الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة الشهيد حمـه لخضـر الوادي
كلية التكنولوجيـــــا

مذكـرة تخرج لنيل شهـادة
ماستـــــر أكاديمـــي

ميدان: العلوم والتكنولوجيا
شعبة : هندسة الطرائق
تخصص: هندسة كيميائية

من إعداد الطالبات:
- انفال فقيري
- خديجة قعري
- خولة حويذق

Prédiction de la toxicité des phénols et thiophenoles en utilisant l'analyse QSAR avec les outils statistiques MLR et ANN

نوقشت يوم: 2021/06/22

لجنة المناقشة:

د. بلغيث يزيد	أستاذ بجامعة الشهيد حمه لخضر بالوادي	رئيسا
د. بلقاسم سويعي	أستاذ بجامعة الشهيد حمه لخضر بالوادي	مقررا
د. منصر سهيلة	أستاذ بجامعة الشهيد حمه لخضر بالوادي	مناقشا

السنة الجامعية: 2021/2020

اهداء

اهدي ثمرة جهدي الى ملائكتي في الحياة...الى مصدر الحب و العطاء...الى أمني وطمأنينتي ...الى من كان دعائهما سر نجاحي وحنانهم بلسم جراحي ***امي (جديدي مبروكة) و ابي (عبد الحفيظ)*** .

الى من تقاسموا معي الحياة بحلوها و مرها و عاشوا حياتي بكل تفاصيلها- فخري و عزتي اخوتي... ***احمد*** (سندي واستقامة ظهري بعد ابي)، ***عواطف***(سري و ملجئي)، ***نوفل، معاذ*** (وفقهم الله و حفضهم من كل سوء) .

الى جدتي (خديجة) حفظها الله وبارك في عمرها ... الى روح جدتي (وريدة) التي مازالت في قلب حفيدتها لم تمت ...الى روح جداي (المولدي جديدي) و (السعيد فقيري) رحمهما الله و اسكنهم فسيح جنانه .

الى منبر العلم الذي افخر به .... الذي تميز بالوفاء و العطاء و تشجيعنا طيلة المسيرة الدراسية ....الى خالي ***جديدي العروسي(سيدي)***

الى رفيقة عمري و توأم روحي ولأنها نادرة اسميتها ذات يوم اختي ***(اية)***...الى تلك البعيدة عن عيني و القريبة من قلبي***(صفاء)***... الى العزيزة على قلبي ***(خلود)***...الى كل **عائلتي** الكبيرة

انفال

اهداء

أهدي ثمرة جهدي إلى التي من بحر حبها أغرقتني ومن نبع حنانها سقتني ،إلى التي في أعلى درجات الانام رأتني ونزعت من طيب خاطرها وأسعدتني، ربتني علمتني أكرمتني امي **(نزيهة فقيري)**

إلى من فارق الراحة لا جلي ولمكارم الاخلاق علمني إلى من القلب يهواه والروح فداه والعين ترتاح لرؤيته أبي**(ميداني)**

إلى ملائكتي في الحياة، إلى معنى الحب و الحنان، الى النور الذى سينير الدرب، إلى من بعثهم الله ليكونوا بلسم الحياتي الى قرة عيني اولادي **(العربي)** و **(ميرال)**

الى من سكب الحب وزرع الورد، الى من كان عونا لي و لاولادي، الى سند ظهري زوجي ***(احمد حمتين)***

إلى من جمعتهم معي ظلمة الرحم، إلى من شاركتهم وشاركونى سنوات الطفولة فخري وعزتى إخوتى وأخواتى: ***(فاطمة الزهراء)، (سليم)، (نورالهدى)، (عبد المؤمن)، (احمد)***

إلى من جمعتها معي ذكريات جميلة، تعلمت منها الكثير، صديقتى العزيزة ***(حنان)***

و الى كل من مد يد العون و ساعدنا من قريب او بعيد و لن ننسلى فضل الجميع على المجهودات و جعلها الله في ميزان حسناتهم و سدد خطاهم

خولة

اهداء

أهدي ثمرة جهدي هذه الى من وضعتني على طريق الحياة الى من بها اكبر وعليها اعتمد الى شمعة تنير ظلمة حياتي الى من كان دعائها سر نجاحي وحنانها بلسم جراحي الى من ارضعتني الحب والحنان امي الحبيبة ***(مريم شيحاني)***

الى صاحب السيرة العطرة والفكرة المستنيرة الى من كلله الله بالهيبة والوقار الى من علمني العطاء بدون انتظار الى من احمل اسمه بكل افتخار فلقد كان له الفضل الاول في بلوغي التعليم العالي والدي العزيز ***(جمال قعري)***

الى توأم روحي ورفيق دربي الى صاحب القلب الطيب والنوايا الصادقة الى من رافقني منذ ان حملنا حقائب صغيرة ومعك سرت الدرب خطوة بخطوة وما تزال ترافقني حتى الان الى من اعتمد عليه في كل كبيرة وصغيرة اخي سندي ومسندي ***(قعري احمد ياسين)***

الى القلوب الطاهرة الرقيقة والنفوس البريئة الى رياحين حياتي إلى من شاركوني الدمعة والابتسامة إلى من كانوا متكئا أستند عليه وخير معيل ألجأ إليه إلى من لم يبخلوا علي بالنصح والإرشاد اخوتي دنيا(سري وملجئي)....(ريحانة ، اماني ، ريان) وفقهم الله

الى الروح التي سكنت روحي زوجي **(سيف مصباحي)**

إلى ملاكي في الحياة، الى النور الذى سينير دربي، إلى من بعثه الله ليكون بلسم حياتي و قرة عيني ابني حبيبي ***(محمد الاسكندر)***حفظه الله ورعاه

خديجة

شكر وعرفان

الحمد لله و الصلاة و السلام على سيدنا محمد المصطفى الامين و بحمد الله حمدا جزيلا الذي وفقنا في مشوار دراستنا

نتقدم بالشكر الجزيل المليء بالاحترام و التقدير الى كل اساتذتنا بكلية العلوم والتكنولوجيا (جامعة حمه لخضر)

يسرنا ان نتقدم بالشكر الى الاستاذ الفاضل سويعي بالقاسم الذي اشرف على انجاز هذا العمل فجزاه الله عن الخير كله و له منا الشكر و التقدير على مجهوده و نصائحه القيمة و توجيهاته السديدة التي كانت عونا لنا في اتمام هذا العمل

وكذلك اشكر كل من ساعدنا و قدم لنا العون و زرع التفاؤل في دربنا ومد لنا يد العون و زودننا بالمعلومات اللازمة لإتمام هذا العمل

الملخص:

يستهدف عملنا التنبؤ بسمية مجموعة من المركبات الكيميائية من عائلة **الفينولات و الثيوفينولات** ، حيث تعتمد هذه الدراسة اساسا على النمذجة الجزيئية.

و يرتكز هذا الهدف على امكانية ايجاد العلاقة بين البنية الجزيئية والخاصية الفيزيوكيميائية (**بنية/سمية**) ، و تمت هذه الدراسة على **51** مركب كيميائي بتطبيق تقنية **QSPR** وذلك باستعمال مصفوفات تعتمد على الواصفات الجزيئية و **EPC50**، شملت هذه الدراسة 5 كتل من الواصفات الجزيئية (2D) وهذا وفق تقنيتين، الانحدار الخطي المتعدد (**MLR**) وحسابات الشبكة العصبية الاصطناعية (**ANN**) وهي من طرق النمذجة الاكثر شيوعا في **QSPR** وهذا باستعمال البرمجية (**Xlsta**) و (**MATlab**) على التوالي. وكانت أحسن نتائج في تقنية (MLR) هي نتائج كتلة (الواصفات الطوبولوجية) R^2=0.907 و MSE=0.115 و كانت نتائج ال(ANN) لنفس الكتلة جيدة حيث ان R لكل من Training و Validation و Teste اكبر من **0.95** .

الكلمات المفتاحية : النمذجة الجزيئية ، EPC50، الواصفات الجزيئية، QSPR، الانحدار(MLR)،الانحدار (ANN).

قائمة المحتويات

الملخص	
قائمة الاختصارات	
قائمة الرموز	
قائمة الأشكال	
قائمة الجداول	
العنوان	**الصفحة**
مقدمة عامة	2
الجزء النظري	
الفصل الأول: دراسة بيبليوغرافية	
I- دراسة بيبليوغرافية	06
مقدمة	06
I-1- عموميات حول المركبات الفينولية و الثيوفينولية	06
I-1-1- المركبات الفينولية	06
I-1-2- المركبات الثيوفينولية	07
I-2- الخصائص الفيزيوكيميائية	08
I-2-1- خاصية السمية	09
I-3- تعريف البنية الجزيئية للجزيء	10
I-4- تعريف النمذجة الجزيئية	12
I-4-1- اسباب استخدام النمذجة	12
I-4-2- طرق النمذجة الجزيئية	13
I-4-2-1- الميكانيكا الجزيئية	13
I-4-2-1-1-انواع طرق الميكانيكا الجزيئية	14
I-4-2-1-2- تطبيق حساب الديناميكا الجزيئية	15
I-4-2-2- طريقة الحل الشاملAb -initio Method	16
I-4-2-2-1- أنواع طرقAb-Initio	17
I-4-2-3- الطرق الشبه عمليةSemi- Empirical Methods	18
I-4-2-3-1-أنواع طرق الشبه عمليةSemi- Empirical Methods	18
I-4-2-4- نظرية الدالة الوظيفية للكثافة:Density Functional Theory (DFT)	20
I-4-2-4-1- انواع نظرية الدالة الوظيفية للكثافة	21
I-5- تعريف عملية التحسين	22
I-6- تعريف الواصف الـجزيئي	23
الفصل الثاني : الطرق والبرامج المستعملة	
II- الطرق و البرامج المستعملة	29
مقدمة	29
II-1-نبذة تاريخية عن QSAR/QSPR	29
II-1-1- تعريف QSPR	30

II-1-2- مبدا QSPR	31
II-2- تقنيات الانحدار الخطي المستعملة للحسابات الاحصائية	32
II-2-1-MLR	33
II-2-1-1- صيغة حساب الانحدار الخطي المتعدد	33
II-2-1-2- معامل الارتباط الخطي	36
II-2-1-3- نوع العلاقة	36
II-2-1-4- قوة العلاقة	37
II-2-1-5- شكل الانتشار	38
II-2-1-6- قياس الارتباط	38
II-2-1-7- معامل التحديد (R^2)	38
II-2-2-ANN	40
II-2-2-1- الهيكل العام للشبكة	41
II-2-2-2- صيغة حساب الشبكات العصبية الاصطناعية	41
II-3- النموذج الإحصائي	42
II-4-البرامج المستعملة	42
II-4-1- Chemdraw	42
II-4-2- HyperChem	43
II-4-3- DRAGON	44
II-4-4- برنامج (XLSTAT)	45
II-4-5- برنامج الماتلاب (Matlab)	46
خاتمة	48
الجزء العملي	
الفصل الثالث : قاعدة البيانات مع تحليل و مناقشة النتائج	
III-1- الخطوات التجريبية للعمل	51
III-2- شرح الخطوات التجريبية للعمل	54
III-3- النتائج	57
مقدمة	57
III-3-1-نتائج التقنية الاولى	57
III -3-1-1-كتلة الواصفات الطوبولوجية (Topological descriptors)	59
III-3-1-2-تحليل و مناقشة نتائج التقنية الاولى	65
III-3-2-نتائج الطريقة الثانية	66
III-3-2-1-كتلة الواصفات الطوبولوجية (Topological descriptors)	66
III-3-2-2-تحليل و مناقشة نتائج الطريقة الثانية	69

الرقم	العنوان	الصفحة
الجزء النظري		
الفصل الاول: حوصلة بيبليوغرافية		
الجدول (I-1)	درجات السمية	09
الجدول (I -2)	الهياكل البنيوية لبعض المركبات المستعملة	11
الجدول (I-3)	اهم كتل الواصفات الجزيئية	26
الفصل الثاني:الطرق والبرامج المستعملة		
الجدول (II-1)	درجات قوة العلاقة	37
الجزء العملي		
الفصل الثالث: قاعدة البيانات		
الجدول (III-1)	يوضح الجدول التالي قائمة قاعدة البيانات	52
الجدول (III-2)	نتائج حساب الواصفات الجزيئية المكونة لقاعدة البيانات باستعمالDragon	55
الجدول (III -3)	نتائج الكتل المدروسة	58
الجدول (**III**-4)	معاملات التعديل الخاصة بالكتلة (الواصفات الطوبولوجية)	60
جدول (III-5)	تحليل الفروق الخاصة بالكتلة	60
جدول(III-6)	نتائج الانحدار لكتلة(الواصفات الطوبولوجية)	67

الرقم	العنوان	الصفحة
الجزء النظري		
الفصل الأول: حوصلة بيبليوغرافية		
الشكل (I -1)	ارتباطات الواصفات الجزيئية في البحث العلمي	25
الفصل الثاني: الطرق و البرامج المستعملة		
الشكل (II-1)	مبدا طريقة QSPR	32
الشكل(II-2)	خط الانحدار الخطي	33
الشكل (II-3)	وحدة تشغيل داخل الشبكة العصبية	40
الشكل (II-4)	برنامج ChemDraw	43
الشكل (II-5)	برنامج Hyperchem	44
الشكل(II-6)	برنامج Dragon	45
الشكل(II-7)	برنامج Xlstat	46
الشكل (II-8)	برنامج Matlab	47
الجزء العملي		
الفصل الثالث: قاعدة البيانات مع تحليل و مناقشة النتائج		
الشكل (III-1)	الخطوات التجريبية	51
الشكل (III-2)	بيان يوضح اسماء معاملات الواصفات الجزيئية الخاصة بنموذج كتلة (الواصفات الطوبولوجية)	61
الشكل (III-3)	بيان يوضح تغيرات (Résidus normalises) بدلالة (Y) بكتلة (الواصفات الطوبولوجية)	62
الشكل (III-4)	بيان يوضح تغيرات (Résidus normalises) بدلالة (Y) الخاصة بكتلة (الواصفات الطوبولوجية)	62
الشكل (III-5)	بيان يوضح تغيرات (Y) بدلالة(Y)Préde الخاصة بكتلة (الواصفات الطوبولوجية)	63
الشكل (III-6)	بيان يوضح تغيرات بدلالة(Résidus normalises) بكتلة(الواصفات الطوبولوجية)	64
الشكل (III-7)	لقطة شاشة لتدريب نموذج الشبكة العصبية بواسطة Matlab لكتلة (الواصفات الطوبولوجية)	66
الشكل(III-8)	يوضح مخطط الاداء لكتلة (الواصفات الطوبولوجية)	67
الشكل(III-9)	بيان يوضح الخطأ (Errors)	68
الشكل(III-10)	بيانات توضح الانحدار R لكل من (Trining) و (Validation)و(test)	68

الرمز	المعنى
Ψ	الدالة الموجية
H	المعامل الهاميلتوني
2D	ثنائي البعد
3D	ثلاثي البعد
R^2	معامل التحديد(مربع معامل الارتباط الخطي)
R^2 ajusté	معامل التحديد المضبوط

الاختصار	المعنى	الكتابة بالعربية
AMBER	Assisted Model Building with Energy Refinement	وهي تقنية إعادة الطاقة
CHARMM	Chemistry at HARvard Macromolecular Mechanics	هي طريقة حسابية تستخدم مصطلحات التكافؤ
OPLS	optimized potential for liquid simulations	هي إمكانية محسنة لمحاكاة السوائل
MMFF	Merck Molecular Force Field	قوة ميرك الجزيئية
UFF	Universal Force Field	تعني القوة العالمية
Dreiding	The Dreiding force field	قوة الجزيء العضوي أو الحيوي
MM^+	Moléculair Mechanics	ميكانيكا الجزيئية
DM	Dynamique Moléculaire	الديناميكا الجزيئية
τ	Temps de relaxation	وقت الاسترخاء
T_0	Température de référence	درجة الحرارة المرجعية
T	Température de systéme	T درجة حرارة النظام
E	The energy, eigenvalue of the Hamiltonian operator	الطاقة قيمة خاصة للمعامل الهاميلتوني
GVB	The Generalized Valence Bond method	رابطة التكافؤ المعممة
CNDO	Complete Neglect of Differential Overlap	الاهمال الكامل للتداخل التفاضلي
INDO	Intermediate Neglect of Differential Overlap	الاهمال الوسيطي للتداخل المفرط
MNDO	Modified Neglect of Diatomic Overlap	إهمال تعديل التداخل ثنائي الذرة
AM1	Austin Model 1	نموذج1لAustin
PM3	Parametric Model number	معيارنموذج رقم3
LDA	Local Density Approximation	تقريب الكثافة المحلية
QSPR	Quantitative Struture - Propert Relationships	العلاقات الكمية(البنية الكيميائية-الخاصية)

QSAR	Quantitative Struture - Activity Relationships	العلاقات الكمية (البنية الكمية- الفعالية)
Kow	Octanol-Water Partition Coefficient	معامل التوزيع(اوكتانول – ماء)
LogKow	Octanol-Water Partition Coefficient Logarithme	لوغاريتم معامل التوزيع(اوكتانول - ماء)
MLR	Multiple Linear Regression	الانحدار الخطي المتعدد
PCR	Principal Component Regression	الانحدار المكون الرئيسي
ANN	Artificial Neural Network	الشبكات العصبية الاصطناعية
PLS	Partial Least squares	انحدار المربعات الصغرى الجزئي
MSE	Mean Squared Error	خطا تربيعي متوسط
DDL	Data Definition Language	لغة تعريف البيانات

مقدمة عامة

مـــــقـــــدمة

منذ الثورة الصناعية اصبحت الآلات معينة بل بديلة عن العنصر البشري فيما يتعلق بتحليل المواد الكيميائية و التنبؤ بخواصه الفيزيوكيميائية الى ان تعدت هذا الامر الى تصنيع المركبات بواسطة نظام حاسوبي خاص و بدون تدخل انسان.

ولكن يهمنا من خلال ما سبق هو التنبؤ بالخواص الفيزيوكيميائية للمركبات الكيميائية، فبالرغم من ان المواد الكيميائية تستخدم لتعزيز وتحسين و تسهيل مجالات عدة في الحياة الا انه بجانب فوائدها و إجابياتها ، توجد امكانية حدوث تأثيرات ضارة بالنسبة للانسان او البيئية ، و من بين هذه الاضرار، سمية بعض المواد فتعتبر السمية من الخواص الفيزيوكيميائية للمادة ، وللتنبؤ بهذه الخاصية تستعمل تقنيات حسابية .

إن النمذجة الجزيئية هي مجال يشمل جميع الأساليب النظرية والتقنيات الحسابية المستخدمة في تصميم نموذج أو تتبع سلوك الجزيئات، وطموح الكيميائي النظري هو أن يكون قادرعلى التنبؤ والتأكيد أو إعادة تفسير التجربة باستخدام النمذجة الجزيئية وفي الواقع فإن مثابرة الباحثين ولا سيما قوة مواردهم الحاسوبية تلعب لصالح الكيمياء النظرية ومجال تطبيقها.

فتعد برامج النمذجة الجزيئية من اهم البرامج المساهمة في تصوير المركب الكيميائي على هيئة ثلاثي الابعاد وتعطينا خصائص الجزيئات و في هذه الدراسة استخدمنا مجموعة من البرمجيات و هي : البرنامج (ChemDraw ultra12.0) لرسم البنية الجزيئية لهذه المركبات العضوية، و البرامج (HyperChem) و (Dragon) لنمذجة الجزيئات وحساب الواصفات، والبرنامج الرياضي (Xlsta) للحساب الاحصائي.

الهدف من هذا العمل هو التنبؤ بسمية مجموعة مكونة من 51 مركب كيميائي من عائلة الفينولات و الثيوفينولات حيث يرتكز على ايجاد العلاقة بين البنية الجزيئة للمركبات الكيميائية والخاصية الفيزيوكيميائية (**بنية/سمية**) بتطبيق تقنية QSPR .

وقد اجريت هذه الدراسة وفق الخطة التالية :

❖ مقدمة عامة حيث تحدثنا فيها عن الخاصية الفيزيوكيميائية المختارة و أهمية النمذجة الجزيئية والتعرف على البرامج المستعملة في الدراسة

❖ الجزء النظري حيث ينقسم هذا الجزء الى فصلين :

- الفصل الاول : تطرقنا فيه على عموميات حول المركبات الفينولية و الثيو فينولية و الخواص الفيزيزيكيميائية و مفاهيم على تقنية النمذجة و طرقها .
- الفصل الثاني : تحدثنا في هذا الفصل على طرق الحساب والبرامج المستعملة في هذه الدراسة.

❖ الجزء الثاني الحسابات والنتائج حيث تم في هذا الجزء تحليل ومناقشة النتائج التي توصلنا إليها، و هو ينقسم إلى فصلين كذلك:

- الفصل الثالث : تطرقنا إلى المركبات العضوية المكونة لقاعدة البيانات ثم شرح خطوات العمل التجريبية بالتفصيل.
- الفصل الرابع : في هذا الفصل تم عرض النتائج المتحصل عليها ومناقشتها

الجزء النظري

الفصل الاول
حوصلة
بيبليوغرافية

I- دراسة بيبليوغرافية:

مقدمة :

في يومنا هذا زاد اهتمام الباحثين في مجال الكيمياء بدراسة الخواص الفيزيوكيميائية المختلفة ،وتندرج الخاصية التي نحن بصدد دراستها ضمن الخواص البيئية و المتمثلة في خاصية السمية.ان اهمية و دور هذه الخاصية تحتم علينا تطبيق طرق و تقنيات للتنبؤ بها تحديدا في المركبات المختارة في مجموعة المعطيات الخاصة بهذا العمل و هي المواد الفينولية و الثيوفينولية التي سيتم تعريفها.

