ التربيعي المتوسط** (Mean Squared Error - MSE): يقيس متوسط مربع الفروق بين القيم التجريبية والقيم المتوقعة.

ثالثا: اختبار أهمية المعاملات:

تم اختبار أهمية المعاملات (Coefficients) باستخدام **اختبار t** لتحديد مدى تأثير كل واصف جزيئي على المحتوى الحراري. تم تحديد المتغيرات التي تسهم بشكل كبير في تحسين دقة النموذج بناءً على قيم p-Value [94].

رابعا: اختبار النموذج باستخدام مجموعة التحقق

أ. تطبيق النموذج على مجموعة التحقق:

بعد تدريب النموذج، تم اختبار أدائه باستخدام مجموعة التحقق. تم استخدام المعاملات المقدرة لتوقع قيم المحتوى الحراري للمركبات في هذه المجموعة ومقارنتها مع القيم التجريبية.

ب. تحليل النتائج:

تم تقييم أداء النموذج باستخدام مجموعة التحقق عبر المؤشرات الإحصائية نفسها المستخدمة مع مجموعة التدريب، مثل **R^2** و**MSE**. كذلك، تم تحليل الفروق بين القيم المتوقعة والقيم الحقيقية للتأكد من دقة النموذج وملاءمته للتعميم على بيانات جديدة [95].

خامسا: تحسين النموذج

a. تحليل بواقي النموذج:

تم تحليل البواقي (Residuals) للتأكد من توزيعها العشوائي حول الصفر وعدم وجود نمط محدد، مما يشير إلى عدم وجود انحياز في النموذج. في حالة اكتشاف أي انحرافات، تم إجراء تعديلات على النموذج لتحسين أدائه [95].

b. استخدام تقنيات تحسين الأداء:

تم تحسين النموذج من خلال تجربة استراتيجيات مثل اختيار النماذج المتعددة (Model Selection)لتحديد النموذج الأنسب. كذلك، تم استخدام تقنيات مثل التقليص (Regularization) لتجنب الإفراط في التكيف(Overfitting) وتحسين القدرة التنبؤية [90].

2.3.3 منهجية الانحدار غير الخطي باستخدام الشبكات العصبية (NeuroOne):

تُعد الشبكات العصبية الاصطناعية (ANN) أداة قوية ومفيدة في تمثيل العلاقات غير الخطية المعقدة بين المتغيرات [96]، مما يجعلها خيارًا مثاليًا لتطبيق نماذج الانحدار غير الخطي. في هذه الدراسة، تم استخدام برنامج NeuroOne5 لتصميم وتدريب الشبكات العصبية بهدف بناء نموذج QSPR للتنبؤ بالخصائص الفيزيائية والكيميائية للمركبات العضوية بناءً على الواصفات الجزيئية [88] [96]. فيما يلي تفاصيل المنهجية المتبعة في هذا الإطار:

1.2.3.3 تحديد مجموعة المركبات العضوية:

تم جمع بيانات 339 مركباً عضوياً تشمل خصائصها الجزيئية وقيمها التجريبية للمحتوى الحراري للتكوين. تم تقسيم هذه البيانات إلى مجموعتين:

- **مجموعة التدريب** (Training Set)**:** تشمل 266 مركبًا وتستخدم لتدريب الشبكة العصبية على تعلم العلاقات بين الواصفات الجزيئية والقيم المستهدفة.
- **مجموعة التحقق** (Validation Set)**:** تضم 72 مركبًا وتستخدم لتقييم أداء النموذج بعد تدريبه.

2.2.3.3 تصميم الشبكة العصبية باستخدام برنامج NeuroOne:

a) هيكل الشبكة العصبية:

تم تصميم الشبكة العصبية في برنامج **NeuroOne** بحيث تتكون من:

- **طبقة إدخال**: تتضمن العقد التي تمثل الواصفات الجزيئية (المتغيرات المستقلة).
- **طبقة مخفية** (أو عدة طبقات): تقوم بمعالجة البيانات واستخراج الأنماط غير الخطية التي تربط بين المتغيرات المستقلة والمخرجات.
- **طبقة الإخراج** :تمثل قيمة المحتوى الحراري المستهدف (المتغير التابع).

يتم تحديد عدد الطبقات المخفية وعدد الخلايا العصبية في كل طبقة بناءً على تجارب متكررة لاختيار الهيكل الأمثل للشبكة، مع الأخذ في الاعتبار حجم البيانات وتعقيد العلاقات المراد تمثيلها.

b) إعدادات التدريب:

تم ضبط إعدادات الشبكة العصبية في **NeuroOne** باستخدام مجموعة من المعلمات الأساسية:

- **معدل التعلم** (Learning Rate): تم تحديده لتحديد مدى سرعة تعديل الأوزان داخل الشبكة خلال عملية التدريب.
- **عدد العُقَد** في الطبقات المخفية: تم ضبطه بناءً على تجارب متعددة لضمان تحقيق أداء متوازن دون الإفراط في التكيف (Overfitting) أو التقليل من التعلم (Underfitting). بحيث استعملنا عقدتين (Deux Neurones) مخفيتين.
- **عدد الدورات** (Epochs): عدد مرات تمرير بيانات التدريب عبر الشبكة لضبط الأوزان تدريجيًا.

3.2.3.3 عملية التدريب

a. استخدام الدالة سيغمويد (Sigmoid):

دالة السيغمويد (Sigmoid) هي إحدى **دوال التنشيط** الشائعة المستخدمة في تدريب الشبكات العصبية، وتُستخدم بشكل خاص لتحويل المدخلات الخطية إلى مخرجات غير خطية. تتميز دالة السيغمويد بأنها تعطي ناتجًا محدودًا بين القيمتين 0 و1، مما يجعلها مناسبة لنماذج التصنيف، خصوصًا تلك التي تتطلب تنبؤات احتمالية [97] [98].

دالة السيغمويد تأخذ المدخلات (input) الموزونة وتحولها إلى قيم غير خطية باستخدام الصيغة التالية: $\sigma(x) = \frac{1}{1+e^{-x}}$.

- حيث x هو مجموع المدخلات الموزونة، وe هو ثابت أويلر (تقريبًا 2.718).
- ناتج الدالة يكون دائمًا بين 0 و1. وعندما تكون x موجبة كبيرة، فإن دالة السيغمويد تقترب من 1، وعندما تكون x سالبة كبيرة، تقترب الدالة من 0.

b. التدريب التدريجي:

تم تنفيذ التدريب عبر عدة دورات (Epochs) لتمكين الشبكة من تحسين أدائها تدريجيًا. خلال كل دورة، يتم تحسين الأوزان المخصصة للروابط العصبية اعتماداً على النتائج المحققة في الدورة السابقة. استمر التدريب حتى الوصول إلى معدل خطأ منخفض ومستقر [99] [100].

4.2.3.3 اختبار النموذج وتحليل الأداء:

أ. تقييم الأداء باستخدام مجموعة التحقق:

تم استخدام مجموعة التحقق (Validation Set) لاختبار أداء الشبكة بعد التدريب. تم تقييم دقة النموذج باستخدام مقاييس متعددة مثل:

- **معدل الخطأ التربيعي المتوسط:** (Mean Squared Error - MSE) لقياس الفارق المتوسط بين القيم التجريبية والقيم المتوقعة [101] [102].
- **معامل التحديد:** (R^2) لقياس مدى قدرة النموذج على تفسير التباين في البيانات.
- **معدل الخطأ المطلق المتوسط:** (Mean Absolute Percentage Error - MAPE) لقياس نسبة الفروق بين القيم المتوقعة والحقيقية بالنسبة إلى القيم الحقيقية [101] [102].

ب. تحسين النموذج

❖ **تعديل بنية الشبكة ومعلمات التدريب:**

تمت إعادة ضبط معلمات الشبكة، مثل عدد العُقَد ومعدل التعلم بناءً على نتائج الاختبار لتحسين أداء النموذج. تم تحسين الهيكل الداخلي للشبكة وإعادة تدريبها لتحسين دقتها وتقليل الفروقات بين القيم المتوقعة والتجريبية [103] [104].

❖ **استخدام تقنيات تحسين الأداء:**

تم استخدام خوارزميات إضافية مثل **Adam** و**RMSProp** لتحسين أداء النموذج وتسريع عملية التعلّم. تم أيضًا استخدام تقنيات **إيقاف التدريب المبكر** (Early Stopping) لمنع الإفراط في التكيف [105].

3.3.3 هيكل الشبكة العصبونية المستخدم في هذه الدراسة:

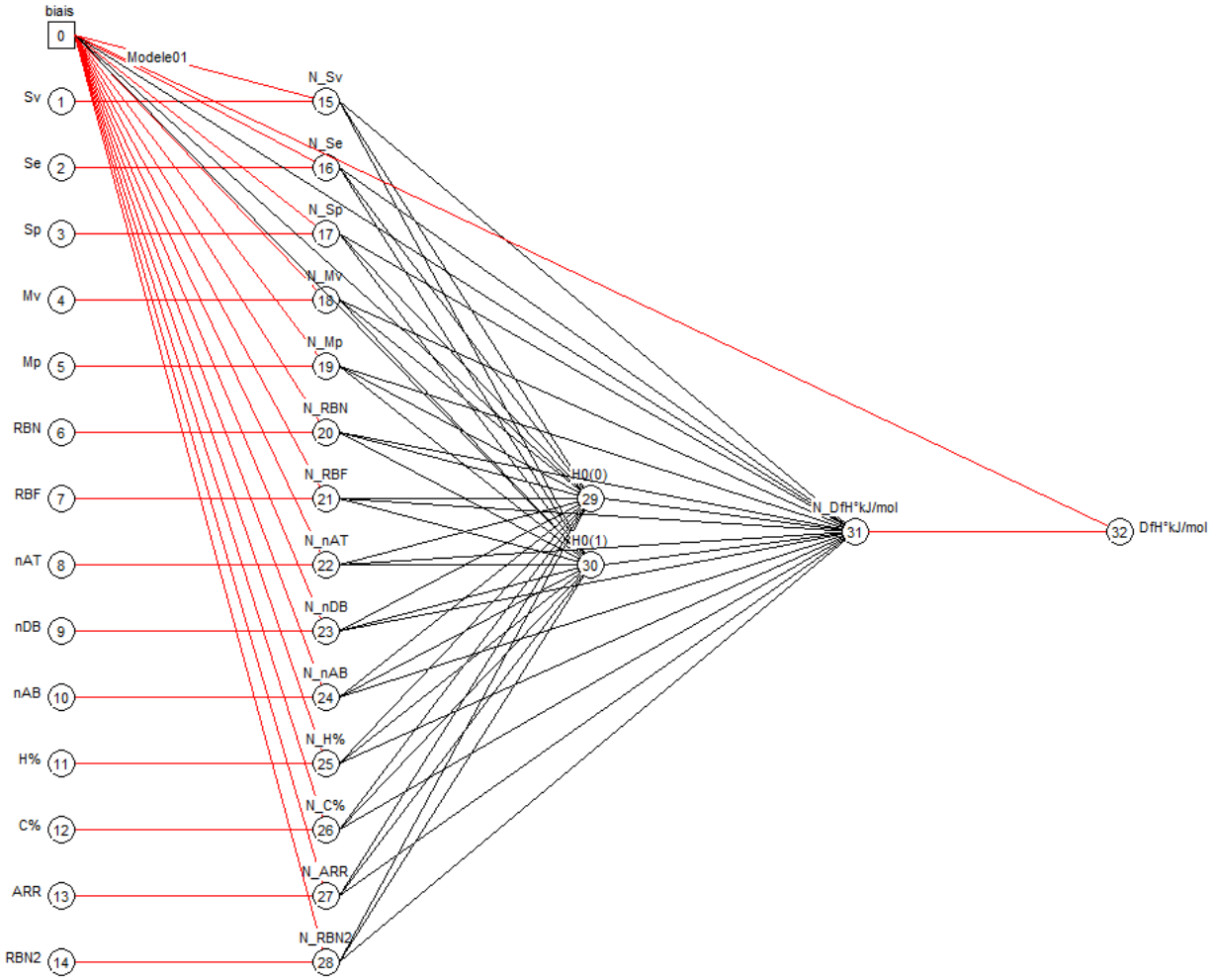
هيكل الشبكة العصبونية (Neural Network) يتميز بتكوين متسلسل من الطبقات التي تتفاعل فيما بينها لمعالجة البيانات المدخلة والوصول إلى المخرجات النهائية. يوضح الرسم البياني ادناه الخطوات الأساسية التي تمر بها البيانات من المدخلات حتى الوصول إلى المخرجات في الشبكة العصبونية. لنوضح هذا الهيكل بشكل مفصل:

1.3.3.3 المدخلات (Input Layer):

البيانات الأولية أو السمات (features) التي يتم إدخالها إلى الشبكة تُمرر عبر **طبقة المدخلات**. كل وحدة أو عقدة (Node) في هذه الطبقة تمثل متغيرًا من متغيرات البيانات المدخلة .على سبيل المثال، في

حالة النموذج الكيميائي، قد تكون المدخلات هي الواصفات الجزيئية (مثل الحجم الجزيئي أو الروابط الكيميائية).