و من بين التقنيات المطبقة نجد تقنية النمذجة الجزيئية و المتمثلة في مجموعة من الخطوات قصد الحصول على معلومات حول هندسة وشكل الجزيء، و بعض الخصائص مثل الشحنية، الاستقطابية، الحجم و السطح الجزئي...، والتي تسمح ايضا بتحديد بنية وطاقة الجزيئات ونشاطها التفاعلي، كما استعملها الصيادلة المهتمين بالعلاقة **بنية/فعالية**(QSAR) و **بنية/خاصية**(QSPR)، و من هنا يمكن توضيح مفاهيم هذه التقنية و الطرق التي تعتمدها ثم التعرف على المراحل التي تمكننا من تحسين بنية الجزيئات، و اخيرا تعريف الواصفات الجزيئية .

I–1 –عموميات حول المركبات الفينولية و الثيوفينولية:

I–1–1– المركبات الفينولية :

بدأ الاهتمام مبكرا بالمركبات الفينولية المستخلصة من النباتات المستعملة في مجال الصناعة ، ومن الاصناف الاكثر شهرة لهذه المركبات هي التانينات و التي استعملت منذ القدم في دباغة الجلود وصناعة الحبر، وكذلك تدخل بعض المركبات الفينولية في صناعة البلاستك والمواد الملونة ، وتعرف هذه المركبات على انها فعالة ذات أوزان جزيئية منخفضة حاوية على حلقة أروماتية تحمل واحدا أو أكثر من مجاميع الهيدروكسيل وهي تؤدي دورا مهما في نمو وتكاثر النباتات، فضلا عن انها تحميها من الاصابة بالأمراض والحشرات وبالتالي تعد عوامل مقاومة طبيعية للنباتات، إذ تجعل جدران الخلايا غير منفذة للماء والغازات، وأنها تكون مسؤولة عن اعطاء صفة الصلابة للنباتات.

تكون اغلب الفينولات ذائبة في الماء وهي توجد مرتبطة مع السكر على هيئة كلايكوسيدات وتنشأ في جدران الخلايا.

لموقع وعدد مجاميع الهيدروكسيل في الفينولات علاقة مع الفعالية المضادة لهذه المركبات تجاه الاحياء المجهرية والفعالية المضادة للأكسدة.

تتواجد المركبات الفينولية في الطبيعة على هيئة فلافونيدات، حوامض فينولية، تانينات ستلبينات، اللكنينات .

تستخدم الفينولات على نطاق واسع في المنتجات المنزلية وكوسيط للتوليف الصناعي، على سبيل المثال: يستخدم الفينول نفسه (بتركيزات منخفضة) كمطهر في المنظفات المنزلية وفي غسول الفم.

ان المحاليل المركزة للفينول تسبب حروقًا شديدة ولكن غير مؤلمة للجلد والأغشية المخاطية رغم السمية العالية لهذا المركب ،فهناك فينولات الأقل سمية مثل (n-hexylresorcinol) حلت محل الفينول نفسه في قطرات السعال وغيرها من التطبيقات المطهرة حيث هيدروكسي تولوين بوتيل (BHT) له سمية أقل بكثير وهو أحد مضادات الأكسدة الشائعة في الأطعمة [1] .

I-1-2- المركبات الثيوفينولية:

المركبات الثيوفينولية هي فئة من المركبات المشتقة من الثيوفينول تحتوي جميعها على مجموعة سلفيدريل-(SHمرتبطة تساهميًا بحلقة عطرية.فالثيوفينول هو مادة سائلة عديمة اللون سامة و كريهة الرائحة وهو أبسط ثيول عطري. الهياكل الكيميائية للثيوفينول ومشتقاته مماثلة للفينولات باستثناء ذرة الأكسجين في مجموعة الهيدروكسيل (-OH) المرتبطة بالحلقة العطرية اين يتم استبدالها بذرة الكبريت. تشير البادئة إلى مركب يحتوي على الكبريت وعند استخدامه قبل اسم كلمة الجذر للمركب الذي يحتوي عادة على ذرة أكسجين ، في حالة "الثيول" يتم استبدال ذرة أكسجين الكحول بذرة كبريت[2].

ان الثيوفينول قابل للذوبان في الكحول و الإثر و البنزين،و قابل للتأكسد في الهواء خصوصا عندما يتم اذابته في هيدروكسيد الأمونيوم الكحولية ناتجا عن ذلك تكون ثنائي كبريتيد ثنائي الفينيل [3] $C_6H_5SSC_6H_5$.

يعد الثيوفينول مهيج وسام عن طريق الابتلاع أو الامتصاص عن طريق الجلد أو الاستنشاق [2].

فالتعرض الشديد لهذه المادة يسبب تهيجات في العينين و الجلد و الجهاز التنفسي و التهاب في الجلد و ازرقاق في الجسم نتيجة نقص الأكسجين و سعال، و ضيق في التنفّس، و ذمة رئوية ، صداع، دوخة و اضطرابات في الجهاز العصبي المركزي و غثيان، و تقيؤ و تلف في الكليتين و الكبد و الطحال.

هناك عدة طرق لتحضير هذه المادة و منها اختزال كلوريد بنزين سلفونيل بالخرصين وتأثير عنصر الكبريت على هاليد فينيل المغنيسيوم أو فينيل ليثيوم phenyllithium متبوعًا بالتحميض (أي تحميض وسط التفاعل) [3].

I-2- الخصائص الفيزيوكيميائية :

يعتمد تحديد او تدوين المادة على خصائصها و يمكن تقسيم هذه الخصائص الى فئتين فيزيائية و كيميائية[26]

-الخصائص الفيزيائية:

الخصائص الفيزيائية هي التي لا تغير الطبيعة الكيميائية للمادة. و تعرف الخاصية الفيزيائية للمادة النقية على انها اي شيء يمكن ملاحظته دون تغيير هوية المادة، تتكون الملاحظات عادة من انواع قليلة من القياسات الرقمية، على الرغم من وجود وصف نوعي (غير رقمي). تشمل الخصائص الفيزيائية : اللون، نقطة التجمد، نقطة الغليان...[4]

-الخصائص الكيميائية

اما الخصائص الكيميائية تُعرّف بأنّها خاصيّة غير مرئيّة للمادة يُمكن تحديدها بالحواس، وهي تحدد سلوك المادة أثناء تفاعل كيميائي، ويمكن تعريفها أيضاً بأنّها قدرة مادة على الاتحاد مع غيرها أو التحول إلى مادة أخرى

فهناك امثلة للخاصية الكيميائية مثل : التأكسد، الذوبان، السمية...[5] ،اي انها تغير الطبيعة الكيميائية للمادة حيث يمكن ملاحظة هذه فقط اثناء تفاعل كيميائي. على سبيل المثال، لا يمكنك معرفة ما اذا كان الكبريت قابل للاحتراق الا اذا حاولت حرقه. [26]

I-2-1-خاصية السمية:

تعتبر السمية خاصية بيولوجية من خواص المادة والتي تعني قدرة المادة على الحاق الضرر بكائن حي بغير الوسائل الميكانيكية [6]..هذه الخاصية تعتمد على التركيب الذري او التركيب الجزيئي للمركب [7]،اي ان السمية هي مقياس قدرة مادة سامة على احداث تأثيرات ضارة و من ضمنها الصحة [8] .

الفرضية القائلة بأنه إذا كان من الممكن فهم العوامل المسؤولة عن سمية مادة كيميائية ، فيمكن عندئذٍ التنبؤ بسمية المواد الكيميائية ذات الصلة دون الحاجة إلى إجراء تجارب على الحيوانات. لسوء الحظ هناك العديد من العوامل بعضها معقد للغاية والتي تحكم ما إذا كانت أبسط المواد الكيميائية الصناعية ستكون سامة، غالبية هذه العوامل (مثل التمثيل الغذائي والتوافر البيولوجي) خارج نطاق دراستنا في هذا الفصل. وبدلاً من ذلك ، فإن تركيز هذا الفصل يسلط الضوء على تقييم الحالة الإلكترونية لمواد كيميائية محتملة السمية ، وكيف تمكن هذه المعلومات المرء من البدء في تحديد النسبة والتنبؤ لاحقًا بجوانب معينة من صحة الإنسان والسموم البيئية[9]. وتنقسم السمية الى انواع : حاد، تحت الحاد، دون المزمن او مزمن [8].

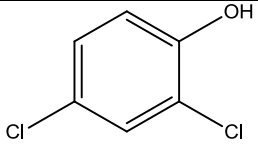
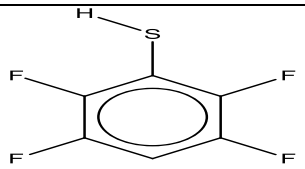
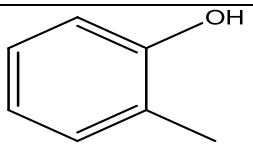
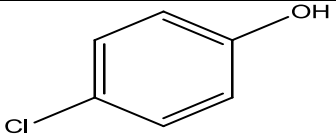
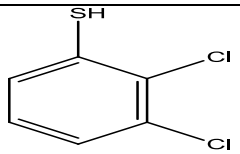
الجدول (I-1): درجات السمية [6]

درجة السمية	التأثير	التفصيل
0	لا تأثير	لا ضرر تحت أي ظرف
1	طفيف	التأثير يزول بسهولة عند التوقف عن التعرض للسبب
2	متوسط	تتضمن تأثيراً قد تعود او لا تعود الى طبيعتها. لكنها غير كافية لتتسبب بإعاقة دائمة او الموت
3	شديد	تتسبب بإعاقة دائمة او الموت بعد التعرض لكميات صغيرة
U	غير معروف	التأثير غير معروف

I-3-تعريف البنية الجزيئية للجزيء:

الهندسة الجزيئية أو البنية الجزيئية للجزيء هي في الأساس شكل أو ترتيب أو هيكل ثلاثي الأبعاد للذرات التي تشكل الجزيء. عندما تتشكل الجزيئات التي تنتج عن ترابط الذرات معا عن طريق الروابط الكيميائية المختلفة. تتحكم المدارات الفرعية المشاركة في الروابط الكيميائية في خواص و خصائص الجزئ تبعا لعوامل كثيرة.على سبيل المثال نجد جزيئات الماء ليست خطية فهو في الواقع على شكل 'V' وهناك زاوية تشكلت بين اثنين من ذرات الهيدروجين وذرة الأوكسجين من ما يقرب من 105° درجة. عندما نرسم الجزيئات في بعدين، فإننا في معظم الوقت نعتقد أن هذه الجزيئات هي مسطحة ولكن في الواقع كانت موجودة في العديد من الأشكال الهندسية و البنيات الجزيئية المختلفة. التركيب الكيميائي والبنية الجزيئية (الهندسة الجزيئية) للجزيء هي من تحدد في الأساس خصائص الجزيء مثل : الذوق، نقطة الغليان، المغناطيسية، الديناميكية، الاستقطاب، اللون، الرائحة ، السمية وجميع الخصائص الأخرى .[10]

الجدول (I -2): الهياكل البنيوية لبعض المركبات المستعملة

المركب	الهيكل
2,4 –Dichlorophenol	
2,3,5,6–tetrafluorothiophenol	
o–cresol	
P– ChlorophenoL	
2,3–Dichlorobenzenethiol	

I-4-تعريف النمذجة الجزيئية :

كلمة "الجزيئية" تعني بوضوح التعامل مع الجزيئات. و يعرف قاموس أوكسفورد النموذج بأنها "وصف مبسط أو مثالي لنظام أوعملية، في المصطلحات الرياضية كثيرا ما يستخدم لتسهيل العمليات الحسابية والتوقعات". تهتم النمذجة الجزيئية بتقليد سلوك أنظمة الجزيئي والجزيئات كما ترتبط هذه النمذجة بشكل ثابت بالنمذجة الحاسوبية ومع ذلك، أحدثت التقنيات الحاسوبية ثورة في النمذجة الجزيئية إلى درجة أن غالبية الحسابات لا يمكن أن تنجز بدون استعمال الحاسوب، هذا لا يعني أن نموذجا أكثر تطورا هو بالضرورة أفضل أي واحد بسيط، ولكن أجهزة الكمبيوتر لديها بالتأكيد مجموعة أوسع من النماذج التي يمكن النظر فيها والنظم التي يمكن تطبيقها .[11]

و يمكن النظر الى النمذجة الجزيئية على انها مجموعة من التقنيات الحسابية القائمة على طرق الكيمياء النظرية و البيانات التجريبية التي يمكن استخدامها لوصف و توقع خصائص الانظمة الجزيئية المعقدة في بيئات مختلفة. يفترض هذا النهج وجود تطابق فريد بين اي خاصية فيزيائية او تقارب كيميائي او نشاط بيولوجي لمركب كيميائي و بنيته الجزيئية [25] . و الغرض منها هو التنبؤ بهيكل وتفاعل الجزيئات او انظمة الجزيئات [12]

تستخدم النمذجة الجزيئية لفحص التركيبات الجزيئية، الدراسة الديناميكية و دراسة الخصائص الفيزيوكيميائية للمركبات كما أنها تستخدم أيضا لفحص أنواع الانشطة البيولوجية [13]

وفي دراستنا هذه استعملنا برنامج Hyperchem كاداة تمكننا من تحديد هندسة الجزيئات و تقدير الخصائص الفيزوكيميائية

I-4-1-اسباب استخدام النمذجة :

- ✓ قلة التكلفة
- ✓ سهولة التنفيذ ودقة الأداء
- ✓ وفرة الجهد والوقت
- ✓ المرونة

✓ إمكانية تطبيق النمذجة على مدى واسع من مجالات الحياة المختلفة[11]

الطرق التي اعتمدت عليها النمذجة الجزيئية نذكر :

1/ طرق الميكانيكا الجزيئية .Molecular Mechanics Methods

2/ طريقة الحل الشامل .Ab -initio Method

3/ الطرق الشبه عملية .Semi-Empirical Methods

4/ نظرية الدالة الوظيفية للكثافة .Density Functional Theory (DFT) [14]

I-4-2- طرق النمذجة الجزيئية:

هناك عدة طرق تستعمل لنمذجة الجزيئات يمكنها التنبؤ بالسلوكيات المعقدة بشكل متزايد للمركبات الكيميائية و البيولوجية المختلفة و تتمحور هذه الطرق في ما يلي :

I-4-2-1 - الميكانيكا الجزيئية :

ظهرت الميكانيكا الجزيئية لأول مرة في عام 1930 لكنها تطورت في الستينات ومع تطوير أجهزة الكمبيوتر والتقدم في الاداء جعل من الممكن تحديد طاقة الجزيء وفقا للإحداثيات الذرية، والبحث عن الحد الادنى للطاقة [13]

تسمح الميكانيكا الجزيئية بدراسة مجموعة من الخصائص من خلال وصف طاقة النظام انطلاقا من مجموع سلسلة مساهمات التي تمثل التفاعلات داخل الجزيئات وبين الجزيئات[13] . حيث تعد من بين أسرع الطرق و ابسطها، حيث تستخدم في معالجة المركبات ذات الاحجام الكبيرة كالبروتينات و الستيرويدات وغيرها، إلا أنها لا تعطى نتائج دقيقة مقارنة بالطرق الاخرى كونها عاجزة عن كشف الخصائص الكيميائية المتعلقة بالإلكترونات، كعزم ثنائي القطب وشحنة الذرات و الأماكن الفعالة . وتعتمد هذه الطرق في حساب الطاقة على تحسين أو تغيير البنية الفراغية للمركبات المدروسة للوصول إلى اقل طاقة ممكنة. فتعالج الجزيئات وكأنها كرات متصلة مع

بعضها البعض عن طريق نوابض لا تأخذ الالكترونات بعين الاعتبار، ولحساب قيمة الطاقة يجب التعرف على طول الرابطة الحقيقية بين كل ذرتين و الزاوية بينهما، وكذلك طاقة الشد و الانحناء للرابطة الجديدة (النابض).

طاقة الجزيء E في طرق الميكانيكا الجزيئية تعبر عن مجموع طاقات الترابط E_{Bond} واللاترابط $E_{Nonbond}$ حيث تعبر طاقة الترابط عن التأثير المباشر عن طريق الترابط كمختلف أنواع طاقة الشد و الانحناء. أما طاقة اللاترابط فتطلق على التأثير غير المباشر بين الذرات كقوى فاندر فالس والقوى الناتجة عن الكهرباء الساكنة، وتكتب طاقة الجزيء على الشكل التالي:

$$E = E_{Bond} + E_{Nonbond}$$

$$E = \sum_{Angles} E_{Bend} + \sum_{Bond} E_{Stretch} + \sum_{Dihedrals} E_{Torision} + \sum_{Pairs} E_{Nonbond}$$

I-4-2-1-1-انواع طرق الميكانيكا الجزيئية :

طرق الميكانيكا الجزيئية تختلف باختلاف التأثير المتبادل بين الذرات الذي يعبر عنه بحقل القوة الذي يكشف عن مجموعة المعاملات و المعادلات التي تصف الذرات و التأثير المتبادل بينهما التي تعبر عن شكل دالة الطاقة المستخدمة في كل طريقة.

حيث تم تطوير هذه الطرق و لاستخدامها في معالجة المركبات بطريقة أكثر من الطرق الاخرى، من بين هذه الطرائق نذكر :

✓ بناء النموذج المدعوم باستخدام تقنية إعادة الطاقة(AMBER) يعتمد على شروط الترابط ويتم تضمينها عبر النتائج الجيدة للبروتينات والاحماض النووية، وتكون غير منتظمة بالنسبة للأنظمة الاخرى.

✓ AMBER: Assisted Model Building with Energy Refinement.

✓ CHARMM: هي طريقة حسابية تستخدم مصطلحات التكافؤ، يتم تطبيقها على الجزيئات الحيوية الضخمة كالبروتينات والاحماض النووية.

✓ CHARMM: Chemistry at HARvard Macromolecular Mechanics.

✓ OPLS:وهي إمكانية محسنة لمحاكاة السوائل تعتمد على مصطلحات التكافؤ، استخدمت على نطاق واسع في نمذجة السوائل و الديناميكا الجزيئية للجزيئات الحيوية.

✓ OPLS: optimized potential for liquid simulations.

✓ MMFF: تعتبر قوة Merck الجزيئية من أكثر القوى التي استخدمت للاغراض العامة، و خاصة الجزيئات العضوية، كما تستخدم لمحاكاة الديناميكا الجزيئية و تحسين الهندسة الجزيئية.

✓ MMFF: Merck Molecular Force Field.

✓ UFF: تعني القوة العالمية تستخدم في معالجة المركبات غير العضوية، تستخدم أربعة شروط للتكافؤ. باستثناء الكهرباء.

✓ UFF: Universal Force Field.

✓ Dreiding:تعبر هذه الطريقة عن قوة الجزيء العضوي أو الحيوي، و استخدمت كذلك لمعالجة الجزيئات الحيوية الضخمة، يعتمد على مصطلحات التكافؤ مثل الكهرباء .

✓ Dreiding: The Dreiding force field.[14]

MM^+ : هو امتداد لحقل القوة MM2 مع إضافة بعض البارامترات إضافية، و هو حقل قوة قوي يمكن تطبيقه على الجزيئات الاكثر تعقيدا مثل المركبات غير العضوية[13]

I-4-2-1-2- تطبيق حساب الديناميكا الجزيئية :

قبل البدء في النمذجة بواسطة الديناميكا الجزيئية من المهم أن يكون نموذج البدء قريب من البنية الفعلية.

- التقليل إلى الحد الادنى بواسطة الميكانيكا الجزيئية : حيث يبدأ بصفة عامة بالتشكيل الامثل بواسطة MM
- الديناميكا الجزيئية : بشكل عام يتم تقسيم حساب DM إلى ثلاث مراحل:

-الحرارة : من أجل تسخين النظام لإحضاره إلى درجة الحرارة المرغوبة (بشكل عام 300K) وهذا الارتفاع في درجة الحرارة يتطلب من 2 إلى 10ps من المحاكاة (2000 الى 10000 خطوات التكامل)

-الاتزان: هي مرحلة مهمة من أجل استقرار درجة حرارة النظام، ثم هناك تبادل بين الطاقة الكامنة والطاقة الحركية. هذه الخطوة يمكن أن تستمر ما بين 10 إلى 20ps حيث يتم حساب درجة الحرارة بعد كل خطوة وفق المعادلة التالية:

$$\frac{dT}{dt} = \tau^{-1}(T_0 - T)$$

τ : يسمى وقت الاسترخاء ويجب أن يكون قيمته صغيرة بما يكفى عادة ما يعادل 10fs .

T_0 : درجة الحرارة المرجعية .

T : درجة حرارة النظام .

-الديناميكية أو الانتاج: وهي المرحلة القابلة للاستغلال والتي تدوم في المتوسط من 60 الى 100ps، أثناء هذه المحاكاة يتم حفظ 0.05 أو 0.1ps [13]

I-4-2-2- طريقة الحل الشامل Ab -initio Method :

تعتمد طرق Ab-initio أساسا على نظريات ميكانيكا الكم و حلولها من البداية (الاعتماد على ثوابت فيزيائية كشحنة وكتلة الالكترون، سرعة الضوء).... ،أي أنها تعتمد على حل معادلة شرودنغر التي تصف حركة الالكترون داخل الجزيء، وعند حلها يتم الحصول على طاقة و دالة موجية في شكل معادلة رياضية تعبر عن التوزيع الالكتروني داخل الجزيء، الذي يمكننا من التعرف على العديد من خصائص الجزيء كحساب قطبية الجزيء(المواضع المحبة للنوى و المواضع الفقيرة المحبة للإلكترونات)، طاقة التأين، حساب البنية الفراغية الجزيئية ...

معادلة شرودنغر هي ركيزة من ركائز الكيمياء النظرية و الحاسوبية، الا أنها غير قابلة للحل الا للأنظمة وحيدة الالكترون، لذلك يستوجب إدخال عدة تقريبات رياضية يتم من خلالها إهمال التأثير

المتبادل بين الالكترونات من اجل حل معادلة شرودنغر. والتي تكتب على النحو التالي :

$$H\,\Psi(r) = E\,\Psi(r)$$

H :المعامل الهاميلتوني (The Hamiltonian operator).

Ψ : الدالة الموجية(The wave function)

E : الطاقة قيمة خاصة للمعامل الهاميلتوني (The energy, eigenvalue of the Hamiltonian operator)

يتم في هذه الطريقة تقسيم معادلة شرودينغر متعددة الالكترونات إلى عدة معادلات بسيطة وحيدة الالكترون ينتج عن حلها دالة موجية تدعى (المدار) وطاقة تسمى (طاقة المدار). وبمقارنة طرق Ab- Initio مع طرق الميكانيكا الجزيئية من حيث السرعة نجد أن طرق Ab-Initio أبطأ بكثير من طرق الميكانيكا الجزيئية، أما من حيث الدقة فتكون أدق.

I-4-2-2-1--أنواع طرق Ab-Initio

نظرا لعدم مراعاة التأثير المتبادل بين الالكترونات تم تطوير عدد من النظريات والطرق التي تنطلق من طريقة هارتري و فوك تقوم بإضافة الترابط الالكتروني، لتحسين وتدقيق طرق الحساب والوصول إلى نتائج أكثر دقة في إيجاد قيمة الطاقة أو الشكل الفراغي للجزيء، و من هذه الطرق نذكر:

✓طريقة رابطة التكافؤ المعممة(GVB) هي طريقة مصغرة من MCSCF حيث تعتمد على زوج من المدارات لكل رابطة جزيئية.

✓ GVB: The Generalized Valence Bond method.

✓CI: Configuration Interaction

من الممكن أن تنشأ دالة موجات CI وتبدأ بحسابات MCSCF بدلا من وظيفة الموجات الديكامترية .تكون هذه الطريقة مكلفة للغاية من حيث الموارد الحسابية.

✓ MCSCF: Multi-Configurational Self-Consistent Field

في هذه الطريقة تتم عملية تحسين المدارات للاستخدام مع الدالة موجية متعددة، تكون أكثر دقة مقارنة بحسابات CI , كما تستخدم للتنبؤ بطاقة الترابط

✓ بالإضافة إلى بعض الطرق الاخرى التي يعبر عنها بأساليب الحسابات المترابطة التي تسبب تحسين الطاقات الحسابية و الهندسة الجزيئية بالنسبة للجزيئات العضوية كنظرية الكتلة المتقاربة (cc : coupled cluster theory)

وطريقة (MPn: Mùller-Plesset Perturbation Theory) حيث n :تمثل تراتيب التصحيح [14]

I-4-2-3- الطرق الشبه عملية Semi- Empirical Methods

تعتمد طريقة الحساب في الطرق الشبه عملية على معادلة شرودنغر، لذلك تعد من الطرق الشبيهة إلى حد كبير لطرق Ab- Initio الا أن في هذا النوع من الحسابات يتم حل معادلة شرودنغر بمعاملات تجريبية، حيث يتم استبدال التكاملات المعقدة في طرق Ab-Initio بمعاملات و تكاملات تجريبية تسحب من قاعدة بيانات التكاملات ذات نتائج عالية الدقة كحسابات الطرق ذات المستوى العالي مثل"DFT" .

وتعد الطرق شبه العملية بطيئة مقارنة بطرق الميكانيكا الجزيئية، الا أنها أسرع نوعا ما من طرق

Ab – Initio

I-4-2-3-1-أنواع طرق الشبه عملية Methods (Semi- Empirical) :

تختلف طرق الحساب في هذه المجموعة باختلاف قاعدة البيانات المتعلقة بالمعاملات والتكاملات التجريبية، نذكر منها :

- الاهمال الكامل للتداخل التفاضلي(CNDO) :هو أبسط إهمال لطرق التداخل(NDO) حيث تميز مدارات التكافؤ باستخدام مجموعة بسيطة من مدارات نوع Slater ، تستخدم طريقة (CNDO) للتوقع

الاولي لحسابات طرق Ab-Initio للهيدروكربونات ,كما تستخدم لدراسة البنية الالكترونية والفراغية للجزيء، ورغم بساطة طريقة الحساب في(CNDO) الا أنه لم يتم استخدام هذه الطريقة على نطاق واسع بسبب محدودية دقة نتائجها لعدم قدرتها على تقدير التفاعلات ثنائية الإلكترون و التنبؤ بدقة الهياكل الجزيئية، وقد تبدو طريقة CNDO معقدة بعض الشيء إلا أنها تمثل تبسيطا شاملا لنظريةHF ، وقد أدخلت بعض التعديلات الطفيفة على هذه الطريقة فتفرعت الى مثل:

CNDO/S ,CNDO/BW CNDO/2, CNDO/1

✓ CNDO: Complete Neglect of Differential Overlap

- الاهمال الوسيطي للتداخل المفرط (INDO) : هي طريقة كانت تستخدم للانظمة العضوية، أما الان فتم استبداله بأساليب أكثر دقة، طريقة INDO لديها إمكانية كبيرة لنمذجة التحليل الطيفي للاشعة فوق البنفسجية والمرئية للمركب نتيجة لمعالجتها للتفاعلات الإلكترونية ذات المركز الواحد بطريقة أفضل، كما تستخدم للتنبؤ بزوايا رابطة التكافؤ بدقة كبيرة و تكون ضعيفة نوعا ما في التنبؤ بالهندسة الجزيئية الكلية

✓ INDO: Intermediate Neglect of Differential Overlap.

- إهمال تعديل التداخل ثنائي الذرة (MNDO) : تم الإبلاغ عن الاهمال الموحد للتداخل التفاضلي بطريقة تستند إلى طريقة NDDO للعناصر C,O,N,H مع الاتفاقيات المحددة من قبل NDDO للتكامل. تعطي طريقةMNDO نتائج نوعية مقبولة لمختلف الانظمة العضوية، الا أن دقة نتائج الطريقتين AM1 و PM3 تغلبت على هذه الطريقة مما جعلتها تتراجع. كونها تعطي نتائج نوعية وكمية غير صحيحة، فيتم التقليل من طاقات التحفيز الالكتروني المحسوبة.

✓ MNDO: Modified Neglect of Diatomic Overlap.

- AM1: Austin Model 1

وهي طريق سريعة جدا شائعة لنمذجة المركبات العضوية حيث تعطي نتائج أكثر دقة على خلاف بقية الطرائق شبه التجريبية. تستخدم لتحسين طاقات التنشيط وتوقع الهندسة الجزيئية. وتكون أكثر دقة لطاقات رابطة الهيدروجين.

- PM3: Parametric Model number

وهو نموذج قوي من نماذج NDDO يستمر استخدامه على نطاق واسع، تستخدم هذه الطريقة تقريبا نفس المعادلات المستعملة في AM1 ،وهي أكثر دقة منها وخاصة في حساب الزوايا رابطة الهيدروجين .كما تستخدم لحساب الخصائص البنيوية، الطاقة الكلية، حرارة التشكل , و قد تؤدي بعض خصائص مجوعة PM3 إلى سلوكيات غير طبيعية لا تزال قيد المعالجة.

- Huckel : تعتبر طريقة هيكل من أبسط وأقدم الطرق شبه التجريبية، تختص في حساب الكترونات التكافؤ p للجزيئات المستوية المترافقة، تقوم بإعادة تمثيل التناظر المداري يمكن أن تعطي معلومات كمية أو نوعية للمركبات المترافقة .وهي نموذج من طرق هاملتون لإلكترون، ترتبط بالطرق التجريبية، لكنها نادرا ما تستخدم لاعتبارها عملية حسابية يمكن تطبيقها يدويا.
- Huckel Extended: نظرية هيكل الموسعة(EHT) وهي من ابرز الطرق التي تشهد استخداما واسعا، تعتمد هذه الطريقة على إهمال الإلكترونات الأساسية التي تكون متغيرة في مداراتها وتأخذ بعين الاعتبار جميع الكترونات التكافؤ. وتعتبر الذرة عند احتلالها لمدرات d أعلى مستوى للمساهمة في مدارات التكافؤ التي يتم تمثيلها في مدارات (STO) slater في إحداثيات قطبية محورها الذرة. ومن سلبيات هذه الطريقة أنها لا تدرس الحركة الدورانية للإلكترون.

I-4-2-4- نظرية الدالة الوظيفية للكثافة Density Functional Theory (DFT) :

هي طريقة عرفت اهتماما كبيرا في أواخر الثمانيات وحتى وقتنا الحاضر، وتعرف الدالة الوظيفية بأنها دالة ينتج عن حلها دالة أخرى أي دالة الدالة، فتعتبر الطاقة تابع لدالة الكثافة ، و تعتمد هذه الطريقة على معادلة شرودنغر كمثيلاتها من طرق Ab- initio و الطرق الشبه تجريبية، الا أنها تستخدم التوزيع الإلكتروني المباشر (دالة الكثافة الإلكترونية) بدلا من الدالة الموجية. ويمكن التعبير عن الطاقة الأرضية $0\ E[\rho(\vec{r}\ )]$ لنظام ما بالمعادلة التالية:

$$E[\rho((\vec{r})] = T_S[\rho((\vec{r})] + J[\rho((\vec{r})] + E_{Xc}[\rho(\vec{r})] + E_{Ne}[P]$$

$T_S[\rho((\vec{r})]$: Kinetic energy of non-interacting system.

(الطاقة الحركية للنظام غير التفاعلي)

$J[\rho((\vec{r})]$: Electron-Electron repulsion energy.

(طاقة صد الإلكترون)

$E_{Ne}[P]$: :External potential .

(الكمون الخارجي)

$E_{Xc}[\rho(\vec{r})]$: Exchange – Correlation energy functional

(طاقة التبادل –الارتباط الوظيفية)

يتم حساب الدالة الوظيفية للترابط والتبادل:$E_{Xc}[\rho(\vec{r})]$ التي تحتوي على مجموعة من المعلومات التي لايمكن الحصول عليها، بجمع كل من الدالة الوظيفية للترابط $E_C[\rho(\vec{r})]$ و الدالة الوظيفية للتبادل $E_X[\rho(\vec{r})]$

علي النحو التالي :

$$E_{Xc}[\rho(\vec{r})] = E_X[\rho(\vec{r})] + E_C[\rho(\vec{r})]$$

I-4-2-4-1- انواع نظرية الدالة الوظيفية للكثافة:

على الرغم من الاختلاف في شكل الدالة الوظيفية التي تناسب جميع الأنظمة لحساب طاقة التبادل و الترابط $E_{Xc}[\rho(\vec{r})]$بين طرائق نظرية الدالة الوظيفية للكثافة. إلا أن الشكل المثالي والصحيح لهذه الدالة لا يزال مبهم. ومن بين تقريبات الدالة الوظيفية نذكر:

▪تقريب الكثافة المحلية (LDA) هو أبسط تقدير للمشكلة الكاملة يعتمد على كثافة الإلكترونات فقط في الأنظمة ذات دوران عالي. استخدمت في حساب طاقة النطاق، أما نتائج الهندسة والحسابات الجزيئية تكون أقل دقة

▪ LDA: Local Density Approximation.

▪الدوال الوظيفية الهجينة: يتم دمج نظرية الدالة الوظيفية للكثافة DFT مع نظرية هارتري فوك HF لوصف الدالة الوظيفية لطاقة التبادل:

▪GCA: Gradient Corrected Approximation.

▪B3LYP :هي الطريقة الأكثر استخدام للحسابات الجزيئية نتيجة للدقة العالية في نتائجها في مجموعة كبيرة من المركبات المختلفة وخاصة المركبات العضوية

▪B3LYP: (Becke's 3 parameter exchange correlation functional which uses 3 parameters) and LYP (The Lee, Yang and Parr correlation functional).[14]

I-5-تعريف عملية التحسين :

نأخذ مثال لفهم عملية التحسين بشكل مبسط :

أرادت بلدية ما عمل مكب نفايات حيث يجب أن يكون هذا المكان بعيدا عن التجمعات السكانية ويجب الأخذ في الحسبان أيضا الأراضي الزراعية أو الموارد التي يمكن استثمارها في هذه المنطقة . لكن أيضا ليس من المنطقي أن يكون بعيدا جدا عن الشاحنات وشبكة الطرق التي تسهل عملية الوصول إلى هذا المكب وأيضا من الضروري تحديد مساحة هذه المنطقة والابتعاد عن خطوط الهاتف والكهرباء وأنابيب المياه ، أي كيف يمكن تحقيق أفضل نتيجة وأقل خسائر .

هذه العمليات لا تتم بشكل عشوائي في الغالب لأنها تستند إلى حسابات رياضية تندرج تحت علم البرمجة الخطية وتستخدم عدة طرق وخوارزميات تصنف باسم تقنيات التحسين أو Optimization Techniquesلتحديد أفضل نتيجة بناء على مجموعة من القيود مثل التي ذكرت سابقا. [17]

فالتحسين الهندسي او Geometry optimization هو اسم الاجراء الذي يحاول العثور على تكوين الحد الادنى من الطاقة للجزئ [18]

أي يمكن تحسين بنية الجزيئات المدروسة للحصول على هياكل جزيئية نموذجية اقل طاقة (اكثر استقرار)[14]

I–6– تعريف الواصف الجزيئي :

في العقود الماضية ، ركزت الكثير من الأبحاث العلمية على كيفية التقاط وتحويل – من خلال مسار نظري – المعلومات المشفرة في بنية جزيئية إلى واحد أو أكثر من الأرقام المستخدمة لإنشاء علاقات كمية بين الهياكل والخصائص، والأنشطة البيولوجية، أو غيرها من الأنشطة التجريبية المناسبة– erties. الواصفات الجزيئية هي تمثيلات رياضية رسمية للجزيء، يتم الحصول عليها من خلال خوارزمية محددة جيدًا، ويتم تطبيقها على إعادة إرسال جزيئية محددة أو إجراء تجريبي محدد جيدًا:

الواصف الجزيئي هو النتيجة النهائية لإجراء منطقي ورياضي يحول المعلومات الكيميائية المشفرة ضمن تمثيل رمزي للجزيء إلى رقم مفيد أو نتيجة لبعض التجارب المعيارية[9]

حيث تلعب الواصفات الجزيئية دور أساسيا في تطوير نماذج الكيمياء والعلوم الصيدلانية وسياسية و حماية البيئة و علم السموم وعلم السموم البيئية والبحوث الصحية ومراقبة الجودة. يتم توفير الدليل عن اهتمام المجتمع العلمي بالواصفات الجزيئية من خلال العدد الهائل من الواصفات المقترحة : تم تحديد اكثر من 5000 اداة مشتقة من نظريات وطرق مختلفة وقابلة للحساب باستخدام برنامج مخصص للبنية الكيميائية.[9]

فهذه الواصفات المقترحة بالاعتماد على النظريات المختلفة تهدف الى توقع نشاط والخواص الفيزيوكيمياوية للجزيئات.

يعتمد محتوى معلومات الواصفات الجزيئية على نوع التمثيل الجزئي وعلى الخوارزمية المعرفة لحسابها.

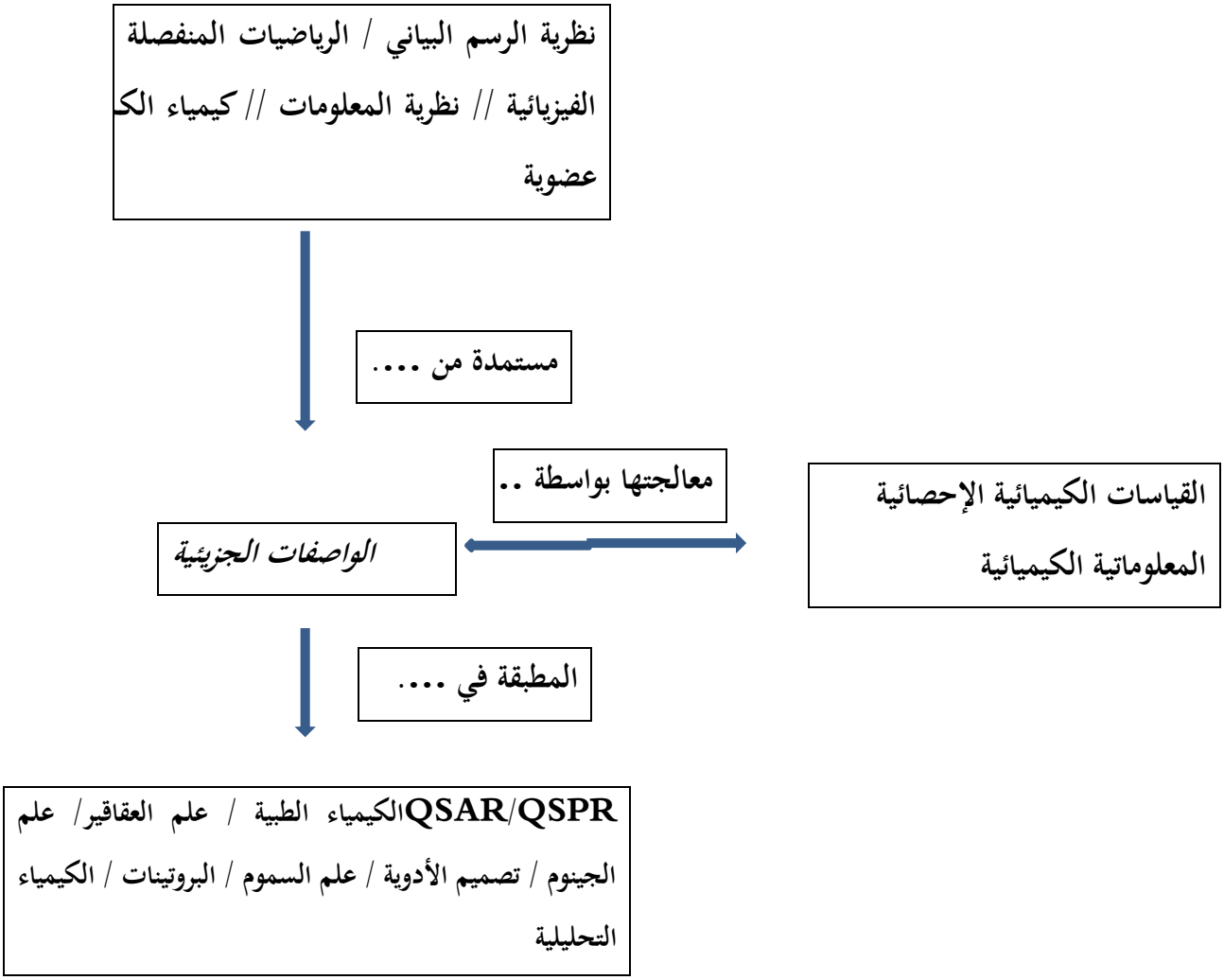
1 – فهناك واصفات جزيئية بسيطة مشتقة لحساب بعض أنواع الذرات أو الأجزاء الهيكلية في الجزيئة

2 –وهناك واصفات طوبولوجية أو تسمى الواصفات ثنائية البعد (2D) مشتقة من خوارزميات طوبولوجية (رسم بياني جزئي) .

3 –واصفات هندسية وتسمى الواصفات ثلاثية البعد (3D) (مشتقة من التمثيل الهندسي).