- كل عقدة في طبقة المدخلات تمثل سمة معينة من البيانات.
- لا يحدث أي تعديل على البيانات في هذه الطبقة؛ فهي فقط تعمل كنقطة دخول للمعلومات.



الشكل 3.3: يوضح الرسم البياني بشكل عام كيفية إدخال البيانات وخطوات المرور المؤدية إلى المخرج للشبكة العصبونية.

2.3.3.3 الطبقة المخفية (Hidden Layer):

تتكون الشبكات العصبونية عادةً من واحدة أو أكثر من الطبقات المخفية بين طبقة المدخلات والمخرجات [106].

- كل عقدة في الطبقة المخفية تتلقى مدخلات من جميع العقد في الطبقة السابقة (سواء كانت طبقة المدخلات أو طبقة مخفية أخرى) ثم تُجري عليها عملية حسابية بناءً على الأوزان (Weights) المرتبطة بكل رابط [106].

- يتم تطبيق دالة تنشيط (Activation Function) مثل دالة السيغمويد أو دالة (Rectified Linear Unit) ReLU على هذه المدخلات الموزونة لتحويلها إلى مخرجات غير خطية [106] [107].
- هذه المخرجات تُرسل بعد ذلك إلى الطبقة التالية.

كل طبقة مخفية تلعب دورًا مهمًا في استخراج الأنماط المخفية في البيانات، حيث تعمل الأوزان ودوال التنشيط على تحويل البيانات إلى أشكال تُساعد على اتخاذ القرارات في الشبكة [108].

3.3.3.3 الطبقة المخرجة (Output Layer):

الطبقة النهائية هي طبقة المخرجات التي تقدم النتيجة النهائية للشبكة بناءً على جميع العمليات التي تمت في الطبقات السابقة.

- يعتمد عدد العقد في هذه الطبقة على عدد الفئات أو المتغيرات المستهدفة [109].
- على سبيل المثال، إذا كانت الشبكة تهدف إلى توقع قيمة معينة، مثل المحتوى الحراري، سيكون هناك عقدة واحدة تمثل هذه القيمة المتوقعة [110].

4.3.3.3 التمرير الأمامي:(Forward Propagation)

هذه العملية تشير إلى تدفق البيانات من طبقة المدخلات مروراً بالطبقات المخفية حتى الوصول إلى طبقة المخرجات. في هذه المرحلة:

- يتم حساب مجموع المدخلات لكل عقدة من خلال ضرب قيم المدخلات في الأوزان الخاصة بها، ثم جمع النتيجة [111].
- يتم تمرير مجموع هذه المدخلات من خلال دالة التنشيط في الطبقة المخفية للوصول إلى ناتج غير خطي [111].
- في النهاية، يتم إنتاج مخرجات الشبكة في الطبقة النهائية [111].

5.3.3.3 التمرير الخلفي (Backpropagation):

بعد الحصول على المخرجات، يتم مقارنة هذه المخرجات بالقيم الحقيقية أو المتوقعة عبر دالة خسارة (Loss Function) لتحديد الفارق (الخطأ). ثم:

- تُستخدم عملية التمرير الخلفي لحساب كيفية تعديل الأوزان في الشبكة لتقليل الخطأ في التنبؤات [112].

- تتم هذه العملية عن طريق حساب التدرج (Gradient) لدالة الخسارة بالنسبة لكل وزن، وتحديث الأوزان باستخدام خوارزمية تحسين مثل خوارزمية الانحدار التدريجي (Gradient Descent) [112].

الخاتمة:

في هذا الفصل، تم استعراض المنهجية التفصيلية التي تم اتباعها لبناء وتقييم نموذج تنبؤي قائم على تقنيات QSPR، بهدف تقدير الخصائص الفيزيائية والكيميائية للمركبات العضوية بدقة. شملت المنهجية خطوات متسلسلة بدءًا من تجميع مجموعة بيانات شاملة من المركبات، وحساب مجموعة متنوعة من الواصفات الجزيئية التي تصف بنيتها، وصولاً إلى بناء وتدريب النموذج التنبؤي. كما تم مناقشة أهمية معالجة البيانات وتحسين أداء النموذج للحصول على نتائج أكثر دقة.

الفصـــل الرابـــع:

النتائـــــــج والمناقشـــــة

الفصل الرابع: النتائج والمناقشة

في هذه الدراسة، تم تطوير نموذجين إحصائيين مختلفين لتقدير المحتوى الحراري القياسي لتكوين المركبات العضوية، وهما :النموذج الانحداري الخطي المتعدد (MLR) الذي تم تحليله باستخدام برنامج XLSTAT، والنموذج القائم على الشبكات العصبية الاصطناعية (ANN) والذي تم تطويره باستخدام برنامج NeuroOne 5. يهدف هذا البحث إلى مقارنة أداء هذين النموذجين في التنبؤ بالأنشطة الكيميائية لمجموعة من المركبات التجريبية.

1.4. نتائج الانحدار الخطي المتعدد:

تم تطبيق تحليل الانحدار الخطي المتعدد (MLR) كأداة إحصائية لتحديد العلاقة الخطية بين المحتوى الحراري القياسي لتكوين ومجموعة من المتغيرات المستقلة. شملت العملية اختيار المتغيرات ذات الأهمية الإحصائية من خلال إجراء تحليل انحدار تدريجي (stepwise regression)، حيث تم إضافة أو حذف المتغيرات بشكل متكرر بناءً على معايير الإحصاء الوصفي) مثل معامل التحديد (R^2 حتى الوصول إلى النموذج الأمثل الذي يشمل جميع المتغيرات ذات التأثير الملحوظ. وقد أسفر هذا التحليل عن النموذج الخطي التالي لوصف المحتوى الحراري القياسي لتكوين:

$$\Delta H^{0}_{f(gaz)} = 750+110,885*Sv-72,421*Se+2343,949*Mv+ 4186,0325*Mp+8,2478*RBN+109,965*RBF+1,127*nDB-6,677*nAB-3293,503*H\%-6480,973*C\%-168,185*ARR \quad \text{..... (4.1)}$$

حيث:

- $\Delta H^{0}_{f(gaz)}$: المحتوى الحراري القياسي.

A. الواصفات البنيوية:

يتضمن الجدول 1.4 المتغيرات المستخدمة في النموذج لحساب المحتوى الحراري القياسي لتكوين للمركبات العضوية في الحالة الغازية. هذه المتغيرات متمثلة في الواصفات البنيوية التي يتم ربطها بمعاملات خاصة بكل واصف، إضافة إلى الثابت A الذي يساوي 750. ويعرض الجدول القيم الخاصة بكل واصف بوحدة كيلو-جول/مول، إلى جانب الوزن المرتبط بكل منها. هذا النموذج يوضح كيفية تأثير كل واصف بنيوي على المحتوى الحراري القياسي، حيث يمكن تلخيص المتغيرات والمعاملات المرتبطة بها كما يلي:

جدول 1.4: يبين العلاقة بين الواصفات البنيوية والمعاملات الحرارية.

الواصف البنيوي (xi)	القيمة (كيلو-جول/مول)	الوزن (wi)
الثابت (A)	750,00	/
Sv	2,12	110,88
Se	-2,35	-72,42
Mv	1,66	2343,95
Mp	3,22	4186,03
RBN	0,04	8,25
RBF	0,01	109,97
nDB	0,00	1,13
nAB	-0,05	-6,68
H%	-2,52	-3293,50
C%	-3,15	-6480,97
ARR	-0,08	-168,18

يمثل هذا الجدول العلاقة بين الواصفات البنيوية والمعاملات الحرارية، مما يساهم في بناء نموذج حسابي يهدف إلى التنبؤ بالمحتوى الحراري القياسي لتكوين المركبات بناءً على خصائصها البنيوية. وتشمل هذه الوصفات:

- **ثابت :A** قيمته الثابتة 750 كيلوجول/مول تمثل قيمة أساسية في النموذج.
- **الوصفات الذرية :(Sv, Se)** تعبر عن مساهمة الذرات المختلفة (مثل الكبريت والهالوجينات) في المحتوى الحراري، مع الأخذ في الاعتبار نوع الذرة وبيئتها الكيميائية.
- **الوصفات الجزيئية :(Mv, Mp)** تمثل مساهمة مجموعات الذرات المختلفة (مثل الميثيل والميثيلين) في المحتوى الحراري، مع مراعاة نوع المجموعة وعددها في الجزيء.
- **الوصفات البنيوية الأخرى :(RBN, RBF, nDB, nAB)** تعكس تأثيرات أخرى على المحتوى الحراري، مثل وجود روابط مزدوجة أو ثلاثية، وتفرع السلسلة، وحلقات الأروماتية.
- **الوصفات العناصرية :(H%, C%, ARR)** تمثل نسبة الهيدروجين والكربون في الجزيء، بالإضافة إلى تأثير ترتيب الذرات في السلسلة.

لكل وصف من هذه الوصفات وزن (wi) يعبر عن أهميته النسبية في تحديد المحتوى الحراري. ويتم حساب المحتوى الحراري النهائي للمركب بجمع مساهمات كل وصف مضروبة في وزنها، بالإضافة إلى الثابت A.

B. معاملات التعديل للمحتوى الحراري القياسي (كيلو جول/مول):

يتم في الجداول التالية تقديم الإحصائيات المتعلقة بمعاملات التوافق الخاصة بنموذج الانحدار الخطي المتعدد المستخدم لتقدير المحتوى الحراري القياسي بناءً على الواصفات البنيوية. هذه الإحصائيات تقدم لمحة عن مدى دقة النموذج وقوة توافقه مع البيانات المستخدمة في التدريب والتحقق. تُظهر الجداول مؤشرات الأداء مثل معامل التحديد (R^2) والقيم المعدلة منه، بالإضافة إلى مؤشرات الخطأ مثل متوسط مربع الخطأ (MCE)، والجذر التربيعي لمتوسط مربع الخطأ (RMCE)، والنسبة المئوية للخطأ المطلق المتوسط (MAPE).