إن أغلب الواصفات لا تستطيع تمثيل التعقيد الجزئي لجزيء ما، أو تمثيل كل الاستجابات الفيزيوكيميائية كنتيجة للتفاعلات الحيوية، تعتبر معظمها تقريبات لكن يمكن اتخاذها كنماذج لدراستها.[19]

فإن الواصفات الجزيئية ، التي ترتبط ارتباطًا وثيقًا بمفهوم التركيب الجزيئي ، تلعب دورًا أساسيًا في البحث العلمي ، كونها النواة النظرية لشبكة معقدة من المعرفة ، كما هو موضح في الشكل [9]:



الشكل (I -1) :ارتباطات الواصفات الجزيئية في البحث العلمي [9]

و لحساب الواصفات الجزيئية للمركبات يمكن استعمال برنامج (DRAGON) وهو عبارة عن تقنية مطبقة لحساب الواصفات الجزيئية، أنجز من طرف العالم Milano ومجموعة بحث (QSAR) . حيث يحتوي على 1664 واصفة موزعة على 20 كتلة منطقية، تشمل العديد من الواصفات الجزيئية ،منها الطوبولوجية ، الهندسية ، الالكترونية ، و الواصفات المتعلقة بالفاعلية البيولوجية والصيدلانية.[19]

الجدول (I-3):اهم كتل الواصفات الجزيئية [19]

البعد	رقم الكتل	اسم الكتلة بالعربية	اسم الكتلة بالعربية	عددها
2D	1	Constitutional descriptors	الواصفات الاساسية	48
2D	2	Topological descriptors	الواصفات الطوبولوجية	119
2D	3	Walk and path counts	مسار وطرق العد	47
2D	4	Connectivity indices	مؤشرات الرابط	33
2D	5	Information indices	مؤشرات المعلومة	47
2D	6	2D autocorrelations	الارتباط الذاتي 2D	96
2D	7	Edge adjacency indices	مؤشرات جوار الحافة	107
2D	8	Burden Eigen value descriptors	واصفات القيم الحرة ل Burden	64
2D	9	Topological charge indices	مؤشرات الشحن الطوبولوجية	21
2D	10	Eigen value-based indices	مؤشرات اساسها قيم حرة	44
2D	11	Functional group counts	مجاميع المجموعات الوظيفية	154
2D	12	Atom-centered fragments	اجزاء مراكز الذرة	120
2D	13	Molecular properties	الخصائص الجزيئية	29
3D	14	Randic molecular profiles	المقاطع الجانبية للجزيئات Randic	41
3D	15	Geometrical descriptors	الواصفات الهندسية	74
3D	16	RDF descriptors	واصفات RDF	150
3D	17	3D-MoRSE descriptors	واصفات 3D-MoRSE	160
3D	18	WHIM descriptors	واصفات WHIM	99
3D	19	GETAWAY descriptors	واصفات GETAWAY	197

3D	20	Charge descriptors	الواصفات الشحنية	14

الخاتمة :

في الاخير تحصلنا في هذا الفصل على مفاهيم حول هندسة و شكل الجزيء و خصائصه الفيزيوكيميائية....و كذلك تقنية النمذجة الجزيئية التي بدورها تستخدم لتجسيد محاكاة سلوك الجزيئات وطرقها، الى ان وصلنا الى الواصفات و هي النتيجة النهائية لاجراء رياضي لتحويل المعلومات الكيميائية الى رموز و انطلاقا من هذه الاخيرة و باستعمالها في الحسابات الاحصائية من خلال برمجيات مختصة (احصائية) نتمكن من بناء علاقة كمية بنية/خاصية (QSPR) الذي هو موضوع الفصل الثاني.

الفصل الثاني
الطرق و البرامج المستعملة

II- الطرق و البرامج المستعملة :

مقدمة :

للتنبؤ بخاصية السمية يمكن تطبيق طرق نظرية بديلة، حيث اتاح تطوير تقنيات النمذجة الجديدة تنفيذ العديد من الاساليب بما في ذلك العلاقات الكمية (QSAR/QSPR) و التي تعتمد معظمها على ايجاد علاقة بين مجموعة من الارقام الحقيقية " تسمى الواصفات الجزئية " و الخصائص او النشاط المراد توقعه . تتيح هذه الطرق تبرير البيانات التجريبية المتاحة و التنبؤ (بخاصية / نشاط) المركبات او المركبات الجديدة التي لا تتوفر لها بيانات تجريبية، وفي السنوات الاخيرة تطور استخدام النمذجة الجزيئية في حساب نماذج (QSPR/ QSAR) باستخدام الانحدار الخطي المتعدد (MLR) و الشبكة العصبية الاصطناعية (ANN).

للتنبؤ بهذه الخاصية تم استعمال عدة برمجيات مختلفة الوظائف لحل هذه المسألة، لذا في هذا الفصل سنتطرق الى الأساليب والطرق النظرية المتبعة بالإضافة الى البرمجيات المستعملة في دراستنا.

II-1- نبذة تاريخية عن QSAR/QSPR :

يرجع تاريخ استخدام العلاقات (QSAR/QSPR) في علم السموم المائية إلى الأعمال التي أنجزت في أواخر القرن التاسع عشر من قبل الباحثين (أوفرتون في زيورخ) و(ماير في ماربورغ) والتي توضح أن قدرة المواد على إحداث تخدير في صغار الضفادع وصغار الأسماك ترتبط ارتباطا مباشرا بمعامل توزعها المقاس بين زيت الزيتون والماء. طرح أوفرتون في درسته عام 1901 فرضية بان هذا الإرتباط يعكس السمية التي تحدث في تركيز أو في حجم الجزيئي القياسي، في نطاق موقع جزيئي معين في داخل الكائن العضوي .

كما خلص إلى أن هذا يناظر التركيز أوالحجم نفسه في مختلف الكائنات، بغض النظر عما إذا كان الامتصاص من الماء أوعن طريق استنشاق غازي. ويعرف هذا الارتباط في علم التخدير بإسم نظرية ماير- أوفرتون .

اقترح كورفن هانس ومعاونوه في كلية بومونا استخدام نظام معامل التوزيع Kow كنظام قياسي للتوزيع، ووجد أن معاملات التوزيع هذه قيم جمعية تركيبية يمكن تقديرها مباشرة على أساس التركيب الكيميائي. وذكر

هؤلاء الباحثين كذلك أنه يمكن تصميم نماذج للعلاقات (QSAR/QSPR) على أساس تحليل إحصائي للانحدار في النتائج. أعلن الباحثون عن طريق تطبيق هذا المنهج في عام 1982 ما مجموعة 138 نموذجا للعلاقات (QSAR/QSPR) على شكل

$$Log(1/C) = A \log K_{ow} + B$$

K_{ow} هو معامل التوزيع (الأكتانول – الماء)

C : التركيز المولي للمادة الكيميائية الذي يسبب استجابة بيولوجية قياسية لتأثير المركبات العضوية غير الإلكتروليتية غير المتفاعلة في حيوانات كاملة، أوأعضاء، أوخلايا، بل وأنزيمات نقية.

وكانت الإنحدارات ومقاطع محاور الإحداثيات متطابقة تقريبا في خمس من هذه المعادلات التي بسمية خمسة كحولات بسيطة أحادية الهيدروكسيل لخمسة أنواع من الأسماك، وكانت في الواقع مطابقة تقريبا لما توصل إليه كونمان في 1971 ، الذي يبدو أنه لم يكن على علم بأعمال هانس السابقة .

وأثبت كونمان وآخرون أن كل هذه المواد البسيطة غير الإلكتروليتية وغير المتفاعلة تعمل من خلال آلية تخدير في اختبار السمية الحادة في الأسماك، مما يؤدي إلى سمية دنيا أوسمية خط الأساس. [11]

II–1–1– تعريف QSPR :

QSPR هي العملية التي من خلالها يتم تأسيس العلاقات الكمية بين البنية الجزيئية لمجموعة مركبات مع الخاصية الفيزيوكيميائية. وذلك بالمراحل الرئيسية الثلاثة لبناء نموذج (حساب) QSPR وهي :

- اختيار الواصفات الجزيئية المناسبة مع إشكالية بنية الكيميائية – الخاصية.
- استغلال قيم الواصفات الجزيئية كمتغيرات لتحديد العلاقة التي تربطها مع الخاصية المدروسة، وذلك باستخدام برامج خاصة.
- وضع معايير الأداء والتحقق من شأنها أن تساعد في اختيار أفضل النماذج للإشكالية المطروحة وتقدير أخطاء التنبؤ.[11]

الهدف من هذه الطريقة هو تحليل البيانات البنيوية للكشف عن عوامل خاصية قياسها. وللقيام بذلك، يمكن استخدام الأدوات التالية:

الانحدار الخطي المتعدد (MLR)

الانحدار المكون الرئيسي (PCR)

انحدار المربعات الصغرى الجزئي (PLS)

الشبكات العصبية الاصطناعية (ANN) [20]

II-1-2- مبدا *QSPR* :

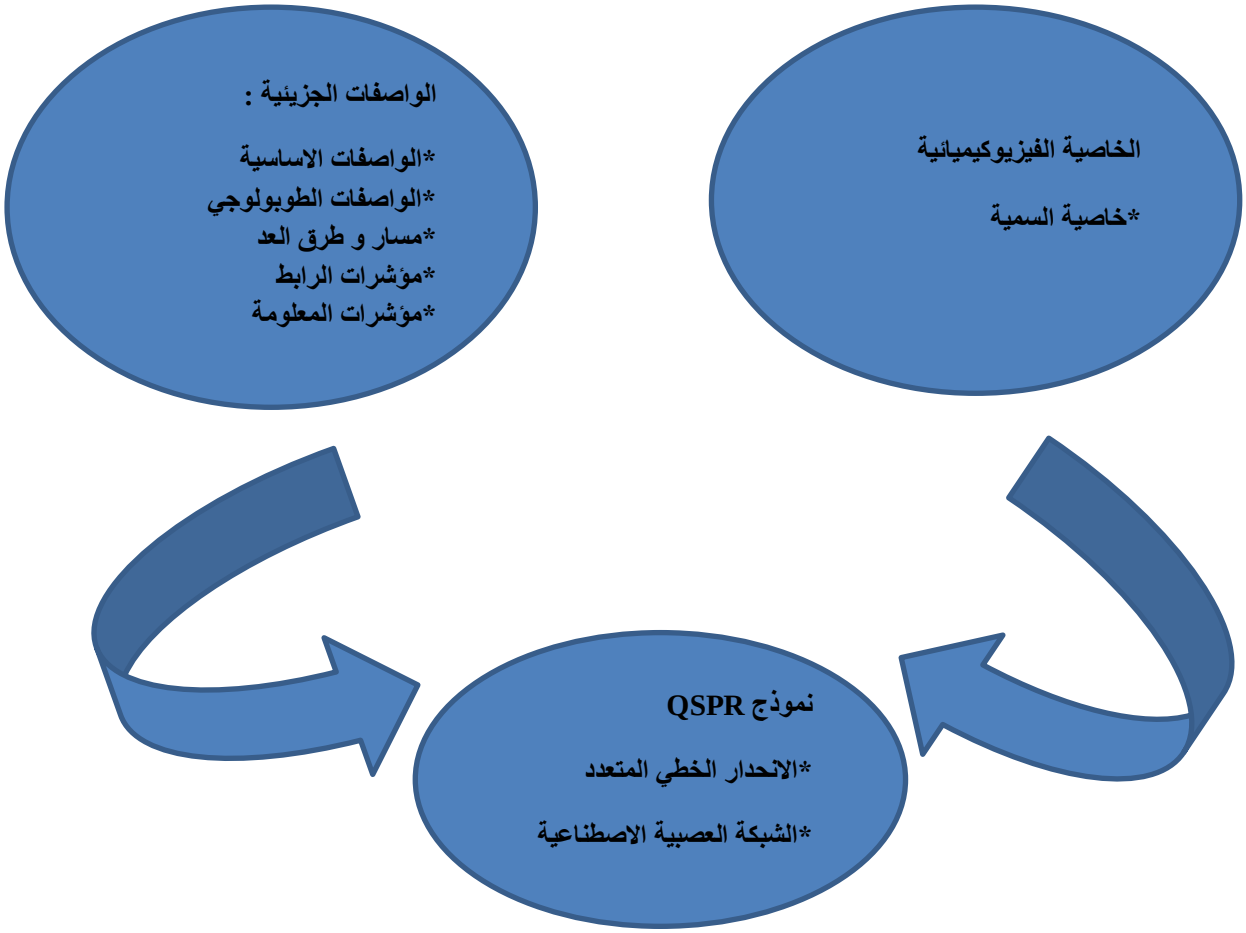
مبدأ طريقة QSPR هي إنشاء العلاقة الرياضية التي تربط كميا خصائص الجزيئات الإلكترونية والهندسية، تعرف بالواصفات الجزيئية (الفعالية البيولوجي، سمية، وخصائص الفيزيوكيميائية، وغيرها) لسلسلة من المركبات الكيميائية متشابهة باستخدام طرق تحليل المعطيات. والشكل العام للنموذج كالتالي :

(الواصفات) f= الخاصية

الهدف من هذه الطريقة هو تحليل البيانات البنيوية للكشف عن عوامل خاصية قياسها. وللقيام بذلك، يمكن استخدام الأدوات التالية :

- الانحدار الخطي.
- أشجار القرار.
- الشبكات العصبية.
- الخوارزميات الوراثية.

بمجرد إنشاء هذه العلاقة والتأكد من صلاحيتها، يمكننا أن نتنبأ بالخصائص جديدة للمركبات مجهولة، والتي هي غير معروفة تجريبيا.[11]



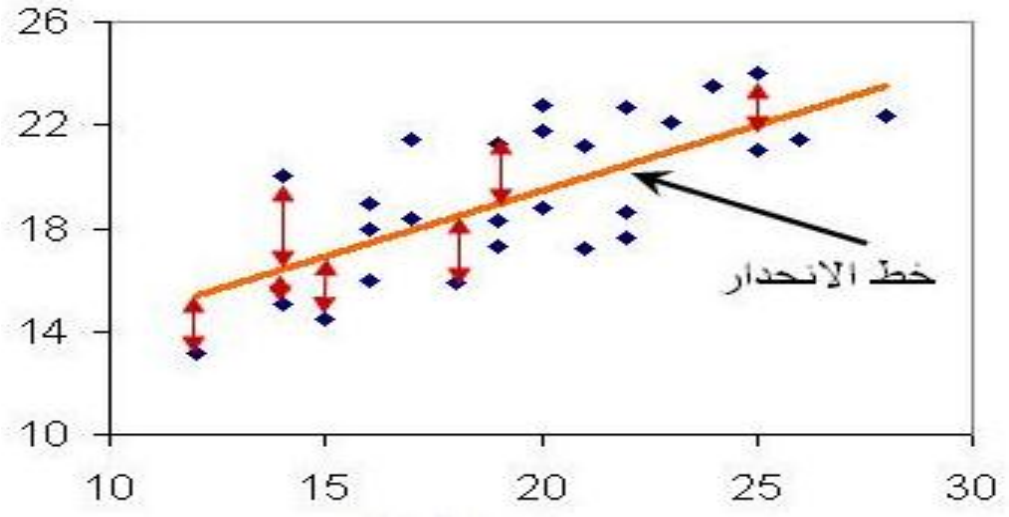
الشكل (II-1):مبدا طريقة QSPR المرجع [14]

II-2- تقنيات الانحدار الخطي المستعملة للحسابات الاحصائية:

يعد الانحدار الخطي المتعدد (MLR) و الانحدار المكون الرئيسي (PCR) و انحدار المربعات الصغرى الجزئي (PLS) و الشبكات العصبية الاصطناعية (ANN) من اكثر طرق النمذجة شيوعا في QSAR/QSPR [20].

MLR-1-2-II

الاسلوب الرياضي الاكثر استخداما هو تحليل الانحدار الخطي المتعدد (MLR) هذا النهج مباشر ينتج عنه نتيجة سهلة الفهم، ولهذا السبب ،يتم انشاء معظم نماذج QSAR/QSPR باستخدام تحليل الانحدار الخطي . يعد هذا الاخير طريقة قوية لإنشاء ارتباط بين المتغيرات المستقلة او التفسيرية (الواصفات الجزيئية X) و المتغير التابع او التوضيحي Y. [25]



الشكل (II-2): خط الانحدارالخطي

أي ان الانحدار يربط علاقة بين المتغير التابع الذي يمثل الخاصية الفيزيوكيميائية والعديد من المتغيرات المستقلة، والتي تتمثل في الواصفات الجزيئية .ويستخدم الانحدار الخطي للتنبؤ بقيمة المتغير التابع، من خلال تمثيل العلاقة بين .المتغير التابع والمتغيرات المستقلة [11]

II-2-1-1- صيغة حساب الانحدار الخطي المتعدد:

$$y_i = a + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_p x_{ip} + \epsilon$$

$i = n$

y_i=المتغير التابع

X_i=المتغيرات التفسيرية

a=تقاطع y– (مصطلح ثابت)

b_p=معاملات الانحدار لكل متغير توضيحي

ϵ= مصطلح خطا النموذج (المعروف ايضا باسم القيم المتبقية) [21]

في الواقع الأمر فإن هذه المعادلة واحدة من جملة معادلات يبلغ عددها (n) ويكون نظام المعادلات كالأتي :

$$\begin{cases} Y_1 = a + b_1X_{11} + b_2X_{21} + \dots + b_kX_{k1} + U_1 \\ Y_2 = a + b_1X_{12} + b_2X_{22} + \dots + b_kX_{k2} + U_2 \\ \cdot \quad \cdot \quad \cdot \cdot \quad \cdot \cdot \quad \dots \quad \cdot \cdot \quad \cdot \\ Y_n = a + b_1X_{1n} + b_2X_{2n} + \dots + b_kX_{kn} + U_n \end{cases}$$

هذه المعادلة تتضمن (k+1) من المعلومات المطلوب تقديرها علما بأن الحد الأول منها a يمثل الحد الثابت، الأمر الذي يتطلب اللجوء إلى المصفوفات والمتجهات لتقدير تلك المعلومات. وعليه يمكن صياغة هذه المعادلات في صورة مصفوفات كالأتي :

$$\begin{pmatrix} Y1 \\ Y2 \\ \cdot \\ Yn \end{pmatrix} = \begin{pmatrix} 1 & X11 & X21 & \dots & XK1 \\ 1 & X12 & X22 & \dots & XK2 \\ & \cdot & \cdot & \dots & \cdot \\ 1 & X1n & X2n & \dots & XKn \end{pmatrix} \begin{pmatrix} b1 \\ b2 \\ \cdot \\ bK \end{pmatrix} + \begin{pmatrix} U1 \\ U2 \\ \cdot \\ Un \end{pmatrix}$$

وباختصار يمكن كتابتها كالأتي:

$$\mathbf{Y = XB + U}$$

حيث ان :

Y :متجه عمودي أبعاده (n+1) ويحتوي على المتغير التابع

X :مصفوفة أبعادها (n × K+1)وتحتوي على المتغيرات المستقلة

B :متجه عمودي أبعاده (K+1×1) يحتوي على المعالم المطلوب تقديرها.

U : متجه عمودي أبعاده (n×1) يحتوي على الأخطاء العشوائية

يمكن استخدام الانحدار الخطي المتعدد في حالة توافر الشروط التالية :

1- أن تكون العلاقة خطية بين المتغيرات المستقلة والمتغير التابع .

2- أن تكون البيانات موزعة توزيعا طبيعيا للمتغيرات المستقلة والمتغير التابع .

3- يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل.

باستخدام الفروض السابقة تصبح أفضل معادلة لتقدير قيمة المتغير التابع بدلالة المتغيرات المستقلة هي:

$$\mathbf{Y_j = a + b_1\, X1_j + b_2 X_{2j} + \ldots + B_{k-1} X_{Kj}}$$

$$\mathbf{j=1,\ 2,\ldots\ n}$$

حيث ان

j :القيمة التقديرية للمتغير التابع

a :القيمة التقديرية لثابت الانحدار الخطي.

b1 : ميل الانحدار y على القيمة التقديرية لمتغير المستقل الأول

b2 :ميل الانحدار y على القيمة التقديرية لمتغير المستقل الثاني

X_{1j}:المتغير المستقل الأول

X_{1j}:المتغير المستقل الثاني

بعد الحصول على نتائج معادلة الانحدار يجب علينا أن نبين هل أن هذه المعاملات مقبولة من الناحية الإحصائية أي معنوية إحصائيا مع التنويه بأن المعنوية تكون لكل معامل على حده.

نعتبر معامل الارتباط الخطي مقياسا جيدا لدراسة العلاقة بين متغيرين و التي تاخذ الشكل الخطي سوف تقتصر دراستنا على الارتباط بين متغيرين (ظاهرتين) فقط

كما سيتم الحصول على إحصائيات تستخدم لمعرفة المعنوية الإجمالية للنموذج ومنها معامل الارتباط الخطي (r) معامل التحديد (R^2) الاختبار الإحصائي(F) .[11]

II-2-1-2- معامل الارتباط الخطي :

معامل الارتباط هو مقياس كمي نسبي يستخدم لتحديد نوع وقوة العلاقة بين متغيرين، ويرمز له بالرمز (r). [11]

II-2-1-3- نوع العلاقة :

تأخذ العلاقة بين المتغيرين ثلاث صور مختلفة حسب إشارة معامل الارتباط كما يلي:

إذا كانت إشارة معامل الارتباط سالبة ($r< 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الآخر، والعكس

إذا كانت إشارة معامل الارتباط موجبة ($r >0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الآخر، والعكس

إذا كان معامل الارتباط قيمته تساوى صفرا ($r = 0$) دل ذلك على عدم وجود علاقة (أي الارتباط) بين المتغيرين

II-2-1-4- قوة العلاقة:

يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن الصفر، حيث أن قيمة معامل الارتباط تقع في المدى ($-1 < r < 1$) وقد صنف بعض الإحصائيين درجات لقوة العلاقة التي يمكن تمثيلها على الجدول التالي:

الجدول (II-1): درجات قوة العلاقة

قيمة معامل الارتباط	المعنى
+1	ارتباط طردي تام
(من 0.90 الى 0.99)	ارتباط طردي قوي جدا
(من 0.70 الى 0.89)	ارتباط طردي قوي
(من 0.50 الى 0.69)	ارتباط طردي متوسط
(من 0.30 الى 0.49)	ارتباط طردي ضعيف
(من 0.01 الى 0.29)	ارتباط طردي ضعيف جدا
0	لا يوجد ارتباط
(من 0.01- الى 0.29-)	ارتباط عكسي ضعيف جدا
(من 0.30- الى 0.49-)	ارتباط عكسي ضعيف
(من 0.50 الى 0.69-)	ارتباط عكسي متوسط
(من 0.70- الى 0.89-)	ارتباط عكسي قوي
(من 0.90- الى 0.99-)	ارتباط عكسي قوي جدا
-1	ارتباط عكسي تام

II-2-1-5- شكل الانتشار :

شكل الانتشار هو عبارة عن تمثيل أزواج قيم الظاهرتين المتغيرين(X ,Y) بمجموعة من النقاط على الرسم. فيتم تمثيل المتغير X على المحور الأفقي، والمتغير Y على المحور الرأسي حيث نرصد لكل زوج من قيم المتغيرين بنقطة على الرسم فنحصل على شكل الانتشار.