فيما يلي ملخص للإحصائيات كما هو موضح في الجدول VI.2، والذي يتضمن معاملات التوافق للنموذج لكل من مجموعة التعلم ومجموعة التحقق:

الجدول 2.4: معاملات التوافق للنموذج الخاصة بالانحدار الخطي.

الإحصائية	عينة التعلم (Apprentissage)	عينة التحقق (Validation)
عدد قيم المركبات العضوية	266	72
معامل التحديد (R^2)	0.9962	0.9602
معامل التحديد المعدل (R^2 ajusté)	0.996	0.9585
متوسط مربع الخطأ (MCE)	2616.9	1757.6
الجذر التربيعي لمتوسط مربع الخطأ (RMCE)	51.156	41.924
النسبة المئوية للخطأ المطلق المتوسط (MAPE)	41.911	41.757

يقدم الجدول 2.4 إحصائيات أداء نموذج الانحدار الخطي المتعدد الذي تم تطويره لتقدير المحتوى الحراري القياسي للمركبات العضوية بناءً على وصفاتها البنيوية .تم تقسيم البيانات إلى مجموعتين :مجموعة تدريب تضم 266 مركباً ومجموعة اختبار مستقلة تضم 72 مركباً.

يشير معامل التحديد (R^2) المرتفع جداً في كلا المجموعتين (0.9962 لمجموعة التدريب و0.9602 لمجموعة الاختبار) إلى أن النموذج يستطيع تفسير نسبة كبيرة جداً من التباين في البيانات، مما يدل على ملاءمة النموذج للبيانات. كما يشير معامل التحديد المُعدل (R^2 adjusted) إلى أن إضافة متغيرات إضافية إلى النموذج لن يؤدي إلى تحسين كبير في أدائه.

يشير متوسط الخطأ التربيعي (MSE) والجذر التربيعي لمتوسط الخطأ التربيعي (RMSE) إلى حجم الأخطاء المتوقعة في التنبؤات .قيم هذه الإحصائيات منخفضة نسبياً، مما يدل على دقة النموذج في التنبؤ بالقيم الحقيقية للمحتوى الحراري.

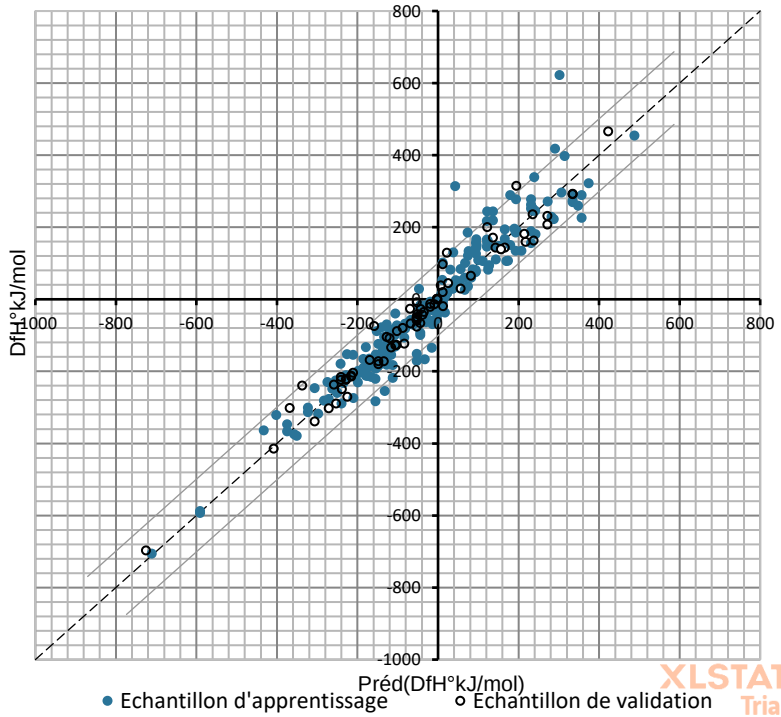
يشير متوسط الخطأ المطلق المئوي (MAPE) إلى متوسط النسبة المئوية للخطأ المطلق في التنبؤات. قيم MAPE منخفضة أيضاً، مما يشير إلى أن الأخطاء النسبية في التنبؤات صغيرة.

بشكل عام، تشير هذه النتائج إلى أن النموذج المقترح يقدم أداءً ممتازاً في تقدير المحتوى الحراري القياسي للمركبات العضوية، ويمكن الاعتماد عليه في التطبيقات المستقبلية.

C. تحليل عام للرسوم البيانية:

يوضح الرسم البياني العلاقة بين القيم التجريبية والمتوقعة لحرارة التكوين القياسية ($\Delta Hf°$) لمجموعة من المركبات الكيميائية، والتي تم الحصول عليها باستخدام نموذج انحدار خطي متعدد يعتمد على 14 متغيرًا بنيويًا. يمثل كل نقطة على الرسم البياني مركباً واحداً، حيث تمثل الإحداثيات الأفقية القيمة التجريبية (المقاسة عمليًا) والإحداثيات الرأسية القيمة المتوقعة (المحسوبة بواسطة النموذج).

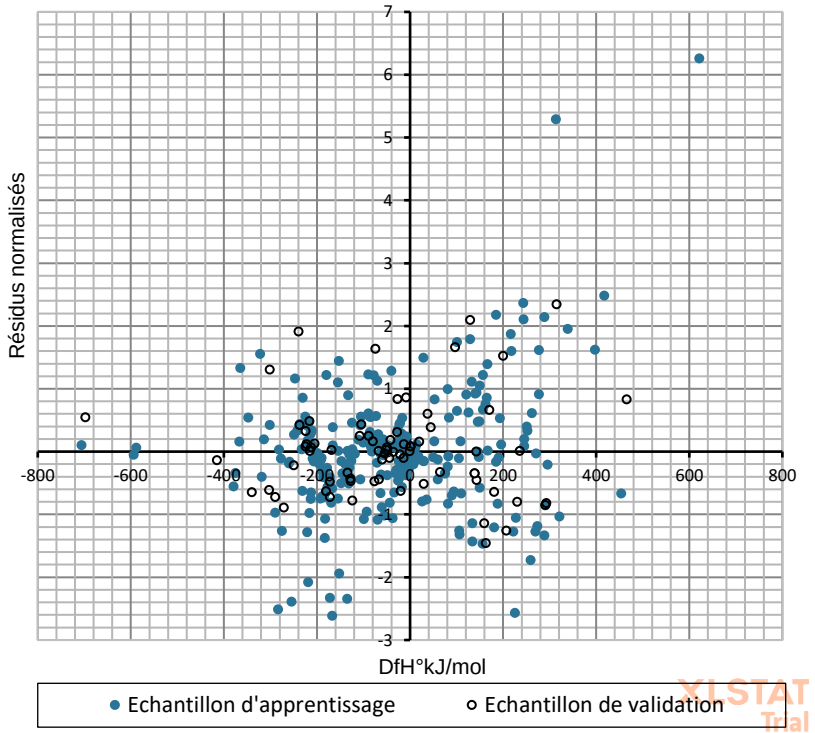
يشير مدى تقارب النقاط من خط الانحدار إلى جودة النموذج في التنبؤ بقيم $\Delta Hf°$. كما يوضح الرسم البياني مجال الثقة الذي يمثل مدى التباين المتوقع في القيم المتوقعة، مما يساعد في تحديد المركبات التي تم التنبؤ بخصائصها بدقة أكبر والتي تتطلب مزيدًا من الدراسة .النقاط التي تقع بعيدًا عن خط الانحدار وخارج مجال الثقة (Intervalle de confiance) تشير إلى وجود انحرافات قد تكون ناجمة عن عوامل لم يتم تضمينها في النموذج أو أخطاء في البيانات التجريبية.



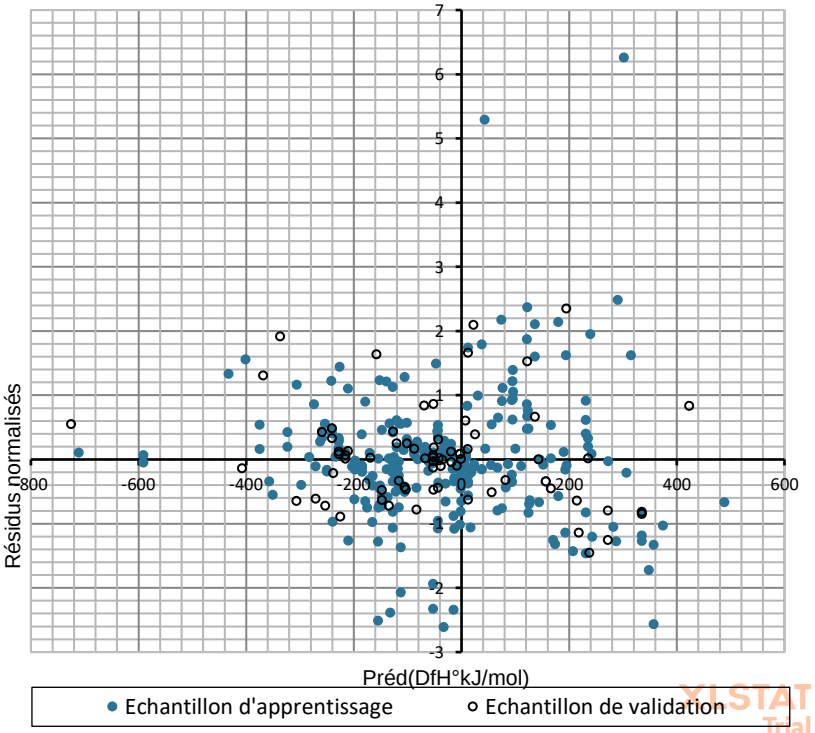
الشكل 1.4: يوضح العلاقة الخطية بين القيم التجريبية والتنبؤية للمحتوى الحراري القياسي لتكوين، مع إبراز مجال الثقة والنقاط التي تشير إلى قوة أو ضعف ارتباط الواصفات البنيوية بالنموذج.

هذا الرسم البياني يُظهر العلاقة بين القيم التجريبية والمتوقعة. يُفترض أن تكون النقاط قريبة من خط مستقيم بزاوية 45 درجة (يمثل الخط المثالي حيث تتطابق القيم التجريبية مع المتوقعة). كلما كانت

النقاط أقرب إلى هذا الخط، كان النموذج أفضل في التنبؤ. نلاحظ وجود نقطتين بعيدتين عن هذا الخط، مما يدل على أن النموذج لم يتنبأ بقيمتهما بدقة. هاتان النقطتان هما 1,8 Cyclotetradecadiyne- و(Cubane) Pentacyclo- [4.2.0.02,5.03,8.04,7] octane.



الشكل 2.4: توضيح كيف تتوزع القيم التجريبية مع البواقي (**Residuals**).



الشكل 3.4: توضيح كيف تتوزع القيم المتوقعة مع البواقي (**Residuals**).

الرسمان البيانيان (الشكل 2.4، الشكل 3.4) يُظهران توزيع القيم المتبقية (الفرق بين القيم التجريبية والمتوقعة) مقابل القيم المتوقعة. يُفترض أن تتوزع النقاط بشكل عشوائي حول الخط الأفقي الذي يمثل قيمة صفرية للبواقي. أي نمط في توزيع البواقي (مثل شكل قمع أو منحنى) يشير إلى وجود مشكلة في النموذج. نلاحظ هنا أن البواقي تتوزع بشكل عشوائي إلى حد كبير، باستثناء النقطتين البعيدتين وهما كالاتي:

- **1,8-Cyclotetradecadiyne (313.8 kJ/mol):**

هذه النقطة تقع أعلى مجال الثقة. هذا يعني أن النموذج قلل من تقدير قيمة ΔHf° لهذا المركب بشكل كبير. قد يكون هذا بسبب خصائص فريدة لهذا المركب لم يتم تمثيلها بشكل جيد في المتغيرات المستخدمة في النموذج.

- **Pentacyclo- [4.2.0.02,5.03,8.04,7] octane (Cubane) (622.2 kJ/mol):**

هذه النقطة أيضاً تقع أعلى مجال الثقة وتعتبر أبعد من النقطة السابقة. هذا يشير إلى أن النموذج قلل من تقدير قيمة ΔHf° لهذا المركب بشكل كبير جداً. وبالمثل، قد يكون هذا بسبب خصائص فريدة لمركب Cubane، مثل الشكل المكعب غير المعتاد، الذي لم يتم تضمينه في النموذج.

1.1.4. تحليل ومناقشة نتائج الانحدار الخطي المتعدد:

تم تحديد الواصفات البنيوية ذات العلاقة الإحصائية القوية بالمحتوى الحراري القياسي للمركبات من خلال تطبيق التحليل الانحداري التدريجي، وهي تقنية إحصائية تعتمد على اختيار المتغيرات الأكثر تأثيراً بشكل متدرج حتى الوصول إلى النموذج الأمثل. هذه الطريقة تساهم كما قال ميرس واخرون (Myers et al., 2016) [113] في تحسين دقة النموذج من خلال إزالة المتغيرات غير المؤثرة بشكل كبير وإبقاء فقط تلك التي تُظهر علاقة جوهرية مع المحتوى الحراري، مما يساعد في بناء نموذج أكثر فعالية.

حقق النموذج النهائي معامل تحديد مرتفع ($R^2 = 0.9962$ لعينة التعلم و0.9602 لعينة التحقق)، مما يعكس قدرة النموذج على تفسير 99.62% من التباين في البيانات المستخدمة للتعلم و96.02% من التباين في البيانات غير المدربة (التحقق). يشير هذا إلى توافق كبير بين التنبؤات والقيم الحقيقية، مما يعني أن النموذج قادر على تمثيل البيانات بشكل دقيق. في الأبحاث السابقة مثل مونتقمري واخرون (Montgomery et al., 2012) [114]، يعتبرون معامل التحديد المرتفع مؤشراً حاسماً على كفاءة النماذج التنبؤية في الكيمياء الحسابية.

تشكل الواصفات البنيوية، مثل نسبة الكربون والهيدروجين ووجود الروابط الثنائية أو الثلاثية، جزءاً أساسياً من المتغيرات المدخلة في النموذج. لكل من هذه الواصفات وزن نسبي يعكس مدى تأثيرها على المحتوى الحراري. فمثلاً، لوحظ أن زيادة نسبة الكربون ترتبط بتقليل المحتوى الحراري، بينما الروابط

الثنائية والثلاثية تزيد من المحتوى الحراري بشكل ملحوظ. هذه النتائج تتماشى مع الدراسات التي تؤكد على أن الطبيعة البنيوية للمركبات تلعب دورًا رئيسيًا في خصائصها الحرارية [115]، [116].

تظهر مؤشرات الأداء مثل الجذر التربيعي لمتوسط مربع الخطأ (RMCE = 51.156 لعينة التحقق) والنسبة المئوية للخطأ المطلق (MAPE = 41.911 لعينة التحقق) أن النموذج يقدم تنبؤات دقيقة للغاية، مع الأخذ في الاعتبار تقارب القيم المتوقعة مع القيم التجريبية. يشير RMCE إلى حجم الخطأ المتوقع لكل مركب، مما يعني أن الفرق بين القيم المتوقعة والتجريبية هو منخفض إلى حد كبير، وهو ما يعزز دقة النموذج.

من خلال تحليل توزع النقاط حول خط الانحدار، يتضح أن معظم القيم المتوقعة تتطابق بشكل كبير مع القيم التجريبية. توزيع النقاط على خط 45 درجة المثالي يوضح مدى قوة النموذج في التنبؤ الدقيق بالقيم الحرارية القياسية. هذا الأداء القوي في التنبؤ يتوافق مع العديد من الدراسات السابقة مثل فراواي (Faraway, et al. 2014) [117] التي أكدت أن النماذج الخطية المتعددة تُظهر دقة عالية في التنبؤات الكيميائية إذا تم اختيار المتغيرات بحذر.

على الرغم من الدقة العالية للنموذج، إلا أنه ظهرت انحرافات ملحوظة لبعض المركبات مثل Cyclotetradecadiyne-1,8 وPentacyclo-octane. تشير هذه الانحرافات إلى أن هذه المركبات تمتلك خصائص فريدة لم تُؤخذ في الاعتبار ضمن النموذج. قد تكون هذه الخصائص ناتجة عن بنيتها الفريدة أو التأثيرات البيئية التي تؤثر على استقرار المركبات. تعكس هذه الانحرافات الحاجة إلى تحسين النموذج أو إدخال مزيد من المتغيرات التي قد تأخذ في الاعتبار هذه الفروق كما جاء في دراسة درابير وسميث (Draper & Smith) [118].

2.4 نتائج الانحدار غير الخطي باستعمال الشبكات العصبونية الاصطناعية (NRN):

تم استخدام تحليل الانحدار غير الخطي (Neural Network Regression - NNR) كأداة إحصائية قوية لنمذجة العلاقة المعقدة وغير الخطية بين المحتوى الحراري القياسي لتكوين المركبات وبين مجموعة من المتغيرات المستقلة .وقد تم تطبيق تقنية الانحدار التدريجي (stepwise regression) لاختيار المجموعة المثلى من المتغيرات المؤثرة في المحتوى الحراري .وبهذه الطريقة، تم التوصل إلى نموذج انحدار عصبي يصف بدقة العلاقة بين المتغيرات المستقلة والمحتوى الحراري، حيث تم تحديد المتغيرات ذات الأهمية الإحصائية بناءً على معايير مثل معامل التحديد (R^2). وقد أسفر هذا التحليل عن النموذج غير الخطي التالي لوصف المحتوى الحراري القياسي لتكوين:

$$\Delta H_{Vap} = -39.755 + 184.944 * \left(w_{30} + \frac{w_{31}}{2}(1+\tanh N) + \frac{w_{32}}{2}(1+\tanh M) + \sum_{i=33, j=0}^{46,13} w_i * (\sigma_j + P_j x_j) \right)$$

$$M = w_{15} + \sum_{i=16, j=0}^{29,13} w_i + \frac{w_{32}}{2}(1+\tanh N) \qquad (4.2)$$

$$N = w_0 + \sum_{i=1, j=0}^{14,13} w_i * (\sigma_j + b_j x_j)$$

في الجداول 3.4 و4.4، تم عرض المتغيرات الخاصة بالنموذج المستخدم لحساب الخاصية المستهدفة، والتي تمثلها الواصفات الجزيئية. توضح هذه الجداول التفاصيل الدقيقة لكل واصف جزيئي، بما في ذلك معاملاته الخاصة، والتي تعكس مدى تأثير كل واصف على التنبؤ بالخاصية المدروسة. تمثل هذه المعاملات القيم العددية المرتبطة بالواصفات المختلفة، حيث يتم تقديرها باستخدام الأساليب الحسابية التي تم الاعتماد عليها في بناء النموذج. توفر هذه الجداول رؤى حول الأهمية النسبية لكل واصف جزيئي ومدى مساهمته في تحسين دقة النموذج، مما يتيح تحليلًا شاملًا لتأثير البنية الجزيئية على الخاصية المراد حسابها.

الجدول 3.4: المتغيرات الخاصة بنموذج الشبكة العصبونية الاصطناعية.

xj	Pj	σj
Sv	0,153267824	-2,159425739
Se	0,09098318	-2,173211325
Sp	0,141623522	-2,17199364
Mv	17,43008037	-9,895438229
Mp	20,36514135	-12,62036245
RBN	0,294568953	-0,459283545
RBF	14,35312081	-0,733240642
nAT	0,087624618	-2,170653621
nDB	1,421310211	-0,62655391
nAB	0,17189118	-0,470412253
H%	12,42638458	-7,649449148
C%	12,42638458	-4,776935428
ARR	3,028190797	-0,529216658
RBN2	0,294568953	-0,459283545
N_DfH	0,005407028	0,214954948

الجدول 4.4: الاوزان الخاصة بنموذج الشبكة العصبونية الاصطناعية.

Poids (wi)	Valeur	Poids (wi)	Valeur
w(0)	0	w(24)	-0,07508
w(1)	-0,02178	w(25)	0,005244
w(2)	0,080971	w(26)	-0,03251
w(3)	-0,00617	w(27)	-0,056
w(4)	0,067246	w(28)	0,273649
w(5)	-0,00194	w(29)	-0,04845
w(6)	7,11E-05	w(30)	0
w(7)	0,001598	w(31)	0,016664
w(8)	0,119949	w(32)	-0,0202
w(9)	0,097959	w(33)	0,121104
w(10)	0,027484	w(34)	0,203005
w(11)	0,168735	w(35)	0,17085
w(12)	0,040736	w(36)	-0,0519
w(13)	0,102593	w(37)	0,027672
w(14)	-0,01566	w(38)	0,0709
w(15)	0	w(39)	0,045157
w(16)	0,037233	w(40)	-0,04187
w(17)	0,027062	w(41)	-0,11378
w(18)	0,080901	w(42)	0,105735
w(19)	-0,03365	w(43)	0,064088
w(20)	0,02545	w(44)	-0,0197
w(21)	0,100041	w(45)	0,085431
w(22)	-0,07612	w(46)	0,039951
w(23)	-0,03709		

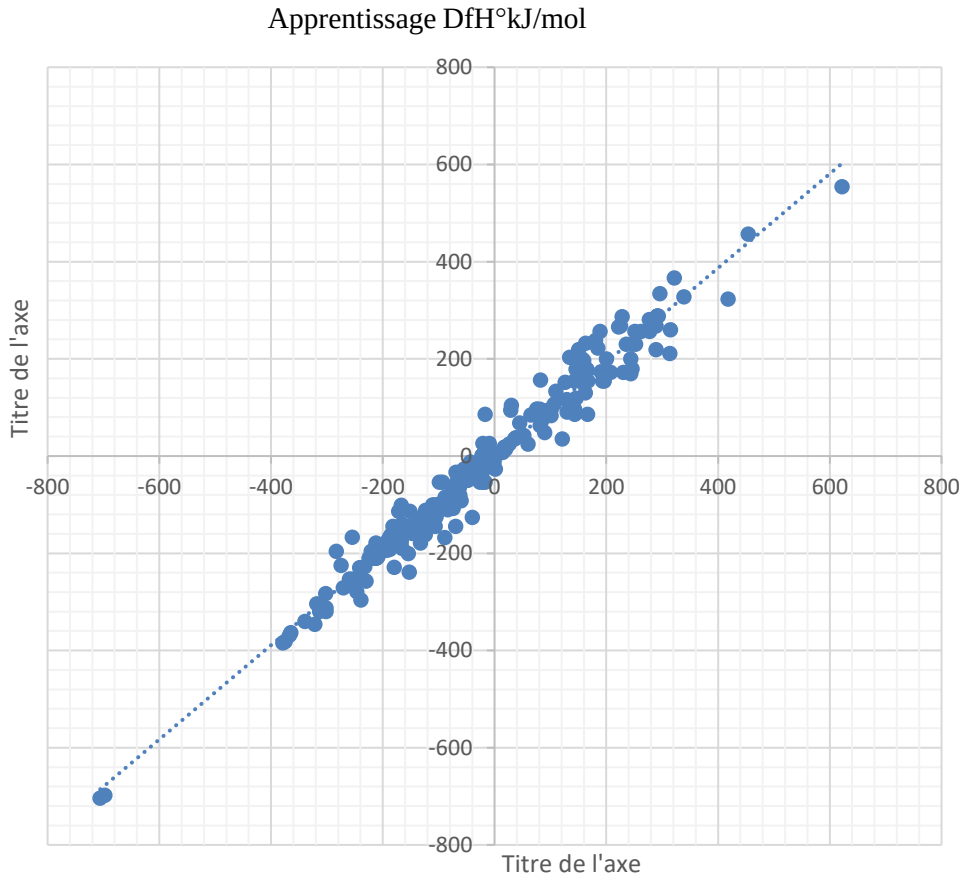
في الجداول أدناه، تم تقديم الإحصائيات المتعلقة بمعاملات التوافق الخاصة بنموذج الانحدار الذي يعتمد على الشبكة العصبية الاصطناعية .(ANN) هذه الإحصائيات توفر تحليلاً دقيقًا لمستوى التوافق بين القيم التنبؤية للنموذج والقيم الفعلية للبيانات المستخدمة. تشمل الإحصائيات الرئيسية مقاييس الأداء مثل معامل الارتباط ومعدل الخطأ، والتي تُمكن من تقييم فعالية النموذج ودقته في التنبؤ. بالإضافة إلى ذلك،

تُوضح الجداول كيفية ضبط المعاملات داخل الشبكة العصبية ومدى تأثيرها على دقة النتائج، مما يساعد على فهم دور كل عنصر في تحسين أداء النموذج.