II-2-1-6- قياس الارتباط :

يمكن قياس الارتباط بين المتغيرين بمعامل الارتباط الخطي بدلالة القراءات لبيانات المتغيرين باستخدام الصيغة التالية :

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\ \sqrt{n\sum Y^2 - (\sum Y)^2}}$$ حيث

n :عدد الأزواج قيم المتغيرين

$\sum XY$: مجموع حاصل ضرب Y في X

$\sum X$: مجموع قيم X

$\sum Y$: مجموع قيم Y

$\sum X^2$: مجموع مربعات قيم X

$\sum Y^2$: مجموع مربعات قيم y

II-2-1-7- معامل التحديد (R^2) :

أما R^2 فهو يسمى بمعامل التحديد والذي يستخدم لمعرفة القوة التفسيرية للنموذج المقدر (المعادلة المقدرة) في حالة الانحدار الخطي المتعدد.

تعريف

R^2 يعرف معامل التحديد بأنه مربع معامل الارتباط ويرمز له بالرمز R^2 حيث $R^2 = r^2$ وهو معامل يستخدم لمعرفة مدى جودة التوفيق لخط الانحدار المقدر للمتغير التابع على المتغير المستقل، وبما أنه عبارة عن مربع معامل الارتباط لذلك فقيمته تكون موجبة محصورة بين الصفر والواحد أي $0 \leq R2 \leq 1$

هو يقيس نسبة التغير في المتغير التابع نتيجة تغير المتغير المستقل، أي أنه يوضح نسبة مساهمة المتغير المستقل في التغير الحاصل في المتغير التابع

- يكون التقدير جيدا كلما اقتربت قيمة R^2 من الواحد.
- اذا كان $R2 = 1$ يعني أنه يوجد علاقة خطية تامة بين المتغيرين التابع والمستقل أي ان المتغير المستقل يساهم بنسبة 100 %في التغير الحاصل في المتغير التابع
- كلما اقترب $R2$ من الصفر كلما دل ذلك على ضعف مساهمة المتغير المستقل في تفسير التغير الحاصل في المتغير التابع

يمكن إيجاد معامل التحديد إما من خلال معامل الارتباط أو من خلال ما يلي[11]

$$\mathbf{R^2 = \frac{\sum_{i=1}^{n} (Y1 - Y)^2}{\sum_{i=1}^{N} (Yi - Y)^2}}$$

حيث ان

Y1 :القرمة التقديرية لمتغير التابع

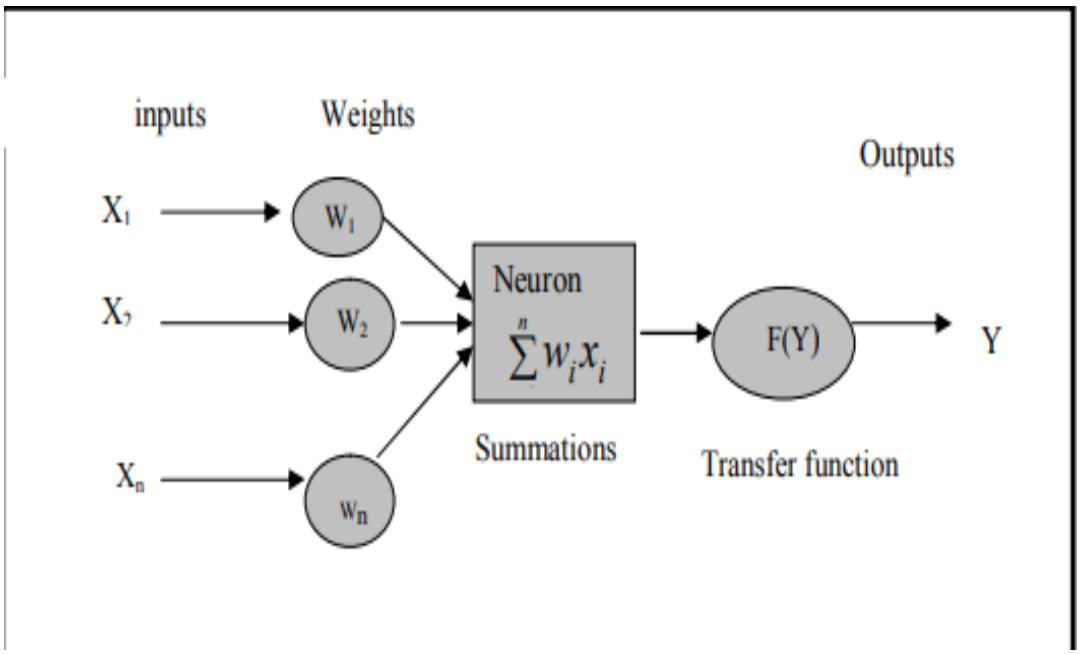
Yi :المتغير التابع.

y: متجه البعدn [11]

ANN–2–2–II :

تعد الشبكات العصبية الاصطناعيةANN Networks Neural Artificial) أحد أهم طرق الذكاء الاصطناعي ، و تتمحور فكرتها حول محاكاة قدرة العقل البشرى على التعرف على الأنماط ، وتمييز الأشياء ، باستخدام الحاسب الآلي من خلال إتباع عملية التعليم الذاتية التي تحدث في العقل ، والتي يتم فيها الاستفادة من الخبرات السابقة في سبيل الوصول إلى أفضل نتائج في المستقبل.[22]

الشبكة العصبية الاصطناعية (ANN) هي جزء من نضام حوسبي مصمم لمحاكاة طريقة تحليل الدماغ البشري للمعلومات و معالجتها . انه اساس الذكاء الاصطناعي و يحل المشكلات التي قد تكون مستحيلة او صعبة وفقا للمعايير البشرية او الاحصائية . تمتلك ANN قدرات التعلم الذاتي التي تمكنها من تحقيق نتائج افضل مع توفر المزيد من البيانات . [23]



الشكل (II–3): وحدة تشغيل داخل الشبكة العصبية

والشكل السابق يبين وحدة تشغيل بسيطة داخل الشبكة العصبية حيث يكون لكل وحدة تشغيل مسار إدخال واحد أو أكثر تنحصر مهمته في نقل المعلومات من العالم الخارجي إلى وحدة التشغيل الذي تقوم بدورها بعملية التجميع البسيطة ، ثم تحويل المعلومات بواسطة دالة تفعيلية تعرف بدالة ال تحويل، بعد ذلك تنقل المعلومات كمخرجات من خلال مسار المخرجات . وبشكل رياضي فان كل خلية عصبية تستقبل عدد من إشارات المدخلات Input signals (X1, X2,… Xn,) والتي تماثل التفرعات العصبية Dendrites،ثم يتم ضرب كل المدخلات في الأوزان (W1, W2,…., Wn,) التي تعد الوسيلة الأساسية لذاكرة الشبكات العصبية الاصطناعية طويلة المدى وتعبر عن درجة أهمية المدخلات ،، وحاصل الجمع summations يتم معالجته باستخدام دالة تحويل F (Y) Transfer function لنحصل منها على إشارات المخرجات (Y) Output Signals [22]

II-2-2-1- *الهيكل العام للشبكة :*

عملية تنظيم الخلايا العصبية في طبقات وكيفية الاتصال بين هذه الخلايا لتكوين الشبكة تسمى بهيكل الشبكة Architectures ،و بوجه عام من الممكن تقسيم هيكل الشبكة العصبية الاصطناعية إلى ثلاثة أنواع رئيسية هي:

- شبكة وحيدة الطبقة ذات التغذية الأمامية
- شبكة متعددة الطبقات ذات التغذية الأمامية
- شبكة متعددة الطبقات ذات التغذية المرتدة [22]

II-2-2-2- صيغة حساب الشبكات العصبية الاصطناعية :

اولا يتم حساب صافي المدخلات وفقا للمعادلة

$$Net_j = \sum_{i=1}^{m} w_{ji}\, x_i$$

وبعد ذلك يتم وضع Net كوسيلة في دالة النقل :

$$[24]\quad y_j = out_j = 1/\{1 + \exp[-\alpha_j(Net_j + \theta_j)]\}$$

II-3-النموذج الإحصائي:

نجد ان Ratcliffe وضع تعريف عن النموذج الذي وصفه على انه "اعادة بناء مبسط للوضع الحقيقي الذي يقلل من مستوى التعقيد فيه ليستطيع المخطط ادراكه وبشكل كاف لتذليل المصاعب"

فالنموذج الاحصائي هو توسع للنماذج التنبؤية تهدف إلى التحقق من صحة وضع نظام ما عن طريق مقابلة نتائج بالواقع الذي يتم تسجيل الإحصاءات عنه.[15]

لذا يمكن القول أن النماذج الإحصائية هي نوع خاص من النماذج الرياضية التي تستخدم كبيانات مصدر ليست فقط البيانات الحالية فقط على الحالة الحالية للكائن، ولكن أيضا البيانات التي تميز حالة أي كائنات أخرى لهذه الفئة .[16]

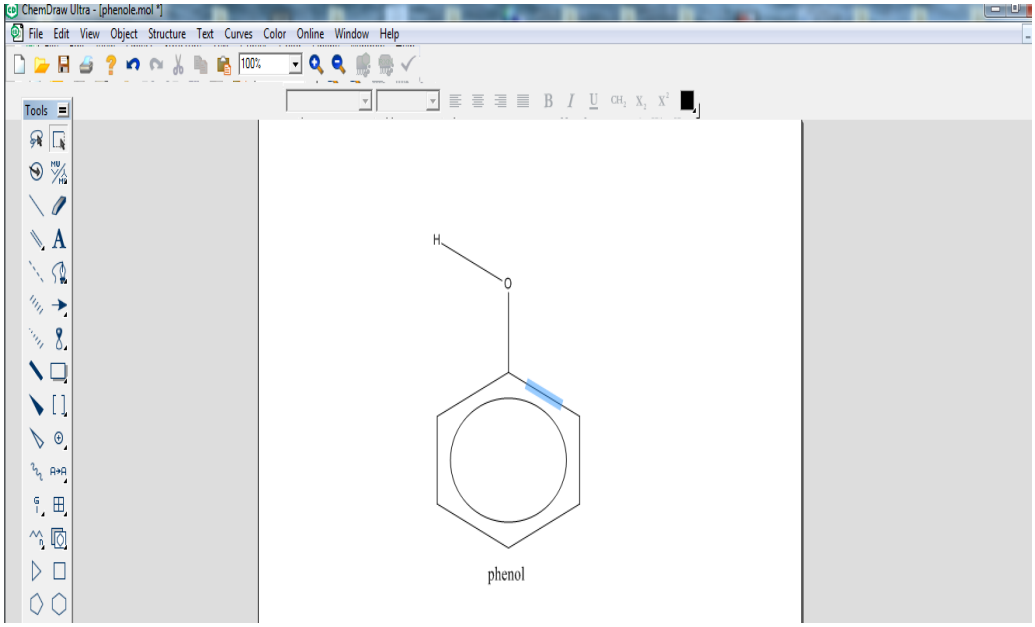
II-4-البرامج المستعملة :

ان رسم الصيغة الكيميائية وخطط التفاعل مهمة متكررة للكيميائيين في جميع مستويات تعليمهم. بينما يتم استخدام الرسم اليدوي بكفاءة اثناء المناقشات والتعلم، الا ان الرسومات الانيقة مطلوبة للتقارير الرسمية و المنشورات و الاطروحات.

يمكن انشاء مثل هذه الرسومات باستخدام العديد من برامج الكمبيوتر، و من الامثلة على ذلك chemdraw

II-4-1- Chemdraw :

هو برنامج سهل الاستخدام يسمح برسم تمثيلات ثنائية الابعاد بسيطة بشكل حدسي وفعال للجزيئات العضوية، وهي متوفرة لاجهزة الكمبيوتر PC وكذلك لمنصة Mac.[29]



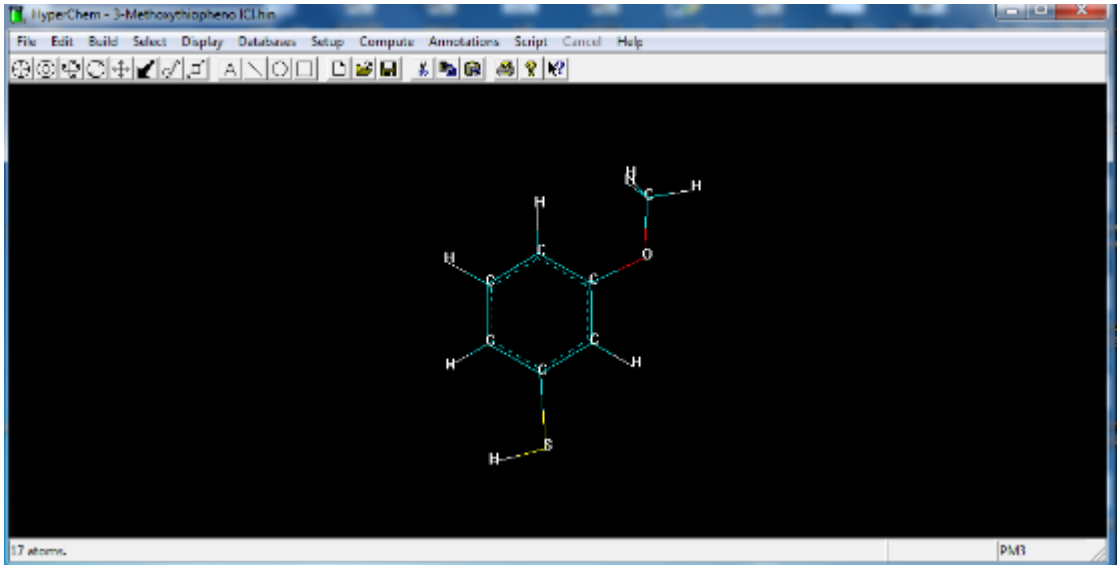
الشكل (II-4): برنامج ChemDraw

تحت مصطلح النمذجة الجزيئية، نجد تقنيات مختلفة لتصور الهياكل الجزيئية و معالجتها و تحليلها و حسابها . من الناحية التخطيطية، يميز المرء تقنيات الرسومات الجزيئية التي تسمح بتمثيل الهيكل ثنائي الابعاد او ثلاثي الابعاد للجزئ على الشاشة و من الامثلة على ذلك HyperChem.

II-4-2- *HyperChem* :

هو أحد البرامج المتقدمة في مجال النمذجة الجزيئية ذو قدرات حسابية عالية تتعلق بالأنظمة الجزيئية البسيطة والمعقدة، وهو قادر على رسم الجزيئة ببعدين وثلاثة أبعاد وبإحداثيات خطية وصولا إلى أفضل شكل مستقر للجزيئة .

برنامج (HyperChem) عبارة عن تقنية من تقنيات النمذجة الجزيئية، وفي هذا الجزء من العمل استعملنا النمذجة الجزيئية التي تتضمن طرق الحساب النظرية (الميكانيكا الجزيئية (MM^+)والطرق شبه عملية (PM3)) التي تمكننا من تحديد هندسة الجزيئات وتقدير الخصائص الفيزيوكيميائية.[11]



الشكل (II-5): برنامج *Hyperchem*

II-3-4- *DRAGON* :

برنامج (DRAGON) عبارة عن تقنية مطبقة لحساب الواصفات الجزيئية، أنجز من طرف العالم (Milano) ومجموعة بحث(QSPR , QSAR) هذه الواصفات تستعمل لتقييم الجزئي (النشاط البنيوي أو خصائص البنيوية) بالإضافة إلى تحليل التشابه وفحص قواعد البيانات الجزيئية[19] .

- و قد تم تصميم (DRAGON) باعتباره برنامج سهل الاستعمال الذي يتولى إجراء العمليات الحسابية للواصف وفقا لتسلسل منطقي بسيط
- الواصفات الجزيئية الناتجة عن برنامج(DRAGON) تستعمل لحساب وتقدير الخصائص الفيزيوكيميائية
- إن القدرة الفعلية لبرنامج(DRAGON) يمكنه من حساب حتى 250251 جزيئة
- الإصدار الأول لبرنامج(DRAGON) يعود إلى سنة 1997 ، و في الوقت الراهن أضيفت له عدة واصفات جزيئية جديدة.
- صمم برنامج (DRAGON) للاستعمال في كلا النظامينWindows و Linux حيث أن هناك إصدارات في Windows يسميان (DRAGON Professional) و DRAGON PLUS) (وإصدار واحد في Linux ويدعى (DRAGON X) .

- يستطيع برنامج(DRAGON) حساب عدة أنواع من الواصفات الجزيئية كنوع الذرة، المجموعات الوظيفية وكذلك الواصفات الطوبولوجية والهندسية وبعض خصائص الجزيئات الفيزيوكيميائية مثل Log Kow ،الشحن الجزيئية، المرونة، الانكسارية، عدد الروابط القابلة للتدوير، السطوح والحجوم الجزيئية، وكذا الأبعاد بين مختلف ذرات الجزيئة.[19]

الشكل(II-6):برنامج Dragon

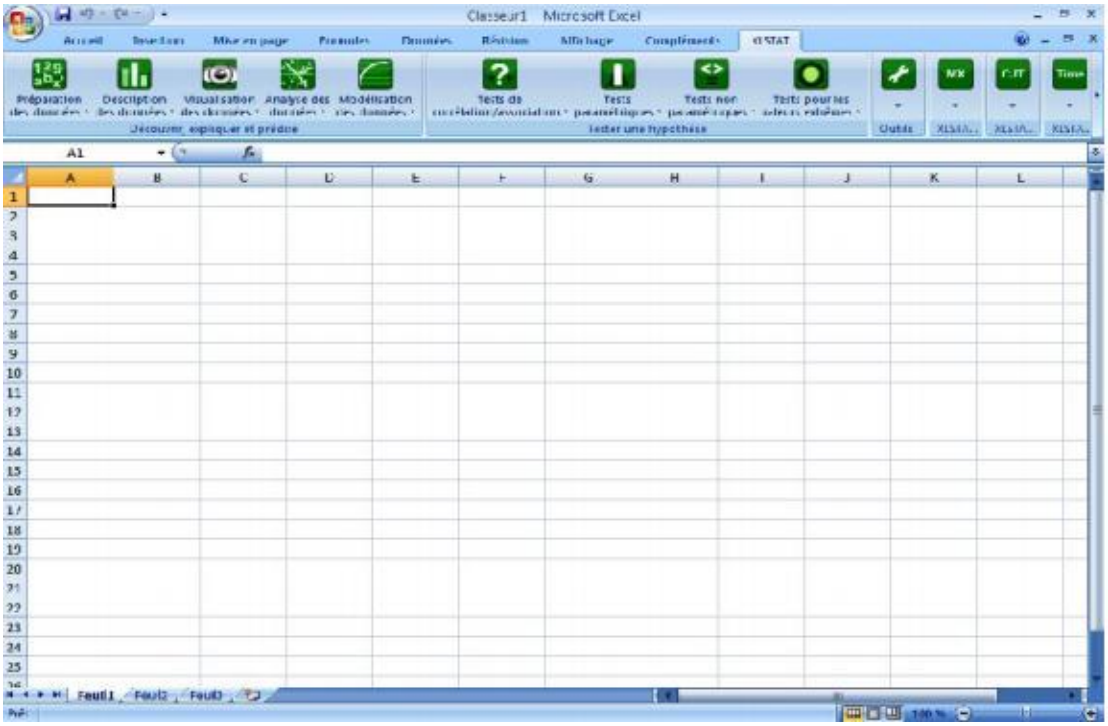
II-4-4- *برنامج (XLSTAT):*

هو برنامج يقوم بتحليل البيانات إحصائيا وهو تابع لبرنامج Excel Microsoft تطور بفضل Thierry Fahmy ،مؤسس Addinsoft .البرنامج متاح على شبكة الإنترنت منذ عام 1996.