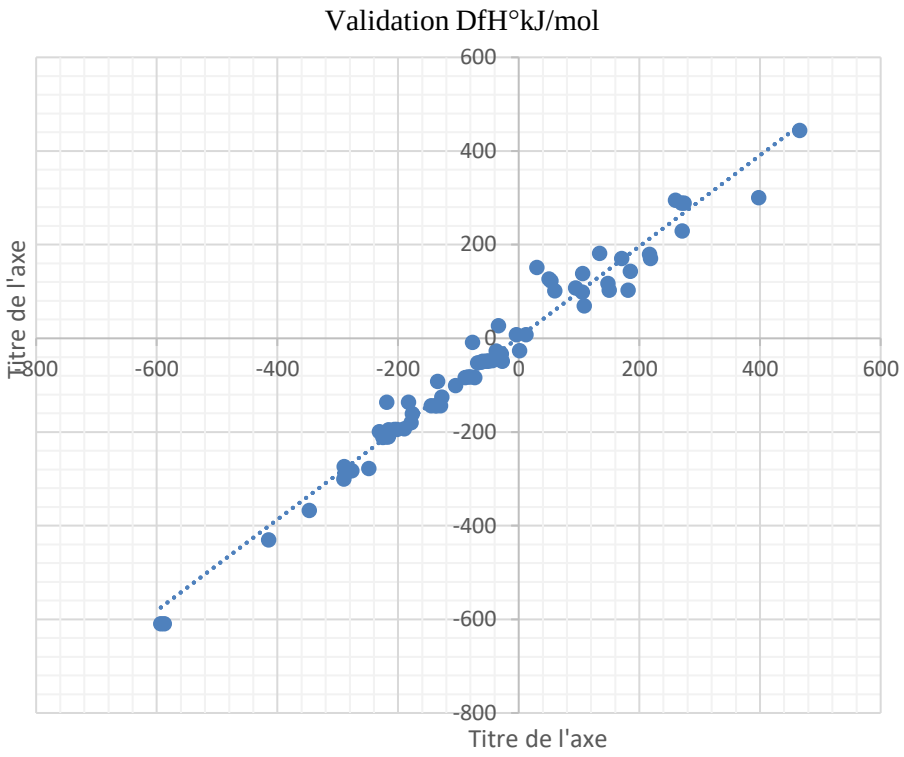
الجدول 5.4: معاملات التوافق للنموذج الخاصة بالانحدار للشبكة العصبونية الاصطناعية.

المعاملات	R^2 Tra set	R^2 Val	EQMA	EQMV
قيمة المعاملات	0,984846	0,984843	488,6246	604,0534

يمثل الرسمان البيانيان في الشكلين 4.4 و5.4 العلاقة غير الخطية المستمدة من نموذج الشبكة العصبية الاصطناعية بين القيم التجريبية والقيم المحسوبة لتدريب النموذج والتحقق من صحة النتائج المتعلقة بحساب المحتوى الحراري القياسي للتكوين .يعكس الرسم البياني مدى تطابق البيانات التجريبية مع التنبؤات الحسابية، موضحًا مجال الثقة لكل نقطة تنبؤية. كما يشير إلى النقاط الأكثر والأقل ارتباطًا بالواصفات الجزيئية (المتغيرات المستقلة)، مما يساعد على تحديد مدى دقة النموذج في التنبؤ بخصائص المركبات المختلفة. يبرز الرسم بشكل واضح النقاط التي تتوافق بشكل جيد مع التوقعات، إضافة إلى النقاط التي تظهر انحرافًا عن القيم المتوقعة، مما يوفر معلومات قيمة حول فعالية النموذج ومدى تأثير كل واصف جزيئي على النتائج التنبؤية.



الشكل 4.4: العلاقة غير الخطية بين القيم التجريبية والتنبؤية لتدريب (Apprentissage).



الشكل 5.4: العلاقة غير الخطية بين القيم التجريبية والتنبؤية لتحقق (Validation).

1.2.4 تحليل ومناقشة نتائج الانحدار غير الخطي باستعمال الشبكات العصبونية الاصطناعية (NRN):

اشارت النتائج المعروضة في هذا الفصل إلى نجاح استخدام نموذج الانحدار غير الخطي بالاعتماد على الشبكات العصبونية الاصطناعية (NNR) في التنبؤ بالمحتوى الحراري القياسي لتكوين المركبات. دعونا نحلل هذه النتائج بمزيد من التفصيل:

- معامل التحديد :(R^2) بلغت قيمة R^2 لكل من مجموعة التدريب (Tra set) ومجموعة التحقق من الصحة0.9848 (Val) ، وهذا يشير إلى توافق عالي جدًا بين القيم التجريبية والقيم المتنبئ بها من قبل النموذج. يدل هذا على قدرة النموذج على تفسير 98.48% من التباين في المحتوى الحراري القياسي لتكوين المركبات. هذه النسبة عالية وتشير إلى أداء ممتاز للنموذج.
- الخطأ المتوسط التربيعي :(EQM) بلغت قيمة EQM لمجموعة التدريب 488.62 ولمجموعة التحقق من الصحة 604.05. يشير الاختلاف البسيط بين قيمتي EQM لمجموعتي التدريب والتحقق إلى أن النموذج لم يعاني من مشكلة فرط التخصيص(overfitting) ، مما يعزز من موثوقية النتائج.

تم استخدام تقنية الانحدار التدريجي (stepwise regression) لاختيار المتغيرات الأكثر تأثيرًا في المحتوى الحراري القياسي للتكوين. يوضح الجدول 3.4 المتغيرات المختارة ومعاملاتها، ولكن يفتقر

إلى شرح طبيعة هذه المتغيرات ودلالتها الفيزيائية والكيميائية. من الضروري توضيح معاني هذه الرموز مثل (Se, Sp, Mv, Mp, RBN, nAT, H%, C%...) لتفسير النموذج بشكل صحيح. فهم طبيعة هذه المتغيرات يسمح بتحديد العوامل الجزيئية التي تؤثر بشكل أكبر على المحتوى الحراري.

يعرض الجدول 4.4 الأوزان الخاصة بالشبكة العصبونية الاصطناعية. هذه الأوزان تعكس أهمية كل متغير في عملية التنبؤ. يلاحظ وجود تباين في قيم الأوزان، مما يشير إلى اختلاف تأثير كل متغير على النتيجة النهائية. من المهم تحليل هذه الأوزان وتفسيرها بناءً على طبيعة المتغيرات المختارة.

يعرض الشكلان 4.4 و5.4 العلاقة بين القيم التجريبية والمتنبئ بها لكل من مجموعتي التدريب والتحقق. يؤكد التوزيع الجيد للنقاط حول الخط المستقيم y=x على دقة النموذج. مع ذلك، يفضل إضافة مجال الثقة (confidence interval) على هذه الرسوم البيانية لتوضيح مدى دقة التنبؤات.

خاتمة

أثبتت الشبكة العصبية الاصطناعية (ANN) قدرتها على إنشاء علاقة أكثر دقة وملاءمة بين الواصفات الجزيئية وخصائص المركبات المدروسة مقارنةً بطريقة الانحدار الخطي المتعد (MLR). ومع ذلك، يُعد التعقيد الذي يصاحب نموذج ANN أحد أبرز التحديات، حيث يتميز هذا النموذج بصعوبة التفسير والتفكيك، مما يقلل من قابليته للفهم المباشر. وعلى النقيض، يُعتبر نموذج MLR أكثر مرونة وسهولة في التفسير، حيث يقدم شرحًا جيدًا ومباشرًا للعلاقة بين الواصفات الجزيئية والخصائص الكيميائية للمركبات. يسمح هذا النهج بتحديد وفهم الخصائص المدروسة للمركبات الجديدة عبر المقارنة مع المركبات السابقة المدروسة باستخدام MLR، مما يعزز من قدرته على التنبؤ بدقة وبساطة.

بالنظر إلى النتائج المُستخلصة، يمكن القول بأن النماذج المقترحة، سواء باستخدام ANN أو MLR، تقدم إسهامًا كبيرًا في تقليل الوقت والتكلفة المرتبطين بعملية التوليف الكيميائي وتحديد الخصائص الكيميائية للمركبات. يعتبر هذا أمرًا مهمًا في تطبيقات تصميم المركبات الكيميائية الجديدة، حيث تسهم هذه النماذج في تسريع العملية وتقليل الاعتماد على التجارب العملية المكلفة والمعقدة.

الخاتمـــة العامـــة

الخاتمة العامة

في خاتمة هذا البحث، تم استعراض وتحليل الأساليب المتقدمة التي تم استخدامها لتقدير المحتوى الحراري القياسي لتكوين المركبات العضوية، بما في ذلك تطبيقات الانحدار الخطي المتعدد والشبكات العصبية الاصطناعية. من خلال تحليل دقيق للواصفات الجزيئية وتقييمها باستخدام تقنيات مثل QSPR، أظهرت الدراسة قدرة هذه النماذج على توفير تقديرات دقيقة وموثوقة للمحتوى الحراري استنادًا إلى مجموعة متنوعة من البيانات المدخلة.

- **نموذج الشبكات العصبية الاصطناعية (ANN):** أظهر قدرة فائقة على التقاط العلاقات غير الخطية بين الواصفات الجزيئية والمحتوى الحراري، مما نتج عنه نموذج تنبؤي عالي الأداء. تشير قيم R^2 المرتفعة (0.9848 لكلا مجموعتي التدريب والتحقق من الصحة) وانخفاض قيم EQM إلى دقة النموذج وقدرته على التعميم. من جهة أخرى كان نموذجه معقدًا ويصعب تفسيره، مما يحد من فهم العلاقة بين الواصفات الجزيئية والمحتوى الحراري.

- **نموذج الانحدار الخطي المتعدد (MLR):** تميز ببساطته وسهولة تفسيره، مما يسمح بفهم واضح للعلاقة بين الواصفات الجزيئية والمحتوى الحراري. يُسهل هذا التفسير فهم العوامل البنيوية التي تؤثر على المحتوى الحراري. حقق نموذج MLR أيضًا أداءً جيدًا مع قيم R^2 مرتفعة (0.9962 لمجموعة التدريب و0.9602 لمجموعة التحقق من الصحة). رغم ذلك، مع افترض وجود علاقة خطية بين الواصفات الجزيئية والمحتوى الحراري، لم يكن دقيقًا تمامًا في بعض الحالات، خاصة مع المركبات ذات البنية المعقدة. كما ظهرت بعض الانحرافات في التنبؤات لبعض المركبات، مثل 1,8-Cyclotetradecadiyne وPentacyclo-[4.2.0.02,5.03,8.04,7]octane (Cubane)، مما يشير إلى محدودية النموذج في التعامل مع بعض الحالات الخاصة.

بشكل عام، تقدم هذه الدراسة نموذجين فعالين للتنبؤ بالمحتوى الحراري القياسي لتكوين المركبات العضوية. يمكن استخدام نموذج ANN عندما تكون الدقة العالية مطلوبة، بينما يُفضل استخدام نموذج MLR عندما تكون سهولة التفسير وفهم العلاقة بين البنية والخاصية أكثر أهمية. يُوصى بإجراء دراسات مستقبلية لمقارنة أداء هذين النموذجين مع أساليب أخرى وتوسيع نطاق تطبيقها على مجموعات بيانات أكبر لتحسين دقة التنبؤات.

قائمــة المراجــع

قائمة المراجع

[1] crossref, « Chooser ». Consulté le: 9 septembre 2024. [En ligne]. Disponible sur: https://chooser.crossref.org/

[2] B. Lian et S. H. Yalkowsky, « Unified Physicochemical Property Estimation Relationships (UPPER) », Consulté le: 9 septembre 2024. [En ligne]. Disponible sur: https://jpharmsci.org/article/S0022-3549(15)30443-3/fulltext

[3] « Thermodynamic modeling of the essential physicochemical interactions between the pore solution and the cement hydrates in chloride-contaminated cement-based materials - ScienceDirect ». Consulté le: 9 septembre 2024. [En ligne]. Disponible sur: https://www.sciencedirect.com/science/article/abs/pii/S0021979718307616?via%3Dihub

[4] E. Pereira, A. J. A. Meirelles, et G. J. Maximo, « Predictive models for physical properties of fats, oils, and biodiesel fuels », *Fluid Phase Equilibria*, vol. 508, p. 112440, mars 2020, doi: 10.1016/j.fluid.2019.112440.

[5] S. Chtita *et al.*, « QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods », *Chemom. Intell. Lab. Syst.*, vol. 210, p. 104266, mars 2021, doi: 10.1016/j.chemolab.2021.104266.

[6] X. Yan, T. Lan, Q. Jia, F. Yan, et Q. Wang, « A norm indexes-based QSPR model for predicting the standard vaporization enthalpy and formation enthalpy of organic compounds », *Fluid Phase Equilibria*, vol. 507, p. 112437, mars 2020, doi: 10.1016/j.fluid.2019.112437.

[7] P. C. S. Costa, J. S. Evangelista, I. Leal, et P. C. M. L. Miranda, « Chemical Graph Theory for Property Modeling in QSAR and QSPR—Charming QSAR & QSPR », *Mathematics*, vol. 9, nº 1, Art. nº 1, janv. 2021, doi: 10.3390/math9010060.

[8] E. S. Domalski et E. D. Hearing, « Estimation of the Thermodynamic Properties of C-H-N-O-S-Halogen Compounds at 298.15 K », *J. Phys. Chem. Ref. Data*, vol. 22, nº 4, p. 805-1159, juill. 1993, doi: 10.1063/1.555927.

[9] R. Nomen, M. Bartra, J. Sempere, E. Serra, J. Sales, et X. Romero, « The estimation of reaction enthalpy for complex molecules using Benson groups », *J. Therm. Anal. Calorim.*, 2005, doi: 10.1007/S10973-005-6929-8.

[10] S. Agarwal, « Engineering Chemistry: Fundamentals and Applications », Higher Education from Cambridge University Press. Consulté le: 8 septembre 2024. [En ligne]. Disponible sur: https://www.cambridge.org/highereducation/books/engineering-chemistry/188D48FC409A9845AC58DC8894A74B0B/fundamentals-of-organic-chemistry/BDD2030C626F72CF24E76B65F951D7A0

[11] J. G. Speight, « Chapter 5 - Properties of Organic Compounds », in *Environmental Organic Chemistry for Engineers*, J. G. Speight, Éd., Butterworth-Heinemann, 2017, p. 203-261. doi: 10.1016/B978-0-12-804492-6.00005-8.

[12] P. M. Dean, « THE NOMENCLATURE OF ORGANIC COMPOUNDS. », *Science*, vol. 63 1637, p. 501, 1926.

[13] L. J. Silverberg, « What is an organic substance? », *Found. Chem.*, vol. 23, nº 3, p. 329-336, oct. 2021, doi: 10.1007/s10698-021-09400-z.

[14] W. R. Fearon, « CHAPTER V - CLASSIFICATION AND CHARACTERISTICS OF ORGANIC COMPOUNDS », in *An Introduction to Biochemistry (Second Edition)*, W. R. Fearon, Éd., Butterworth-Heinemann, 2014, p. 72-75. doi: 10.1016/B978-1-4832-0036-1.50012-4.

[15] J. E. Rice, « Chapter 1 - Bonding in Organic Compounds », in *Organic Chemistry Concepts and Applications for Medicinal Chemistry*, J. E. Rice, Éd., Boston: Academic Press, 2014, p. 1-26. doi: 10.1016/B978-0-12-800739-6.00001-2.

[16] M. Abe et R. Akisaka, « Is π-Single Bonding (C–π–C) Possible? A Challenge in Organic Chemistry », *Chem. Lett.*, vol. 46, nº 11, p. 1586-1592, nov. 2017, doi: 10.1246/cl.170711.

[17] E. D. Głowacki, M. Irimia-Vladu, S. Bauer, et N. S. Sariciftci, « Hydrogen-bonds in molecular solids – from biological systems to organic electronics », *J. Mater. Chem. B*, vol. 1, nº 31, p. 3742-3753, juill. 2013, doi: 10.1039/C3TB20193G.

[18] O. Takahashi, Y. Kohno, et M. Nishio, « Relevance of Weak Hydrogen Bonds in the Conformation of Organic Compounds and Bioconjugates: Evidence from Recent Experimental Data and High-Level ab Initio MO Calculations », *Chem. Rev.*, vol. 110, nº 10, p. 6049-6076, oct. 2010, doi: 10.1021/cr100072x.

[19] M. Nishio, « The CH/π hydrogen bond in chemistry. Conformation, supramolecules, optical resolution and interactions involving carbohydrates », *Phys. Chem. Chem. Phys.*, vol. 13, nº 31, p. 13873-13900, juill. 2011, doi: 10.1039/C1CP20404A.

[20] Linus. Pauling, « THE NATURE OF THE CHEMICAL BOND. APPLICATION OF RESULTS OBTAINED FROM THE QUANTUM MECHANICS AND FROM A THEORY OF PARAMAGNETIC SUSCEPTIBILITY TO THE STRUCTURE OF MOLECULES », *J. Am. Chem. Soc.*, vol. 53, nº 4, p. 1367-1400, avr. 1931, doi: 10.1021/ja01355a027.

[21] « Coordination chemistry at carbon | Nature Chemistry ». Consulté le: 8 septembre 2024. [En ligne]. Disponible sur: https://www.nature.com/articles/nchem.248

[22] N. G. Schmidt, E. Eger, et W. Kroutil, « Building Bridges: Biocatalytic C-C-Bond Formation toward Multifunctional Products », *ACS Catal.*, vol. 6, nº 7, p. 4286-4311, juill. 2016, doi: 10.1021/acscatal.6b00758.

[23] S. Sultan et B. A. Shah, « Carbon-Carbon and Carbon-Heteroatom Bond Formation Reactions Using Unsaturated Carbon Compounds », *Chem. Rec.*, vol. 19, nº 2-3, p. 644-660, 2019, doi: 10.1002/tcr.201800095.

[24] M. Murakami et N. Ishida, « Potential of Metal-Catalyzed C–C Single Bond Cleavage for Organic Synthesis », *J. Am. Chem. Soc.*, vol. 138, nº 42, p. 13759-13769, oct. 2016, doi: 10.1021/jacs.6b01656.

[25] Y. Xia et G. Dong, « Temporary or removable directing groups enable activation of unstrained C–C bonds », *Nat. Rev. Chem.*, vol. 4, p. 600-614, nov. 2020, doi: 10.1038/s41570-020-0218-8.

[26] J. Hong, M. Li, J. Zhang, B. Sun, et F. Mo, « C−H Bond Carboxylation with Carbon Dioxide », *ChemSusChem*, vol. 12, nº 1, p. 6-39, 2019, doi: 10.1002/cssc.201802012.

[27] « A structural approach to calculate physical properties of pure organic substances: The critical temperature, critical volume and related properties - Grigoras - 1990 - Journal of Computational Chemistry - Wiley Online Library ». Consulté le: 9 septembre 2024. [En ligne]. Disponible sur: https://onlinelibrary.wiley.com/doi/10.1002/jcc.540110408

[28] « Single view », International Journal of Current Science Research and Review. Consulté le: 9 septembre 2024. [En ligne]. Disponible sur: https://ijcsrr.org/single-view/

[29] J. G. Speight, « Chapter 5 - Properties of Organic Compounds », in *Environmental Organic Chemistry for Engineers*, J. G. Speight, Éd., Butterworth-Heinemann, 2017, p. 203-261. doi: 10.1016/B978-0-12-804492-6.00005-8.

[30] « Single view », International Journal of Current Science Research and Review. Consulté le: 9 septembre 2024. [En ligne]. Disponible sur: https://ijcsrr.org/single-view/

[31] « Properties of Pure Organic Compounds », in *Studies in Physical and Theoretical Chemistry*, vol. 75, Elsevier, 1992, p. 131-574. doi: 10.1016/B978-0-444-89217-1.50006-5.

[32] « Edelweiss Publications-About us », Edelweiss Publications Inc. Consulté le: 9 septembre 2024. [En ligne]. Disponible sur: http://edelweisspublications.com/

[33] S. L. Khursan, A. S. Ismagilova, A. A. Akhmerov, et S. I. Spivak, « Constructing Homodesmic Reactions for Calculating the Enthalpies of Formation of Organic Compounds », *Russ. J. Phys. Chem. A*, vol. 90, nº 4, Art. nº 4, avr. 2016, doi: 10.1134/S0036024416030201.

[34] R. Nomen, M. Bartra, J. Sempere, E. Serra, J. Sales, et X. Romero, « The estimation of reaction enthalpy for complex molecules using Benson groups », *J. Therm. Anal. Calorim.*, vol. 82, nº 3, Art. nº 3, nov. 2005, doi: 10.1007/s10973-005-0937-6.

[35] « Relative energy of organic compounds III. Components of the heat of reactions | Structural Chemistry ». Consulté le: 10 septembre 2024. [En ligne]. Disponible sur: https://link.springer.com/article/10.1007/s11224-009-9467-3

[36] A. Akhmetyanova et A. Ismagilova, « Determination of the Standard Enthalpy of Formation for Organic Compound 3,3- Dimethyloxetane by the Homodesmotic Method », in *2023 5th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, nov. 2023, p. 622-625. doi: 10.1109/SUMMA60232.2023.10349572.

[37] A. A. Zavitsas, D. W. Rogers, et N. Matsunaga, « Heats of Formation of Organic Compounds by a Simple Calculation », *J. Org. Chem.*, vol. 75, nº 19, p. 6502-6515, oct. 2010, doi: 10.1021/jo1012898.