نختار هذا البرنامج لأنه يحتوي على العديد من الميزات:

- XLSTATهو برنامج موثوق به جدا ومستقر جدا
- يمكننا من إعادة تنظيم هيكل قاعدة البيانات وتسهيل تحليلها

- وفي XLSTAT نستطيع تحليل المكونات الرئيسية (PCA) وذلك بطريقة الانحدار الخطي (بسيطة أو متعدد)
- يمكننا XLSTAT من تقييم جودة البيانات، ويحدد أخطاء حسابية تساعدنا على تحسين عملنا.
- يتضمن XLSTAT بعض الأدوات التي تسهل التصور من البيانات
- XLSTAT يقدم تشكيلة واسعة من الاختبارات الإحصائية التي تسمح لنا باختبار الفرضيات[11]

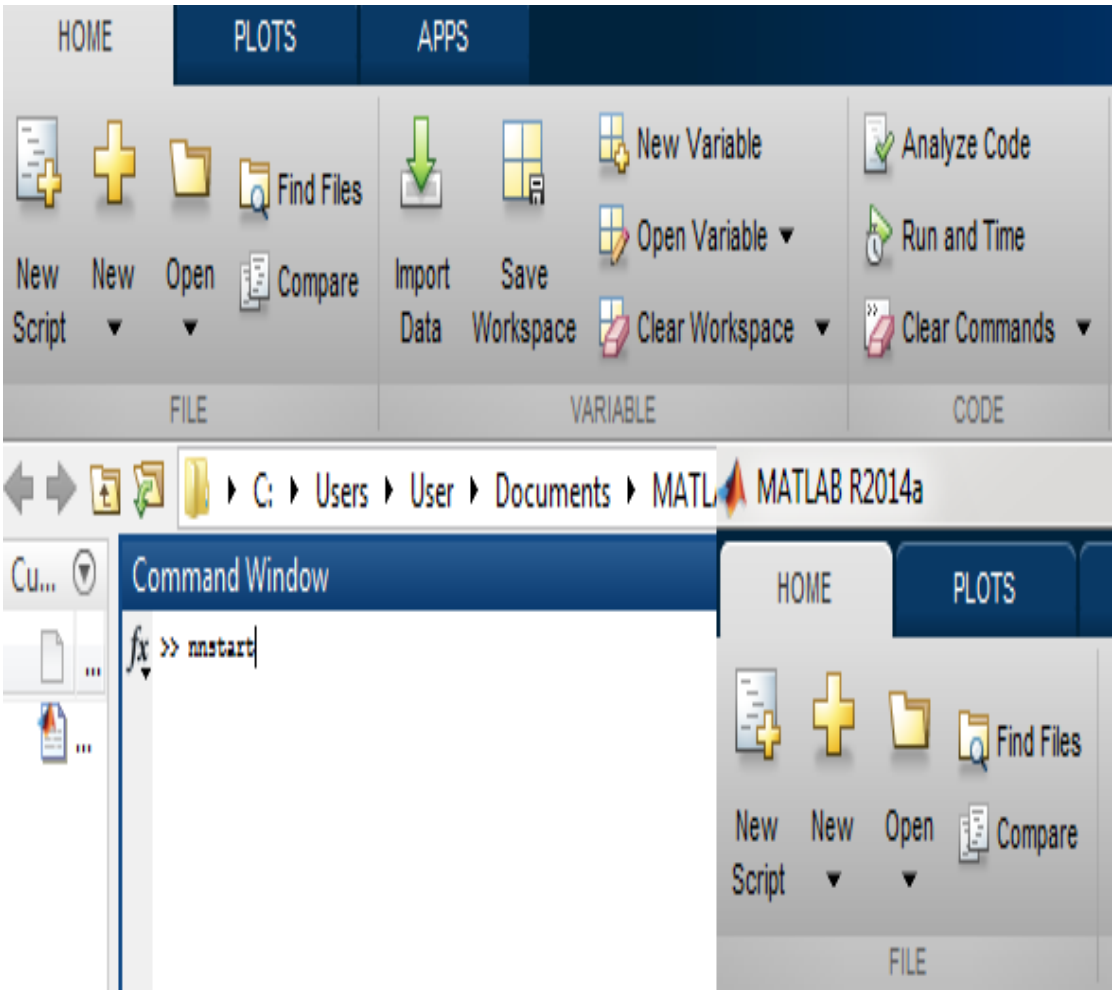


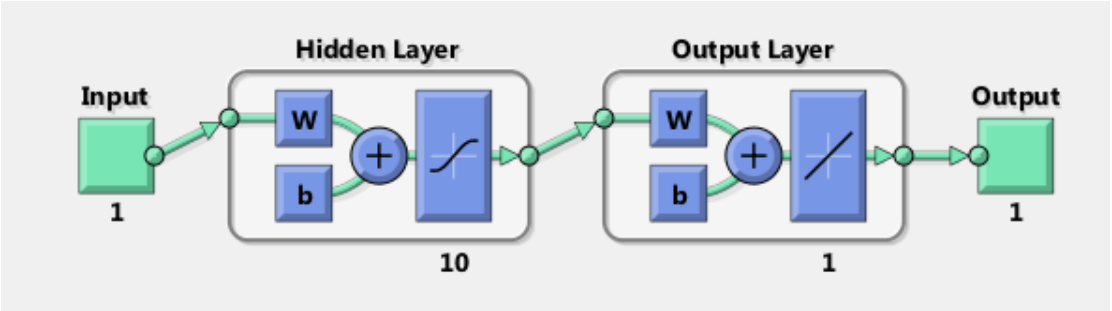
الشكل(II-7): برنمج Xlstat

II-4-5- برنامج الماتلاب (Matlab) لحساب الشبكات العصبية ANN :

يمكن برمجة الشبكات العصبية باستخدام لغات البرمجة المعروفة مثل لغة سي C او TOOLs مثل الماتلاب [27]

و سيتم إستخدام الماتلاب (MATLAB) في دراستنا لتطبيق شبكة عصبية إصطناعية والسبب في إختياره هو سهولة تطبيق الشبكات به من خلال مجموعة من الخطوات البسيطة والتي لا تحتاج إلي معرفة قوية بالبرمجة ولذلك يُتيح للعديد من الباحثين تطبيق الشبكات بأنفسهم حتي وإن كان لديهم خلفية ضعيفة عن البرمجة[28].





الشكل (II-8): برنامج Matlab

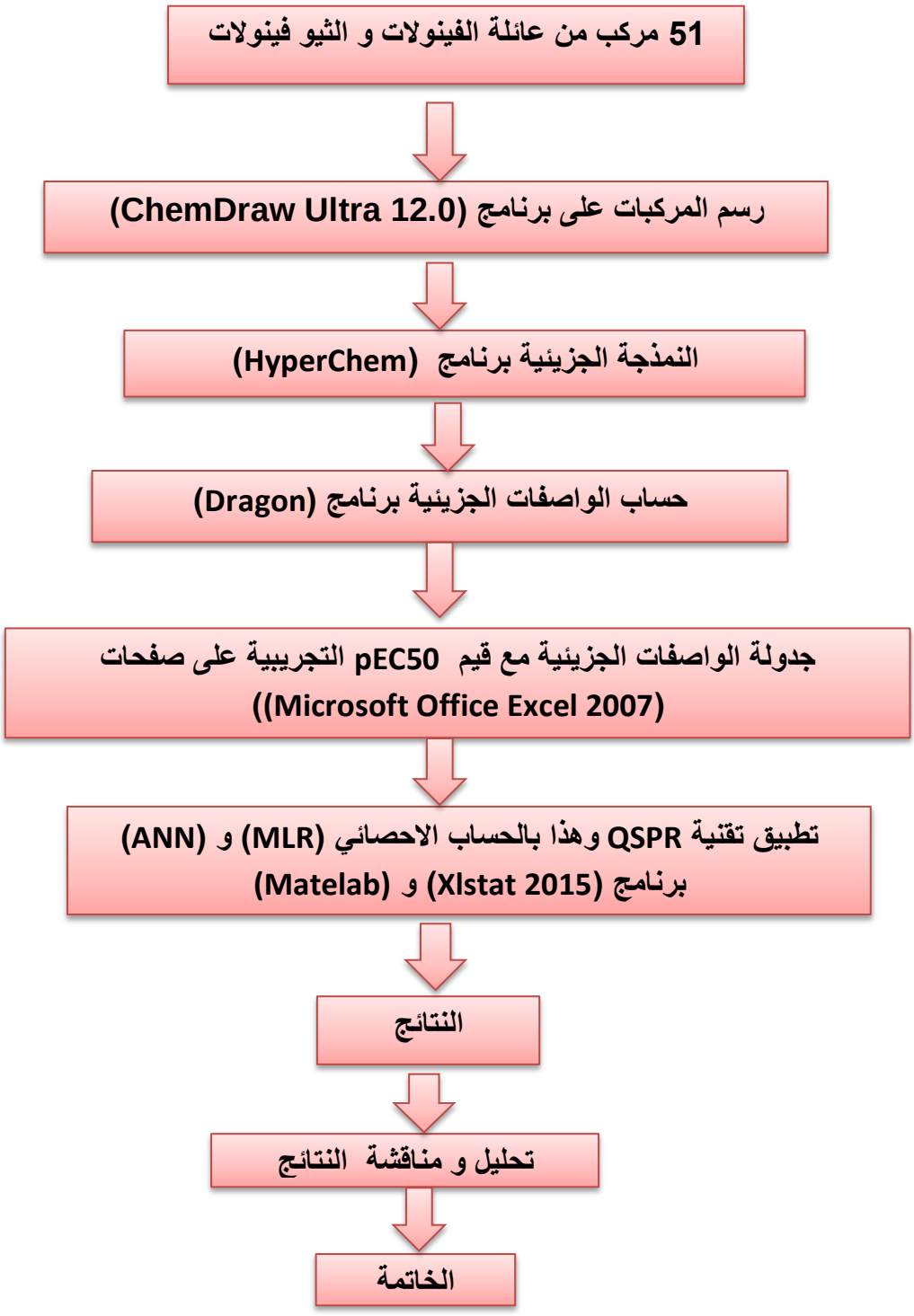
خاتمة :

في نهاية هذا الجزء، تمكنا من خلاله ان نسلط الضوء على كل الجوانب المتعلقة بالتنبؤ بسمية المركبات الكيميائية عن طريق تحليل QSPR بالادوات الاحصائية ANN وMLR , حيث عرضنا كل المفاهيم و العلاقات و الطرق و البرامج المتعلقة بدراستنا ، في الجانب النضري لكي تساعنا في الجانب التطبيقي .

الجزء التطبيقي

الفصل الثالث
قاعدة البيانات مع تحليل و مناقشة النتائج

III-1- الخطوات التجريبية للعمل :



الشكل (III-1):الخطوات التجريبية للعمل

جدول (**III-1**): يوضح الجدول التالي قائمة قاعدة البيانات

الرقم	اسم المركب	السمية(**PEC50**)
01	2,3,5,6-Tetra fluorothiophenol	4.86
02	2,3-Dichlorothiophenol	4.91
03	2,3-Dimethylphenol	3.6
04	2,4,6-Trinitrophenol	2.51
05	2,4-Dichlorophenol	4.01
06	2,4-Dichlorothiophenol	5.59
07	2,4-difluorophenol	2.33
08	2,4-Difluorothiophenol	5.15
09	2,4-Dimethylthiophenol	4.77
10	2,5-Dichlorothiophenol	5.16
11	2,5-Dimethylthiophenol	4.66
12	2,6-Dichlorothiophenol	4.99
13	2,6-Dimethylthiophenol	4.64
14	2-Amino-4-chlorothiophenol	5.37
15	2-aminothiophenol	4.72
16	2-Bromothiophenol	4.88
17	2-chlorothiophenol	4.9
18	2-Fluorothiophenol	4.78
19	3,4-Dimethoxythiophenol	4.72
20	3,4-Dimethylthiophenol	4.95
21	3,5-Dichlorothiophenol	5.11
22	3-Bromothiophenol	4.57

23	3-chlorothiophenol	5.03
24	3-fluorothiophenol	5.06
25	3-Methoxythiophenol	4.26
26	4-Aminothiophenol	4.66
27	4-Bromothiophenol	5.6
28	4-chlorothiophenol	4.99
29	4-fluorophenol	2.22
30	4-Fluorothiphenol	4.97
31	4-tert-Butylcachol	5.87
32	4-Tertiary butythiophenol	5.34
33	Catechol	3.14
34	Hydroquinone	3.14
35	m-Cresol	3.31
36	m-Methylthiophenol	4.6
37	m-Nitrophenol	3.31
38	o-Aminophenol	3.34
39	o-chlorophenol	3.43
40	o-cresol	3.35
41	o-Methilthiophenol	4.49
42	o-Nitrophenol	3.48
43	p-Aminophenol	3.27
44	p-chlorophenol	3.88
45	P-Cresol	3.71
46	Pentafluorothiophenol	4.34
47	Phenol	2.72
48	P-Methylthiophenol	5.89

49	p-Nitrophenol	3.72
50	Resorcinol	2.22
51	Thiophenol	5.71

III-2- شرح الخطوات التجريبية للعمل :

لإنجاز هذا العمل قمنا بعدة حسابات تجريبية علمية وهذا باستعمال برمجيات مختلفة، وقد تضمن هذا العمل مجموعة من الخطوات والمتمثلة في التسلسل التالي:

1-اخترنا 51 مركب من عائلة الفينولات و الثيوفينولات مع القيم التجريبية للخاصية الفيزيوكيميائية المختارة وهي السمية من المراجع العلمية .

2- رسم البنية الجزيئية لهذه المركبات على برنامج CHEMDRAW ULTRA 12.0 و يحفظ الملف بشكل (MDL MOLFILE (*.MOL

3- فتح الملف في برنامج النمذجة الجزيئية HYPERCHEM لحساب طاقات الاستقرار وذلك بطرق (الميكانيكا الجزيئيةMM+ والطرق نصف التجريبية PM3)ويحفظ الملف بشكل (*.HIN)

وقد أجريت النمذجة ببرنامج (HYPERCHEM.0.7)

4-حساب الواصفات الجزيئية باستعمال برنامج DRAGON وذلك باختيار 5 مجموعات (كتل) ثنائية الأبعاد (2D)

-وقد أجريت الحسابات ببرنامج (4.5).DRAGON

5-إنشاء جداول تجمع بين الخاصية المختارة PEC50 والواصفات الجزيئية التي تم حسابها وذلك على صفحات برنامج EXCEL و هذا لكل كتلة .

وقد أجريت الجدولة على برنامج (MICROSOFT OFFICE EXCEL 2007)

6- بعد تكملة الجدول نقوم ب:

* حذف كل الواصفات الثابة في نفس العمود (0=)

* حذف كل الواصفات الجزيئية التي تشكل دالة خطية مع الواصفات الجزيئية الأخرى تسمى هذه الطريقة بإزالة الارتباطات الخطية.

بعد هذه الخطوة نتحصل على ما يلي :

*الجدول (**III-2**): نتائج حساب الواصفات الجزيئية المكونة لقاعدة البيانات باستعمال Dragon*

الكتل	عدد الواصفات في الكتلة	عدد الواصفات بعد الحساب
الواصفات الاساسية	48	29
الواصفات الطوبولوجية	119	84
مسار و طرق البعد	47	40
مؤشرات الرابط	33	33
مؤشرات المعلومات	47	45

7- تطبيق تقنية QSPR وهذا للحساب الإحصائي بطريقة الانحدار الخطي المتعدد (MLR) باستخدام البرامج الإحصائي Xlstat وهذا لدراسة درجة الإرتباط وحساب قيم معاملات التحديد (R^2) التي تدلنا على درجة الألفة بين القيم التجريبية pEC50 والقيم المحسوبة (النظرية) لـ pEC50 بالإضافة إلى نوعية وشكل معادلة النموذج .

إن معامل التحديد يأخذ القيم بين (0 ،1)حيث كلما اقترب معامل التحديد إلى 1 يعني هذا وجود علاقة جيدة بين الخاصية المختارة والواصف، وذلك يؤدي إلى الوصول لنموذج أفضل.

-وقد أجريت الحسابات ببرنامج (Xlstat2015)

و طريقة الشبكات العصبية الاصطناعية (ANN) تم اجراء التجربة على nnstart المقدم من MATlab لدراسة الانحدار و نوعية وشكل النموذج الاحصائي .

8-تم تصميم معادلة النموذج الرياضي الخاص لكل كتلة على حدى والذي يربط بين الخاصية المختارة ومجموعة الواصفات الجزيئية المرتبطة بها .

9- يتم اختيار أفضل نموذج رياضي للانحدار الخطى المتعدد على أساس أعلى قيمة لمعامل التحديد (R^2) .

10- يتم اختيار افضل نموذج رياضي للشبكة العصبية الاصطناعية على اساس قيمة (R) .

III-3- النتائج

مقدمة:

في هذا العمل استعملنا تقنيتين لإنشاء نماذج QSPR على مجموعة اختبار مكونة من 51 مركب من عائلة الفينولات و الثيوفينولات. استخدمنا 5 كتل من الواصفات الجزيئية 2D الاساسية(constitutionnal)، الطوبولوجية (Topological descriptors)، مسار وطرق العد (Walk and path counts)، مؤشرات الرابط (Connectivity indices)، مؤشرات المعلومة (Information indices).

*التقنية الاولى وهي الانحدار الخطي المتعدد (MLR). تمت الحسابات خطوة خطوة (Wise step) بمساعدة برنامج XLSTAT اصدار 2015، استنادا على مصفوفات الواصفات الجزيئية و المحسوبة باستخدام برنامج , DRAGONحيث اخذنا pEC50 كمتغير تابع بينما المتغيرات المستقلة هي الواصفات المحسوبة والتي تختلف من كتلة الى اخرى. تم بناء (حساب) النموذج باستعمال 39 مركب و12 مركب للتحقيق (30% . Validation)

تم قياس دقة النموذج بمعامل التحديد R^2 و قيمة الخطأ (MSE) و معامل فيشر(F).

*التقنية الثانية وهي الشبكات العصبية الاصطناعية(ANN) حيث استعملت هذه التقنية بمساعدة برنامج Matlab اصدار 2014 و كذلك استنادا على نفس مصفوفات الواصفات المحسوبة ،حيث تكون مجموعة البيانات المؤخوذة 70% من اجمالي عدد المركبات و 15% للتحقيق (Validation) و 15%للاختبار (test).

*III-3-*1-نتائج التقنية الاولى: سجلت النتائج على الجدول التالي :

الجدول (*III* -*3*): نتائج الكتل المدروسة

البعد	اسم الكتلة	R^2	R^2 ajusté	MSE
2D	الواصفات الاساسية	0.761	0.725	0.305
	الواصفات الطوبولوجية	0.907	0.883	0.115
	مسار و طرق العد	0.086	0.061	0.850
	مؤشرات الربط	0.704	0.678	0.336
	مؤشرات المعلومات	0.034	0.008	0.640

انطلاقا من الجدول السابق نلاحظ ان كتلة الواصفات الطوبولوجية (***Topological descriptors***) لها اكبر قيمة لمعامل التحديد R^2=0.907 و R^2Ajusté =0.883

اي ان افضل علاقة تم الحصول عليها باستخدام هذه الطريقة

III-3-1-1-كتلة الواصفات الطوبولوجية (Topological descriptors) :

صيغة معادلة النموذج:

الشكل العام لمعادلة النموذج تكون على النحو التالي :

$$Y = a + \Sigma\ (bj \times xj)$$

a: قيمة الثابت

xj :الواصفة

bj :قيمة المعامل

معادلة النموذج :

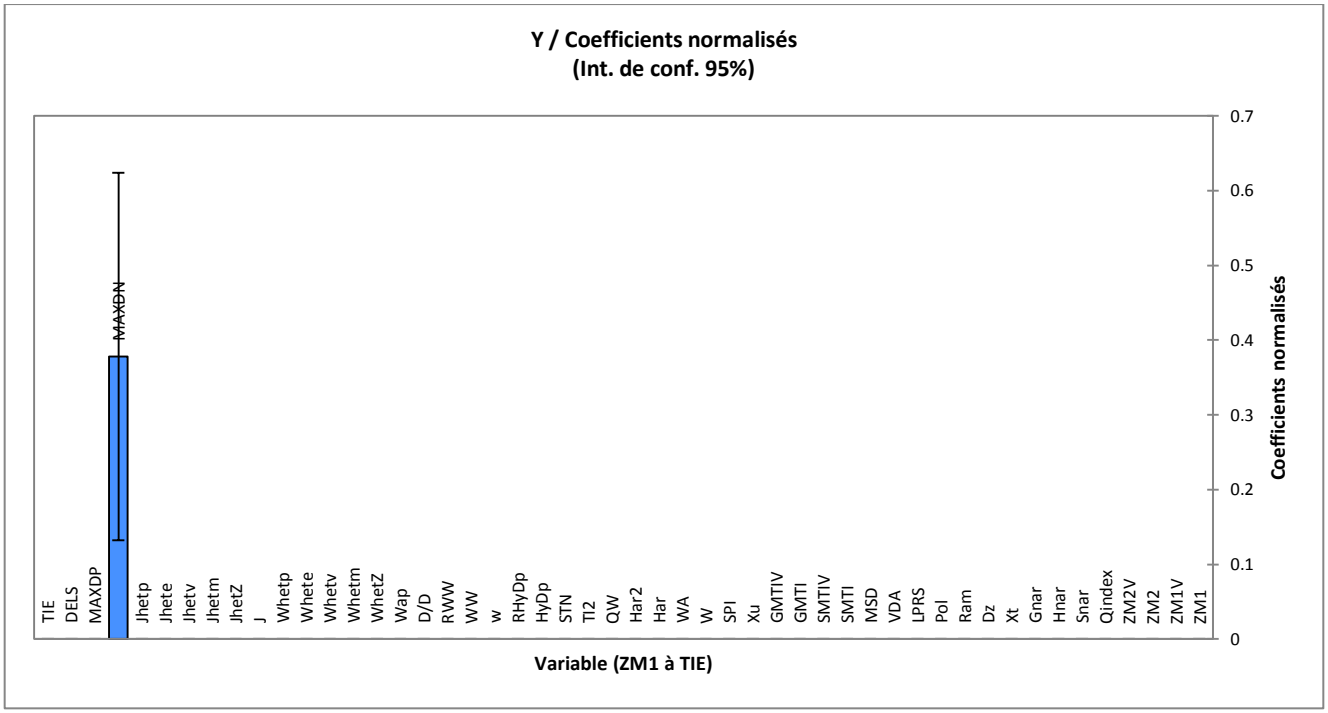
$$Y = -6.71375101258288+0.492581375998827*MAXDN+6.9806868189401*BLI-1.91369932080102*PW2+1.72390291041734*PW4+1.05184206699936*AECC+0.271878935393334*T(N..S)-1.76569047670135E-02*T(O..O)+0.129523034955461*T(O..F)$$

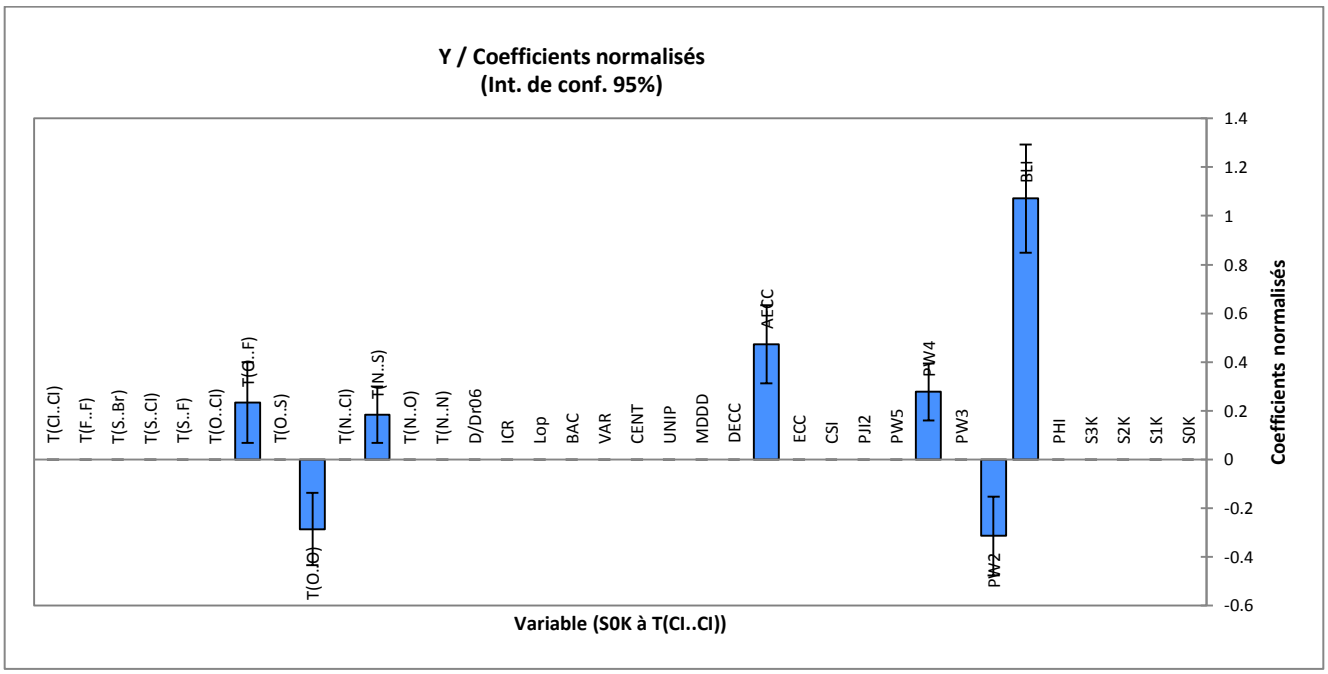
الجدول III-4) معاملات التعديل الخاصة بالكتلة (الواصفات الطوبولوجية)

Observations	Somme des poids	DDL	R^2
39.000	39.000	30.000	**0.907**
R^2 ajusté	RMCE	MAPE	*DW*
0.883	0.340	5.181	**1.974**
Cp	AIC	SBC	PC
9.000	-76.453	-61.481	**0.148**

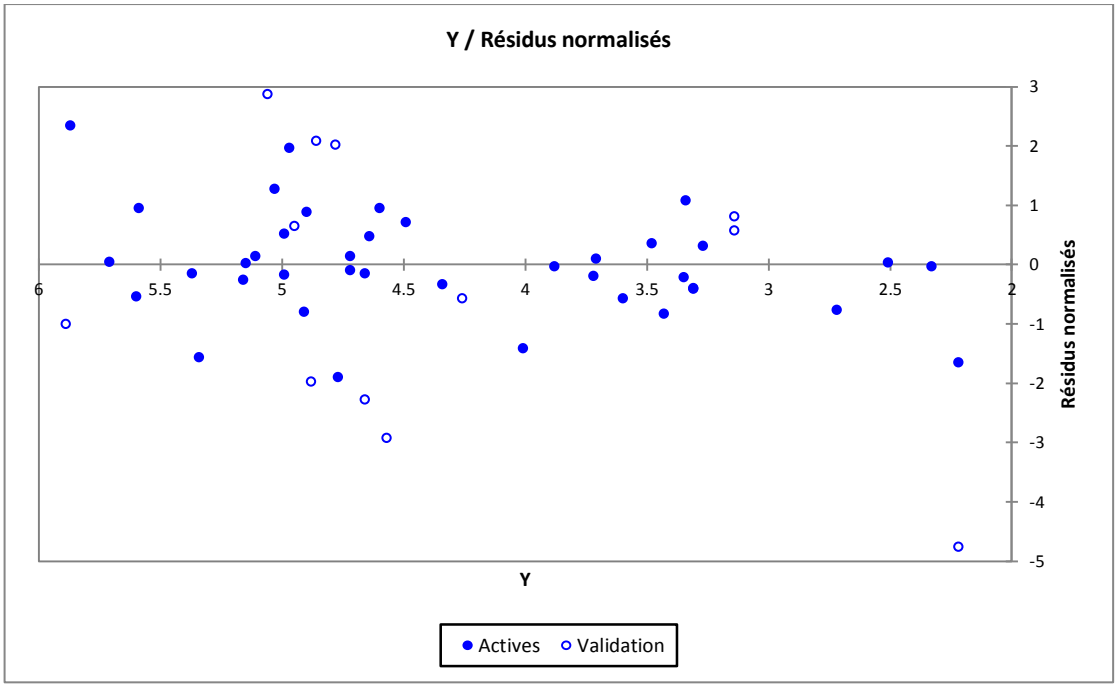
الجدول (III-5): تحليل الفروق الخاصة بالكتلة

Pr > F	F	Moyenne des carrés	Somme des carrés	DDL	Source
< 0.0001	36.747	4.240	33.920	8	**Modèle**
		0.115	3.461	30	**Erreur**
			37.381	38	**Total corrigé**

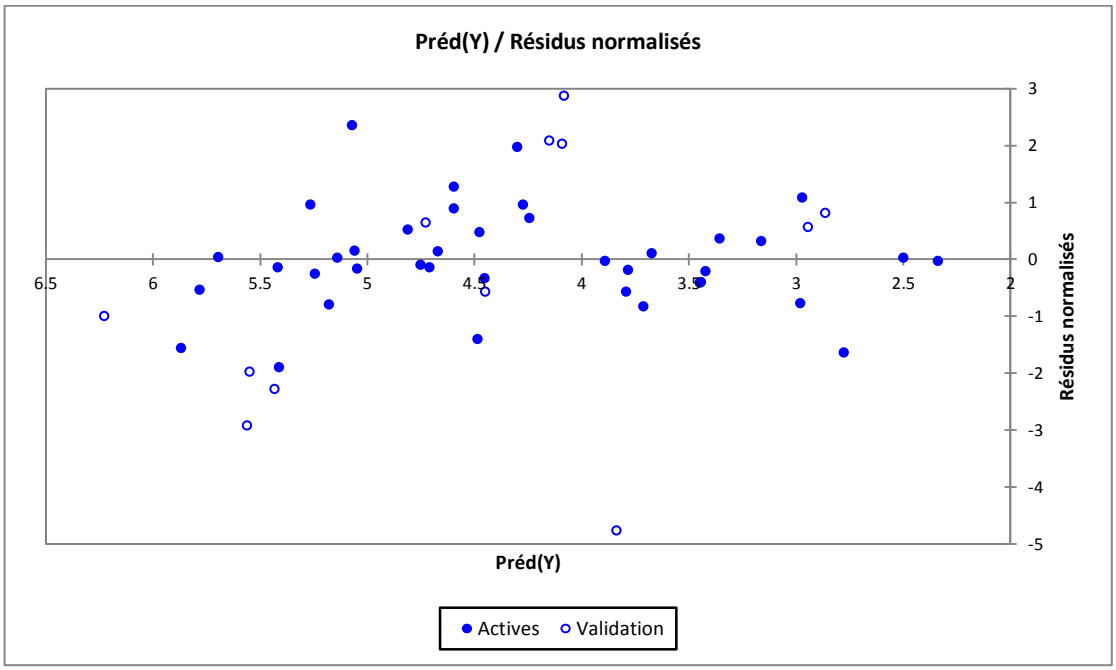




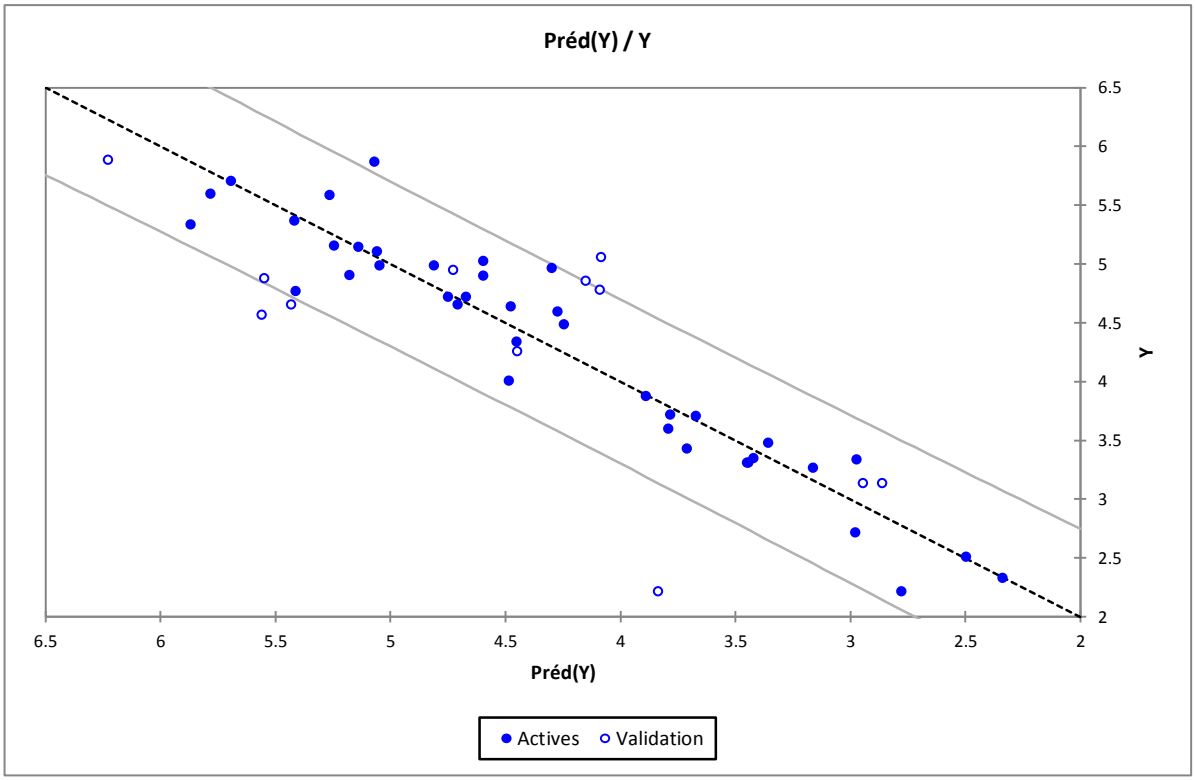
الشكل (*III*-2) بيان يوضح اسماء معاملات الواصفات الجزيئية الخاصة بنموذج كتلة (**الواصفات الطوبولوجية**)



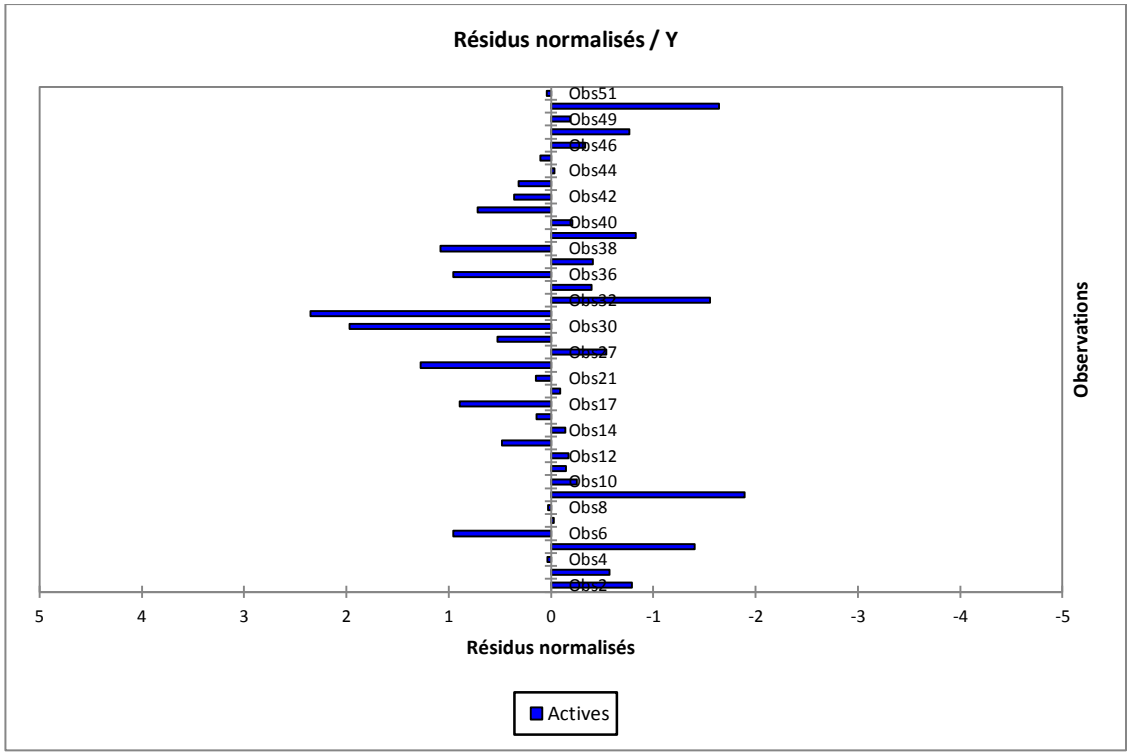
الشكل (***III***-3) : بيان يوضح تغيرات (Résidus normalises) بدلالة (Y) الخاصة بكتلة (**الواصفات الطوبولوجية**)

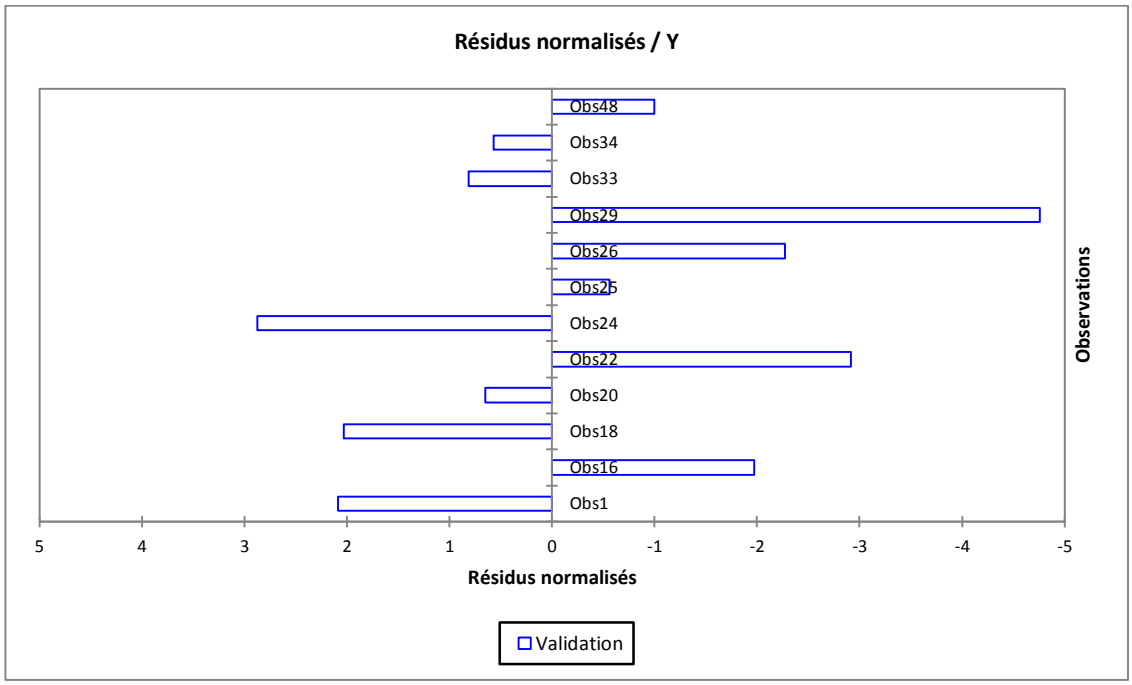


الشكل (***III***-4):بيان يوضح تغيرات (Résidus normalises) بدلالة ((Y) Préd) الخاصة بكتلة (**الواصفات الطوبولوجية**)



الشكل (***III–5***) :بيان يوضح تغيرات (Y) بدلالة Préde(Y) الخاصة بكتلة (**الواصفات الطوبولوجية**)





الشكل (***III–6***): بيان يوضح تغيرات (Y) بدلالة (Résidus normalises) الخاصة بكتلة (**الواصفات الطوبولوجية**)

III-3-1-2- تحليل و مناقشة نتائج التقنية الاولى :

اظهر تحليل QSPR ان الارتباط بين PEC505 مع الواصفات الجزيئية لمجموعة من المركبات الفينولية والثيوفينولية ، يعطي نموذجا مع معطيات احصائية جيدة، للتنبؤ بسمية هذه المجموعة ، في حين تم الحصول على افضل نموذج باستخدام كتلة (الواصفات الطوبولوجية) من الخمس كتل المدروسة (الواصفات الاساسية، مسار و طرق العد، مؤشرات الربط، مؤشرات المعلومات)، و اتضح هذا من خلال معامل التحديد (R^2) و ال (MSE) و الذي يعطينا فكرة عن النسبة المئوية لتغير المتغير المنمذج و النسبة المئوية للمخلفات (Residu).حيث كلما اقترب معامل التحديد إلى 1 ،كلما كان النموذج أفضل.

في هذه الحالة لدينا R^2=0.907 و R^2 ajusté =0.883

وبالتالي فإن النسبة المئوية لتغير المتغير المنمذج = % 90.7

وكذلك المخلفات (Résidu) =

$$Residue = 1- R^2 = 1-0.907= 0.098= 100 \times 0.098=9.3\%$$

و يوضح بيان ((Y) / (Y)Préde) الشكل (***III-5***) الارتباط الخطي بين السمية (Y) المرصودة و قيم السمية المتوقعة (Y)Préde ، ان 90.7% من جزيئات قاعدة البيانات داخل مجال الثقة .

و نلاحظ في بيان (Résidus normallises /(Y)) و (Résidus normalises/ (Y) Préd) الشكل(***III***-3) و (***III***-4) الذي يمثل تباين المخلفات المعيارية كدالة للنشاط التجريبي ان 90.7%توجد في المجال (2-,2+) و ان كل قيمة خارج هذا المجال غير موثوق فيها.

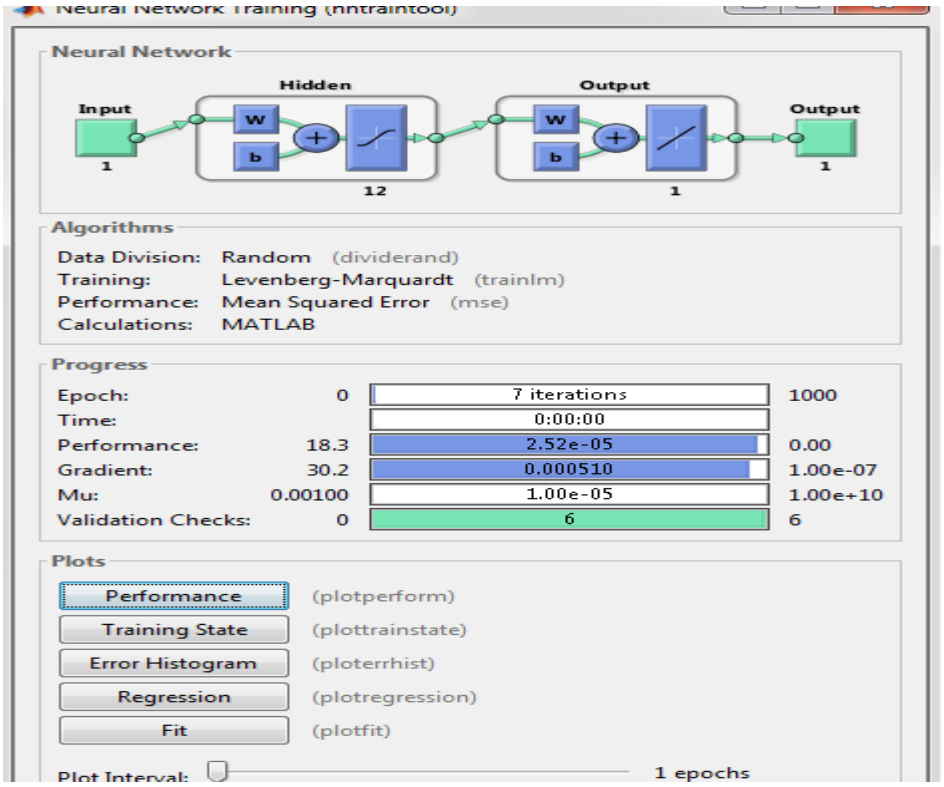
فمن خلال مقارنة نتائج كتلة (الواصفات الطوبولوجية) و نتائج الكتل الاخرى من حيث معامل التحديد (R^2) بالخصوص والتحقق من جميع معطيات هذه المجموعة، استنتجنا ان افضل نموذج تنبئي بتحليل (QSPR) العلاقة الكمية بين البنية الجزيئية و الخاصية الفيزيوكيميائية بطريقة الانحدار الخطي المتعدد (MLR) ل 51 مركب المختارة من عائلة الفينولات و الثيوفينولات، هو نموذج كتلة الواصفات الطوبولوجية (Topological descriptors) .