[38] B. Ruscic, « Active Thermochemical Tables: Sequential Bond Dissociation Enthalpies of Methane, Ethane, and Methanol and the Related Thermochemistry », *J. Phys. Chem. A*, vol. 119, nº 28, p. 7810-7837, juill. 2015, doi: 10.1021/acs.jpca.5b01346.

[39] A. N. Kizin, P. L. Dvorkin, G. L. Ryzhova, et Y. A. Lebedev, « Parameters for calculation of standard enthalpies of formation of organic compounds in the liquid state », *Bull. Acad. Sci. USSR Div. Chem. Sci.*, vol. 35, nº 2, Art. nº 2, févr. 1986, doi: 10.1007/BF00952920.

[40] S. Rayne et K. Forest, « Estimated Gas-Phase Standard State Enthalpies of Formation for Organic Compounds Using the Gaussian-4 (G4) and W1BD Theoretical Methods », *J. Chem. Eng. Data*, vol. 55, nº 11, p. 5359-5364, nov. 2010, doi: 10.1021/je100768s.

[41] F. Gharagheizi, « Prediction of the Standard Enthalpy of Formation of Pure Compounds Using Molecular Structure », Consulté le: 10 septembre 2024. [En ligne]. Disponible sur: https://www.publish.csiro.au/ch/CH08522

[42] K. Argoub, A. M. Benkouider, A. Yahiaoui, et F. Bagui, « Estimation and uncertainty analysis of standard enthalpy of formation in the liquid state by third-order-group-contribution method », *Fluid Phase Equilibria*, vol. 520, p. 112644, oct. 2020, doi: 10.1016/j.fluid.2020.112644.

[43] A. Salmon et D. Dalmazzone, « Prediction of Enthalpy of Formation in the Solid State (at 298.15K) using Second-Order Group Contributions. Part 1. Carbon-Hydrogen and Carbon-Hydrogen-Oxygen Compounds », *J. Phys. Chem. Ref. Data*, vol. 35, nº 3, p. 1443-1457, sept. 2006, doi: 10.1063/1.2203111.

[44] E. Paulechka et A. Kazakov, « Efficient DLPNO–CCSD(T)-Based Estimation of Formation Enthalpies for C-, H-, O-, and N-Containing Closed-Shell Compounds Validated Against Critically Evaluated Experimental Data », *J. Phys. Chem. A*, vol. 121, nº 22, p. 4379-4387, juin 2017, doi: 10.1021/acs.jpca.7b03195.

[45] Z. Wan, Q. Wang, et J. Liang, « Accurate prediction of standard enthalpy of formation based on semiempirical quantum chemistry methods with artificial neural network and molecular descriptors »,

Consulté le: 10 septembre 2024. [En ligne]. Disponible sur: https://www.authorea.com/users/330249/articles/457092-accurate-prediction-of-standard-enthalpy-of-formation-based-on-semiempirical-quantum-chemistry-methods-with-artificial-neural-network-and-molecular-descriptors?commit=e4b2feba1aee78496cee09cd8613fea081fb1ccd

[46] A. G. Mercader et E. A. Castro, « Partial-Order Ranking and Linear Modeling: Their Use in Predictive QSAR/QSPR Studies », in *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons, Ltd, 2012, p. 149-174. doi: 10.1002/9783527645121.ch5.

[47] A. A. Toropov, A. P. Toropova, E. Benfenati, et A. Manganaro, « QSPR modeling of enthalpies of formation for organometallic compounds by SMART-based optimal descriptors », *J. Comput. Chem.*, vol. 30, nº 15, p. 2576-2582, 2009, doi: 10.1002/jcc.21263.

[48] M. Bagheri, K. Yerramsetty, K. A. M. Gasem, et B. J. Neely, « Molecular modeling of the standard state heat of formation », *Energy Convers. Manag.*, vol. 65, p. 587-596, janv. 2013, doi: 10.1016/j.enconman.2012.07.014.

[49] W. J. Welsh, W. Tong, E. R. Collantes, J. S. Chickos, et S. G. Gagarin, « Enthalpies of sublimation and formation of polycyclic aromatic hydrocarbons (PAHs) derived from comparative molecular field analysis (CoMFA): Application of moment of inertia for molecular alignment », *Thermochim. Acta*, vol. 290, nº 1, p. 55-64, janv. 1997, doi: 10.1016/S0040-6031(96)03048-1.

[50] D. T. Stanton, « On the importance of topological descriptors in understanding structure–property relationships », *J. Comput. Aided Mol. Des.*, vol. 22, nº 6, Art. nº 6, mars 2008, doi: 10.1007/s10822-008-9204-9.

[51] « Many-Body Descriptors for Predicting Molecular Properties with Machine Learning: Analysis of Pairwise and Three-Body Interactions in Molecules | Journal of Chemical Theory and Computation ». Consulté le: 11 septembre 2024. [En ligne]. Disponible sur: https://pubs.acs.org/doi/10.1021/acs.jctc.8b00110

[52] W. Pronobis, A. Tkatchenko, et K.-R. Müller, « Many-Body Descriptors for Predicting Molecular Properties with Machine Learning: Analysis of Pairwise and Three-Body Interactions in Molecules », *J. Chem. Theory Comput.*, vol. 14, nº 6, p. 2991-3003, juin 2018, doi: 10.1021/acs.jctc.8b00110.

[53] M. Karelson, V. S. Lobanov, et A. R. Katritzky, « Quantum-Chemical Descriptors in QSAR/QSPR Studies », *Chem. Rev.*, vol. 96, nº 3, p. 1027-1044, janv. 1996, doi: 10.1021/cr950202r.

[54] I. Paster, M. Shacham, et N. Brauner, « Investigation of the Relationships between Molecular Structure, Molecular Descriptors, and Physical Properties », *Ind. Eng. Chem. Res.*, vol. 48, nº 21, p. 9723-9734, nov. 2009, doi: 10.1021/ie801318y.

[55] A. Vatani, M. Mehrpooya, et F. Gharagheizi, « Prediction of Standard Enthalpy of Formation by a QSPR Model », *Int. J. Mol. Sci.*, vol. 8, nº 5, p. 407-432, mai 2007, doi: 10.3390/i8050407.

[56] A. L. Teixeira, J. P. Leal, et A. O. Falcao, « Random forests for feature selection in QSPR Models - an application for predicting standard enthalpy of formation of hydrocarbons », *J. Cheminformatics*, vol. 5, nº 1, p. 9, févr. 2013, doi: 10.1186/1758-2946-5-9.

[57] J. Jover, R. Bosque, J. A. Martinho Simões, et J. Sales, « Estimation of enthalpies of formation of organometallic compounds from their molecular structures », *J. Organomet. Chem.*, vol. 693, nº 7, p. 1261-1268, avr. 2008, doi: 10.1016/j.jorganchem.2008.01.021.

[58] L. C. Yee et Y. C. Wei, « Current Modeling Methods Used in QSAR/QSPR », in *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons, Ltd, 2012, p. 1-31. doi: 10.1002/9783527645121.ch1.

[59] H. Hong *et al.*, « Mold2 Molecular Descriptors for QSAR », in *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, John Wiley & Sons, Ltd, 2012, p. 65-109. doi: 10.1002/9783527645121.ch3.

[60] X. Kang, J. Qian, J. Deng, U. Latif, et Y. Zhao, « Novel molecular descriptors for prediction of H2S solubility in ionic liquids », *J. Mol. Liq.*, vol. 265, p. 756-764, sept. 2018, doi: 10.1016/j.molliq.2018.06.113.

[61] A. Tropsha, P. Gramatica, et V. K. Gombar, « The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models », *QSAR Comb. Sci.*, vol. 22, nº 1, p. 69-77, 2003, doi: 10.1002/qsar.200390007.

[62] S. Ahmadi, S. Ketabi, et M. Jebeli Javan, « Molecular Descriptors in QSPR/QSAR Modeling », in *QSPR/QSAR Analysis Using SMILES and Quasi-SMILES*, A. P. Toropova et A. A. Toropov, Éd., Cham: Springer International Publishing, 2023, p. 25-56. doi: 10.1007/978-3-031-28401-4_2.

[63] « Statistical Methods in QSAR/QSPR | SpringerLink ». Consulté le: 22 octobre 2024. [En ligne]. Disponible sur: https://link.springer.com/chapter/10.1007/978-3-319-17281-1_2

[64] H. Modarresi, H. Modarress, et J. C. Dearden, « QSPR model of Henry's law constant for a diverse set of organic chemicals based on genetic algorithm-radial basis function network approach », *Chemosphere*, vol. 66, nº 11, p. 2067-2076, févr. 2007, doi: 10.1016/j.chemosphere.2006.09.049.

[65] A. Rybinska, A. Sosnowska, M. Barycki, et T. Puzyn, « Geometry optimization method versus predictive ability in QSPR modeling for ionic liquids », *J. Comput. Aided Mol. Des.*, vol. 30, nº 2, Art. nº 2, févr. 2016, doi: 10.1007/s10822-016-9894-3.

[66] « Chtita: Modélisation de molécules organiques hétérocy... - Google Scholar ». Consulté le: 20 septembre 2024. [En ligne]. Disponible sur: https://scholar.google.com/scholar_lookup?title=Mod%C3%A9lisation%20de%20mol%C3%A9cules%20organiques%20h%C3%A9t%C3%A9rocycliques%20biologiquement%20actives%20par%20des%20m%C3%A9thodes%20QSARQSPR%20-%20Recherche%20de%20nouveaux%20medicaments&publication_year=2017&author=S.%20Chtita

[67] G. Mukherjee, A. Braka, et S. Wu, « Quantifying Functional-Group-like Structural Fragments in Molecules and Its Applications in Drug Design », *J. Chem. Inf. Model.*, vol. 63, nº 7, p. 2073-2083, avr. 2023, doi: 10.1021/acs.jcim.3c00050.

[68] M. Karelson, V. S. Lobanov, et A. R. Katritzky, « Quantum-Chemical Descriptors in QSAR/QSPR Studies », *Chem. Rev.*, vol. 96, nº 3, p. 1027-1044, mai 1996, doi: 10.1021/cr950202r.

[69] D. Thusius, « Simultaneous determination of equilibrium constants and thermodynamic functions by means of relaxation amplitude measurements », *Biochimie*, vol. 55, nº 3, p. 277-282, mai 1973, doi: 10.1016/S0300-9084(73)80126-9.

[70] « Nonlinear regression methods for simultaneous property measurement | AIAA Journal ». Consulté le: 24 septembre 2024. [En ligne]. Disponible sur: https://arc.aiaa.org/doi/10.2514/3.5829

[71] F. Luan, H. T. Liu, Y. Gao, et X. Y. Zhang, « QSPR model to predict the thermal stabilities of second-order nonlinear optical (NLO) chromophore molecules », *Mol. Simul.*, vol. 35, nº 3, p. 248-257, mars 2009, doi: 10.1080/08927020802378928.

[72] *Nonlinear Computer Modeling of Chemical and Biochemical Data*. 1996. Consulté le: 24 septembre 2024. [En ligne]. Disponible sur: https://shop.elsevier.com/books/nonlinear-computer-modeling-of-chemical-and-biochemical-data/rusling/978-0-12-604490-4

[73] J. L. Haas et J. R. Fisher, « Simultaneous evaluation and correlation of thermodynamic data », *Am. J. Sci.*, vol. 276, nº 4, p. 525-545, avr. 1976, doi: 10.2475/ajs.276.4.525.