III-3-2-نتائج الطريقة الثانية :

اجرينا طريقة الشبكات العصبية الاصطناعية (ANN) لإنشاء نموذج QSPR ل51 مركب من عائلة الفينولات والثيوفينولات بالخمس كتل المذكورة اعلاه، و تبعا لما تحصلنا في التجربة الاولى بطريقة MLR من افضل كتلة اعطت افضل نموذج، سنجري طريقة ANN على نفس الكتلة .

تم اجراء التجربة على nnstart المقدم من MATlab مع ادخال مجموعة بيانات وهي قيم الواصفات الجزيئية للمركبات المختارة و PEC50 (Y) ، تكون مجموعة البيانات المؤخوذة 70% من اجمالي مجموعة البيانات و 15% لتحقق و 15%للاختبار من اجمالي مجموع البيانات بوضع عدد التدريب 12، الى ان نتوصل على النتائج التالية :

III-3-2-1-كتلة الواصفات الطوبولوجية (Topological descriptors)

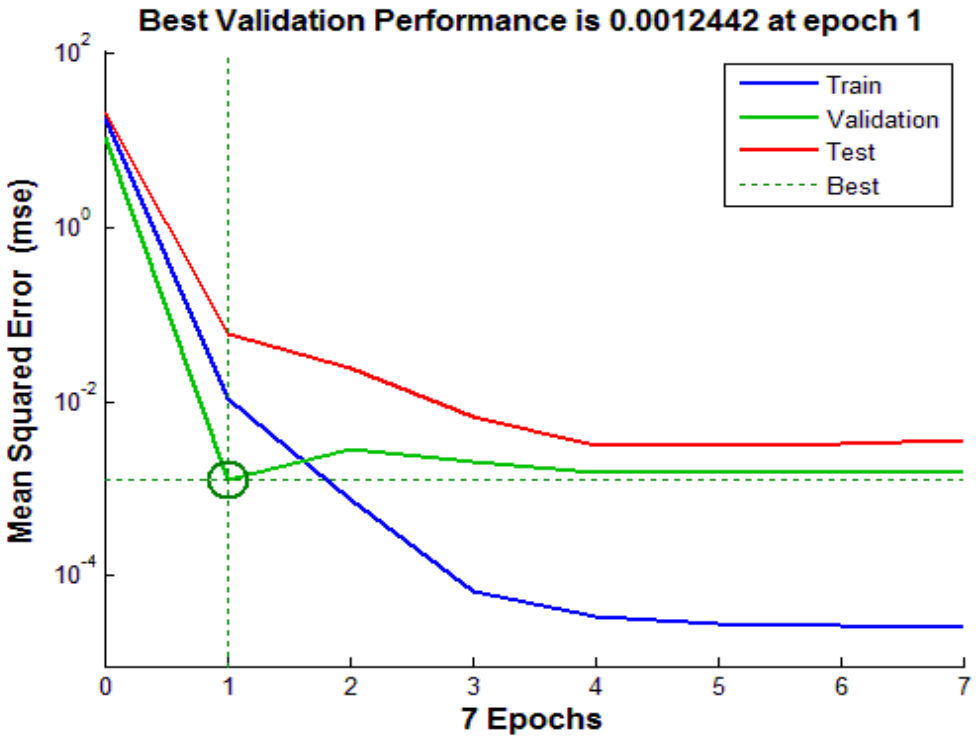


الشكل (*III-7*): لقطة شاشة لتدريب نموذج الشبكة العصبية بواسطة Matlab لكتلة (الواصفات الطوبولوجية)

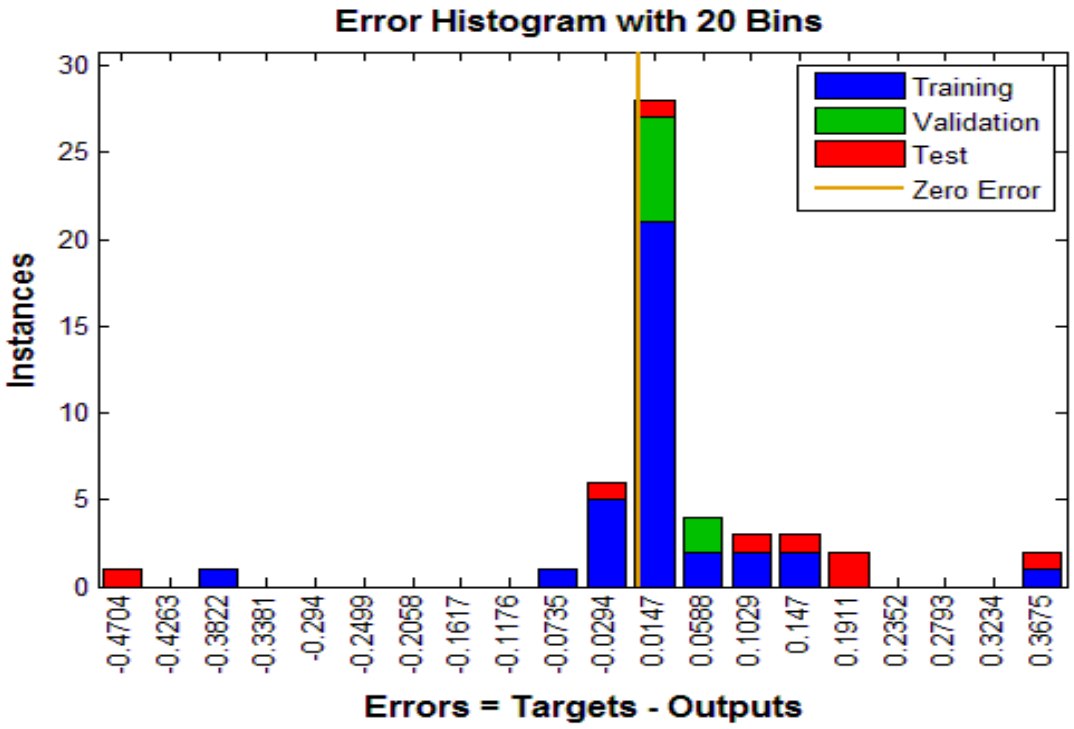
الجدول(III-6): نتائج الانحدار لكتلة(الواصفات الطوبولوجية)

Type	R
Training	0.99626
Validation	0.99878
Teste	0.988081
All	0.99386

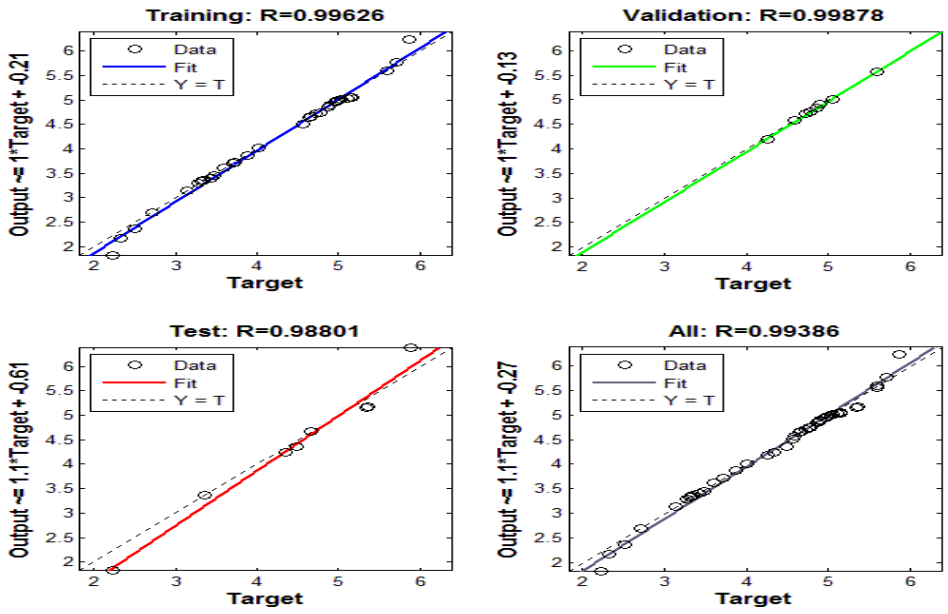
من الجدول السابق نلاحظ ان كتلة الواصفات الطوبولوجية (***Topological descriptors***) لها قيم ال R اكبر من 0.95 ، هذا يعني ان استجابة الشبكة العصبية الاصطناعية بالكتلة الطوبولوجية مرضية.



الشكل(***III***-8) :يوضح مخطط الاداء لكتلة (الواصفات الطوبولوجية)



الشكل(*III*-9): بيان يوضح الخطأ (Errors)



الشكل(*III*-10):بيانات توضح الانحدار R لكل من(Trining)و(Validation)و(test)

III-3-2-2-تحليل و مناقشة نتائج الطريقة الثانية :

يوضح الشكل (***III-7***) تفاصيل الخوارزمية المطبقة ، نوع الخطأ المحدد لتدريب الشبكة ، و يمكننا ايضا رؤية مخطط الاداء و مخطط الانحدار منه. و يعطينا ايضا قيمة الخطأ الذي تم الحصول عليه اثناء تدريب الشبكة ، و هذا موضح في كل من الشكل(***III-8***) و الشكل (***III***-9) و الشكل (***III-10***).

بعد اجراء تحليل QSPR باستخدام مجموعة الكتل المختارة الخمس(الواصفات الاساسية ،الواصفات الطوبولوجية، مسار و طرق العد، مؤشرات الربط، مؤشرات المعلومات) استنتجنا ان كتلة الواصفات الطوبولوجية هي اكثر كتلة ملائمة اي ذات افضل نموذج ، من خلال قيم الانحدار R حيث ان قيم الانحدار اكبر من 0.95 ،لذا نستطيع القول ان استجابة الشبكة العصبية الاصطناعية مرضية.

في هذه الحالة لدينا :

$$R_{trining}=0.99626 \text{ و } R_{validation}= 0.99878 \text{ و } R_{test}=0.988081$$

و منه نستنتج ان نموذج كتلة الواصفات الطوبولوجية تعطي نتيجة مرضية للتنبؤ بالخاصية الفيزيوكيميائية (السمية)، ل51 مركب المختار بطريقة الشبكة العصبية الاصطناعية باستخدام كتلة الواصفات الطوبولوجية .

خاتمة عامة

من خلال الدراسة التنبؤية التي اجريناها تم نمذجة سمية 51 مركب من عائلة (الفينولات والثيوفينولات) باستخدام تحليل QSPR بطريقتين: الطريقة الاولى طريقة الانحدار الخطي المتعدد(MLR) والطريقة الثانية طريقة الشبكات العصبية الاصطناعية (ANN).

بالنسبة الى طريقة MLR :

بعد اجراء الحسابات على 5 كتل(1)،(2)،(3)،(4)،(5) توصلنا الى تحقيق العلاقة التالية:

(الواصفات الجزيئية) f = الخاصية (السمية) بالشكل التالي $y = a + \Sigma (bj \times xj)$ حيث كان افضل نموذج تم الحصول عليه هو الخاص بالكتلة الطوبولوجية (2) (Topologicaldescriptor) والذي تميز ب ($R_{ajuste}=0.883$, $R^2=0.907$)

أما بالنسبة لنتائج الكتلة (1) فتميزت ب ($R_{ajuste}=0.725$, $R^2=0.761$)

أما بالنسبة لنتائج الكتلة (3) فتميزت ب ($R_{ajuste}=0.061$, $R^2=0.086$)

أما بالنسبة لنتائج الكتلة (4) فتميزت ب ($R_{ajuste}=0.678$, $R^2=0.704$)

أما بالنسبة لنتائج الكتلة (5) فتميزت ب ($R_{ajuste}=0.008$, $R^2=0.034$)

وبناء عن هذه النتائج يمكن استنتاج ما يلي:

- ✓ تعتبر الدراسة بتقنية ال QSPR باستعمال طريقة (MLR) حلا موثوقا لأنها تسمح للكيميائيين بإيجاد قيم ل Y لمركبات مجهولة خاصة المركبات الكيميائية الخطيرة ذات السمية العالية التي يصعب التعامل معها مخبريا.
- ✓ أنه يمكن دراسة العلاقة الكمية التي تربط بين البنية الجزيئية - الخاصية الفيزيوكيميائية (السمية) وذلك باستخدام طريقة الانحدار الخطي المتعدد (MLR).

أن القدرة التنبؤية تعتمد على نوع وعدد المركبات الكيميائية المكونة لقاعدة البيانات.

ولتحسين النتائج يمكن اقتراح ما يلي:

- ✓ إلغاء أو حذف المركبات غير مرتبطة.
- ✓ جمع أكثر عدد من الجزيئات لتوسيع قاعدة البيانات لإعطاء نتائج أفضل.
- ✓ يجب إجراء تحسين (Optimisation) كامل للمركبات للوصول إلى أقصى درجة استقرار.

بالنسبة الى طريقة ANN :

بعد اجراء التطبيق الحسابي على افضل كتلة تحصلنا عليها في التجربة الاولى بطريقة MLR

توصلنا الى ان هذه الكتلة (Topologicaldescriptor) هي كتلة ملائمة اي ذات نموذج ج من خلال قيم الانحدار R

$R_{trining}$=0.99626 و $R_{validation}$= 0.99878 و R_{test}=0.988081

بمعنى انها تعطي نتيجة مرضية للتنبؤ بالخاصية الفيزيوكيميائية (السمية).

وبعد ما تحصلنا على النتائج السابقة نقارن بين الطريقتين نجد ان R=0.99 لنموذج الشبكات العصبية الاصطناعية (ANN) اكبر من R^2 =0.907 لنموذج الانحدار الخطي المتعدد (MLR) ومنه نستنتج ان نموذج ANN افضل اداة فعالة للتنبؤ بالسمية من نموذج MLR .

[1]: e3arabi/العلوم /ما-هي-الفينولات؟/

[2]: m.ar.dakentechjp /news/thiophenol-24415506.html

[3]:chemistrysources /ثيوفينول-،-الفينول-الكبريتي/ thiophenol/

[4] :Amrani Hayat,Lebssisse Chaima,master,Développement de modèles QSPR pour la prédiction d'enthalpie de décomposition des composés organiques à l'aide des descripteurs moléculaires, page4, 2018.

[5]: تسنيم معابرة /mawdoo3/ الخواص_الكيميائية_للمادة /2018 7سبتمبر

[6]: kinderfac.mans.edu.eg/files/03_poisons.pdf

[7]:futura-sciences/sante/definitions/medecine-toxicite-6517

[8]:Bouazidi/unive-setif.dz/ 3eme biochemie/ cours/Toxicologie/2019-2020

[9]:Jerzy Leszczynski, Tomasz Puzyn, Mark T.D. Cronin /Department of Chemistry, Jackson State University, U.S.A./ Recent Advances in QSAR Studies/ 2010.

[10]: Bqua/ما هي بنية جزيئية – تعريف أشكال الجزيئات/ يونيو 8، 2021

[11]: **التوزيع**: معامل LogKow (باستعمال) بن عمارة سناء/مذكرة ماستر/ نمذجة العلاقة (البنية الكيميائية - QSPR 2015 طريقة للمركبات العضوية/

[12]:unilim.fr/pages_perso/jean.debord/model/model.pdf

[13] : صبرينة كباس، راضية هبال/دراسة بواسطة النمذجة الجزيئية للبنية والخصائص الفيزيوكيميائية لبعض الجزيئات النيكليوزيدية المضادة للفيروسات/مذكرة ماستر/2018.

[14] : قعري عبد الحي، شراحي عبد الرزاق ،سلطاني جمال/دراسة تأثير المذيبات على الخصائص المضادة مذكرة ماستر/جامعة قاصدي مرباح ورقلة كلية الرياضيات /لألكسدة لحمض االسكوربيك: دراسة مقارنة وعلوم المادة قسم الكيمياء/2019

[15].قيس مجيد عبد الحسين علوش/مفهوم واهمية النماذج/كلية الادارة و الاقتصاد/ جامعة بابل/2013

[16]: schkola4kotovo.ru/ar/chto-takoe-statisticheskaya-model-statisticheskoe-modelirovanie

[17]:mohamed3okasha.blogspot.com/2016/03/optimization-techniques

[18]:GeomOpt_Volodya_2.pdf/Introduction to Geometry Optimization/Computational Chemistry lab/ 2009

[19] تـخة مباركة/مساهمة في دراسة العلاقة (بنية كيميائية- رائحة) باستعمال طريقة آلات الشعاع تطبيق على عائلة البيرازين/جـامعة قـاصدي مربـاح ورقـلة كلــية العلوم والعلوم الهندسيـة قــسم هندسة الطـــرائق/ماجيستير/2008 (SVM)الساند

[20]:Jahanbakhsh Ghasemi , Saadi Saaidpour ,Steven D. Brown/ Journal of Molecular Structure: THEOCHEM 805 (2007) 27–32

[21]:Adam Hayes /Multiple Linear Regression (MLR)/ Mar 30, 2021

[22]: جمال احمد الشوادفى،د-الوهاب السيد حجاج/ المجلة العلمية بقطاع كلية التجارة –جامعة الازهر: 2013د-

[23]:JAKE FRANKENFIELD/Artificial Neural Network (ANN)/ Aug 28, 2020

[24]:Jure Zupan/National institute of chemistry/Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them/January 1994

[25]:file:///C:/Users/scc-2014/Downloads/Documents/These-Bouarra-Nabil.pdf

[26]:MAZOUZI Aziza, NOUI Chahira/Méthode de prédiction des propriétés physiques et thermodynamiques des composes organiques purs basées sur le concept des interactions binaires des groupes organiques fonctionnels/UNIVERSITE KASDI MARBAH OUARGLA Faculté des Sciences et de la Technologie et Science de la matière Département de Génie des Procédés/Masteur/2013

[27] : محمد تاج السر، سهام الصادق،فوزية سليمان، حكمة بشارة، زينب منصور، صابرين يوسف، روضة صابر، زهراء أحمد،ريهام محمد ، نزار عبد اهلل،سعيد محمد/مارس 2013

[28]:Ahmed Gad/دليلك العربي لتعلم الشبكات العصبية الإصطناعية/23,11,2017

[29]:staff.emu.edu.tr /INTRODUCTION TO CHEMDRAW ULTRA 12.0 /Documents/courses/ITEC107/LECTURE7.pdf

الملخص:

يستهدف عملنا التنبؤ بسمية مجموعة من المركبات الكيميائية من عائلة الفينولات و الثيوفينولات ، حيث تعتمد هذه الدراسة اساسا على النمذجة الجزيئية.

و يرتكز هذا الهدف على امكانية ايجاد العلاقة بين البنية الجزيئية والخاصية الفيزيوكيميائية (بنية/سمية) ، و تمت هذه الدراسة على 51 مركب كيميائي بتطبيق تقنية QSPR وذلك باستعمال مصفوفات تعتمد على الواصفات الجزيئية و EPC50، شملت هذه الدراسة 5 كتل من الواصفات الجزيئية (2D) وهذا وفق تقنيتين، الانحدار الخطي المتعدد (MLR)وحسابات الشبكة العصبية الاصطناعية (ANN) وهي من طرق النمذجة الاكثر شيوعا في QSPR وهذا باستعمال البرمجية (Xlsta) و (MATlab) على التوالي. وكـانت أحـسن نـتائج في تقنية (MLR) هي نتائج كـتلة (الواصفات الطوبولوجية) R^2=0.907 و MSE=0.115 و كانت نتائج ال(ANN) لنفس الكتلة جيدة حيث ان R لكل من Training و Validation و Teste اكبر من 0.95 .

الكلمات المفتاحية : النمذجة الجزيئية ، EPC50، الواصفات الجزيئية، QSPR، الانحدار(MLR)،الانحدار (ANN).

Résumé:

Notre travail a pour but de la préduction de la toxicité d'un ensemble de composantes chimiques des familles phénols et éthiophénols. Cette étude se base sur la modélisatio partielle.

L'objectif est la possibilité de trouver la relation entre la structure partielle et la spécificité physicochimique (structure\toxicité) . Notre étude est faite sur 51composantes chimiques par l'application technique QSPR, et cela avec l'utilisation des matrices se basant sur les descripteurs partiels et EPC50. Cette étude comporte 5 masses de descripteurs partiels (2D), selon deux techniques : la régression linéaire multiple (MLR) et les calculs de réseaux de neurones artificiels (ANN) qui sont parmi les méthodes de modélisation les plus connues, dans (QSPR) et cela par l'utilisation successive des logiciels (XLSTAT) et (MATLab) . Les bons résultats de la technique(MLR) étaient ceux de l'ensemble des descripteurs topologiques R^2 = 0.907 et MSE=O.115. Les résultats de l'ANN du même ensemble étaient très bons parce que R pour training, validation et teste était plus que 0.95.

Mots clé :

Modélisation partielle, EPC50, descripteurs partiels, QSRR, régression (MLR), régression (ANN).