[74] R. D. Snee, « Validation of Regression Models: Methods and Examples », *Technometrics*, vol. 19, nº 4, p. 415-428, nov. 1977, doi: 10.1080/00401706.1977.10489581.

[75] M. W. Browne, « Cross-Validation Methods », *J. Math. Psychol.*, vol. 44, nº 1, p. 108-132, mars 2000, doi: 10.1006/jmps.1999.1279.

[76] I. Koo, N. Lee, et R. M. Kil, « Parameterized cross-validation for nonlinear regression models », *Neurocomputing*, vol. 71, nº 16, p. 3089-3095, oct. 2008, doi: 10.1016/j.neucom.2008.04.043.

[77] H. Kaneko et K. Funatsu, « Criterion for Evaluating the Predictive Ability of Nonlinear Regression Models without Cross-Validation », *J. Chem. Inf. Model.*, vol. 53, nº 9, p. 2341-2348, sept. 2013, doi: 10.1021/ci4003766.

[78] G. Cano *et al.*, « Automatic selection of molecular descriptors using random forest: Application to drug discovery », *Expert Syst. Appl.*, vol. 72, p. 151-159, avr. 2017, doi: 10.1016/j.eswa.2016.12.008.

[79] « Replacement Method and Enhanced Replacement Method Versus the Genetic Algorithm Approach for the Selection of Molecular Descriptors in QSPR/QSAR Theories | Journal of Chemical Information and Modeling ». Consulté le: 24 septembre 2024. [En ligne]. Disponible sur: https://pubs.acs.org/doi/10.1021/ci100103r

[80] « Classification and regression tree analysis for molecular descriptor selection and retention prediction in chromatographic quantitative structure–retention relationship studies - ScienceDirect ». Consulté le: 24 septembre 2024. [En ligne]. Disponible sur: https://www.sciencedirect.com/science/article/abs/pii/S0021967303000049?via%3Dihub

[81] « How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space | Journal of Chemical Information and Modeling ». Consulté le: 24 septembre 2024. [En ligne]. Disponible sur: https://pubs.acs.org/doi/10.1021/ci800249s

[82] « Effect of Molecular Descriptor Feature Selection in Support Vector Machine Classification of Pharmacokinetic and Toxicological Properties of Chemical Agents | Journal of Chemical Information and Modeling ». Consulté le: 24 septembre 2024. [En ligne]. Disponible sur: https://pubs.acs.org/doi/10.1021/ci049869h

[83] R. Mercado, S. M. Kearnes, et C. W. Coley, « Data Sharing in Chemistry: Lessons Learned and a Case for Mandating Structured Reaction Data », *J. Chem. Inf. Model.*, vol. 63, nº 14, p. 4253-4265, juill. 2023, doi: 10.1021/acs.jcim.3c00607.

[84] J. Galgonek et J. Vondrášek, « A comparison of approaches to accessing existing biological and chemical relational databases via SPARQL », *J. Cheminformatics*, vol. 15, nº 1, p. 61, juin 2023, doi: 10.1186/s13321-023-00729-5.

[85] T. J. O'Donnell, *Design and Use of Relational Databases in Chemistry*. Boca Raton: CRC Press, 2008. doi: 10.1201/9781420064438.

[86] N. O. of D. and Informatics, « WebBook de Chimie NIST ». Consulté le: 1 novembre 2024. [En ligne]. Disponible sur: https://webbook.nist.gov/chemistry/

[87] A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, et G. Schneider, « Generative Recurrent Networks for De Novo Drug Design », *Mol. Inform.*, vol. 37, nº 1-2, p. 1700111, 2018, doi: 10.1002/minf.201700111.

[88] A. Cherkasov *et al.*, « QSAR Modeling: Where Have You Been? Where Are You Going To? », *J. Med. Chem.*, vol. 57, nº 12, p. 4977-5010, juin 2014, doi: 10.1021/jm4004285.

[89] C. B. Santiago, J.-Y. Guo, et M. S. Sigman, « Predictive and mechanistic multivariate linear regression models for reaction development », doi: 10.1039/C7SC04679K.

[90] A. Golbraikh et A. Tropsha, « Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection », *J. Comput. Aided Mol. Des.*, vol. 16, nº 5, Art. nº 5, mai 2002, doi: 10.1023/A:1020869118689.

[91] H. Chun et S. Keleş, « Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection », *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 72, nº 1, p. 3-25, janv. 2010, doi: 10.1111/j.1467-9868.2009.00723.x.

[92] R. Garg, « Ridge Regression in the Presence of Multicollinearity », *Psychol. Rep.*, vol. 54, nº 2, p. 559-566, avr. 1984, doi: 10.2466/pr0.1984.54.2.559.

[93] G. Liang, J. Xu, et L. Liu, « QSPR analysis for melting point of fatty acids using genetic algorithm based multiple linear regression (GA-MLR) », *Fluid Phase Equilibria*, vol. 353, p. 15-21, sept. 2013, doi: 10.1016/j.fluid.2013.06.008.

[94] D. L. J. Alexander, A. Tropsha, et D. A. Winkler, « Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models », *J. Chem. Inf. Model.*, vol. 55, nº 7, p. 1316-1322, juill. 2015, doi: 10.1021/acs.jcim.5b00206.

[95] K. Roy et P. Pratim Roy, « Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques », *Eur. J. Med. Chem.*, vol. 44, nº 7, p. 2913-2922, juill. 2009, doi: 10.1016/j.ejmech.2008.12.004.

[96] G. B. Goh, N. O. Hodas, et A. Vishnu, « Deep learning for computational chemistry », *J. Comput. Chem.*, vol. 38, nº 16, p. 1291-1307, 2017, doi: 10.1002/jcc.24764.

[97] M. Tanaka, « Weighted sigmoid gate unit for an activation function of deep neural network », *Pattern Recognit. Lett.*, vol. 135, p. 354-359, juill. 2020, doi: 10.1016/j.patrec.2020.05.017.

[98] S. Elfwing, E. Uchibe, et K. Doya, « Sigmoid-weighted linear units for neural network function approximation in reinforcement learning », *Neural Netw.*, vol. 107, p. 3-11, nov. 2018, doi: 10.1016/j.neunet.2017.12.012.

[99] S. Sveleba, I. Katerynchuk, I. Kunyo, O. Semotyjuk, S. Velgosh, et V. Brygilevych, « The Application of Optimization Learning Methods for Multilayered Neural Networks », in *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)*, sept. 2023, p. 525-528. doi: 10.1109/ACIT58437.2023.10275354.

[100] R. Rawat, J. K. Patel, et M. T. Manry, « Minimizing validation error with respect to network size and number of training epochs », in *The 2013 International Joint Conference on Neural Networks (IJCNN)*, août 2013, p. 1-7. doi: 10.1109/IJCNN.2013.6706919.

[101] D. Chicco, M. J. Warrens, et G. Jurman, « The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation », *PeerJ Comput. Sci.*, vol. 7, p. e623, juill. 2021, doi: 10.7717/peerj-cs.623.

[102] S. Ajala, H. M. M. Jalajamony, et R. E. Fernandez, « A CNN Regression Approach for Real-Time Estimation of Dielectrophoretic Force », *ECS Meet. Abstr.*, vol. MA2022-02, nº 62, p. 2290, oct. 2022, doi: 10.1149/MA2022-02622290mtgabs.

[103] D. K. Singh et N. Rawat, « Decoding Optimization Algorithms for Convolutional Neural Networks in Time Series Regression Tasks », *Int. J. Inf. Technol. Comput. Sci.*, vol. 15, nº 6, p. 37.

[104] L. Prechelt, « Automatic early stopping using cross validation: quantifying the criteria », *Neural Netw.*, vol. 11, nº 4, p. 761-767, juin 1998, doi: 10.1016/S0893-6080(98)00010-0.

[105] Z. Zhang, « Improved Adam Optimizer for Deep Neural Networks », in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, juin 2018, p. 1-2. doi: 10.1109/IWQoS.2018.8624183.

[106] E. Oostwal, M. Straat, et M. Biehl, « Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation », *Phys. Stat. Mech. Its Appl.*, vol. 564, p. 125517, févr. 2021, doi: 10.1016/j.physa.2020.125517.

[107] M. Kaloev et G. Krastev, « Comparative Analysis of Activation Functions Used in the Hidden Layers of Deep Neural Networks », in *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, juin 2021, p. 1-5. doi: 10.1109/HORA52670.2021.9461312.

[108] M. Tanaka, « Weighted sigmoid gate unit for an activation function of deep neural network », *Pattern Recognit. Lett.*, vol. 135, p. 354-359, juill. 2020, doi: 10.1016/j.patrec.2020.05.017.

[109] D. A. Winkler et T. C. Le, « Performance of Deep and Shallow Neural Networks, the Universal Approximation Theorem, Activity Cliffs, and QSAR », *Mol. Inform.*, vol. 36, nº 1-2, p. 1600118, 2017, doi: 10.1002/minf.201600118.

[110] R. Tabaraki, T. Khayamian, et A. A. Ensafi, « Wavelet neural network modeling in QSPR for prediction of solubility of 25 anthraquinone dyes at different temperatures and pressures in supercritical carbon dioxide », *J. Mol. Graph. Model.*, vol. 25, nº 1, p. 46-54, sept. 2006, doi: 10.1016/j.jmgm.2005.10.012.

[111] S. Amari, « Backpropagation and stochastic gradient descent method », *Neurocomputing*, vol. 5, nº 4, p. 185-196, juin 1993, doi: 10.1016/0925-2312(93)90006-O.

[112] M. Z. Rehman et N. M. Nawi, « The Effect of Adaptive Momentum in Improving the Accuracy of Gradient Descent Back Propagation Algorithm on Classification Problems », in *Software Engineering and Computer Systems*, J. Mohamad Zain, W. M. bt Wan Mohd, et E. El-Qawasmeh, Éd., Berlin, Heidelberg: Springer, 2011, p. 380-390. doi: 10.1007/978-3-642-22170-5_33.

[113] « Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 4th Edition | Wiley », Wiley.com. Consulté le: 27 septembre 2024. [En ligne]. Disponible sur: https://www.wiley.com/en-us/Response+Surface+Methodology%3A+Process+and+Product+Optimization+Using+Designed+Experiments%2C+4th+Edition-p-9781118916032

[114] D. C. Montgomery, E. A. Peck, et G. G. Vining, « Introduction to linear regression analysis; 2012; Hoboken ». Wiley: zbMath ID. Consulté le: 27 septembre 2024. [En ligne]. Disponible sur: https://scholar.google.com/scholar?cluster=18315925468310579698&hl=en&oi=scholarr

[115] « Theoretical Calculations of Carbon−Oxygen Bond Dissociation Enthalpies of Peroxyl Radicals Formed in the Autoxidation of Lipids | Journal of the American Chemical Society ». Consulté le: 28 septembre 2024. [En ligne]. Disponible sur: https://pubs.acs.org/doi/10.1021/ja034182j

[116] M. R. Ibrahim, Z. A. Fataftah, P. von Ragué Schleyer, et P. D. Stout, « Calculation of enthalpies of hydrogenation of hydrocarbons », *J. Comput. Chem.*, vol. 8, nº 8, p. 1131-1138, 1987, doi: 10.1002/jcc.540080810.

[117] J. J. Faraway, « Regression for non-Euclidean data using distance matrices », *J. Appl. Stat.*, vol. 41, nº 11, p. 2342-2357, nov. 2014, doi: 10.1080/02664763.2014.909794.

[118] N. R. Draper, « Applied regression analysis bibliography update 1994-97 », *Commun. Stat. - Theory Methods*, vol. 27, nº 10, p. 2581-2623, janv. 1998, doi: 10.1080/03610929808832244.