

الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET  
DE LA RECHERCHE SCIENTIFIQUE

جامعة صالح بونبندر قسنطينة 3

UNIVERSITE SALAH BOUBNIDER CONSTANTINE 3

كلية الطب - قسم الصيدلة

FACULTE DE MEDECINE - DEPARTEMENT DE  
PHARMACIE



Mémoire de fin d'études en vue de l'obtention du diplôme de  
DOCTEUR EN PHARMACIE

**DEVLOPPMENT D'UN OUTIL  
D'INTERPRETATION DES SPECTRES  
INFRAROUGE**

Encadré par :

Pr. LALAOUNA Abd El Djalil

Présenté par :

- AIT HABBOUCHE Aicha

- BRAHMI Soulef

- MESLI Nessrine

- RAHMANI Ilhèm

Membres de jury :

Dr Titel F; pharmacien maitre-assistant en Chimie Analytique

Dr Djaou M; pharmacienne maitre-assistante en Chimie Analytique

Dr Gueroui M; pharmacien maitre- assistant en Chimie Thérapeutique

ANNEE UNIVERSITAIRE : 2023 - 2024

# Table des matières

Liste des figures .....	I
Liste des tableaux .....	IV
Liste des annexes .....	V
Liste des abréviations et des acronymes.....	VI
Liste des définitions.....	VIII

## ETUDE BIBLIOGRAPHIQUE

### CHAPITRE I. SPECTROSCOPIE INFRAROUGE

1. Domaine spectrale .....	2
2. Principe .....	3
2.1. Fondement théorique de la spectroscopie infrarouge .....	3
2.2. Vibrations et rotations moléculaires en IR .....	6
2.2.1. vibrations moléculaires .....	6
a. Molécules diatomiques .....	6
b. Molécules polyatomiques .....	7
c. Modes de vibrations dans un groupe d'atomes .....	8
2.2.2. Rotations moléculaires.....	10
2.2.3. Couplage rotation-vibration .....	10
3. Interprétation des spectres.....	11
3.1.Facteurs déterminants le spectre infrarouge .....	13
4. Application .....	13
4.1.Analyse quantitative .....	13
4.2.Analyse qualitative .....	14
4.2.1. Identification de matières premières.....	14
4.2.2. Polymorphisme .....	14
4.3.Analyse fonctionnelle .....	14
4.3.1. Principales régions du spectre.....	14
a. Région des vibrations de valence des principaux groupes fonctionnels. 14	
b. Empreinte.....	16
c. Zone « dite des aromatiques » .....	17

5. Avantages et limite .....	17
5.1. Avantage .....	17
5.2. Limites .....	17
5.3. Comparaison entre la spectroscopie IR et les autres méthodes spectrales.....	18

## **CHAPITRE II. CHIMIOMETRIE**

1. Historique .....	19
2. Définition .....	19
3. Classification des méthodes chimiométrique .....	20
3.1. Analyse qualitative .....	20
3.1.1. Analyse non supervisée .....	20
3.1.2. Analyse supervisée .....	20
3.1.3. Méthodes exploratoires des données .....	20
3.2. Analyse quantitatif.....	21
3.2.1. Méthodes de régression .....	21
4. Etapes de l'analyse chimiométrique.....	23
5. Présentation des données.....	23
5.1. Principales approches pour la gestion de la qualité des données en pratique.....	24
5.2. Qualité de l'échantillon .....	25
5.3. Évaluation préliminaire des données .....	25
5.4. Prétraitement des données .....	26
5.4.1. Méthodes de prétraitement par variable (longueur d'onde) .....	27
5.4.2. Méthodes de prétraitements spectroscopiques .....	27
5.5. Sélection des variables .....	28
6. Évaluation des méthodes chimiométriques .....	29
6.1. Définition .....	29
6.2. Taille et partition des ensembles des données .....	29
6.3. Méthodes d'évaluation de modèle chimiométriques .....	31
6.4. Critères de performance.....	32
7. Application de la chimiométrie .....	34

## **CHAPITRE III. RESEAU DES NEURONES ARTIFICIELLES**

1. Historique .....	36
---------------------	----

2. Définition .....	36
2.1. Du neurone biologique au neurone formel .....	37
3. Structure de neurone .....	39
4. Fonction d'activation .....	40
5. Architecture de réseaux de neurones .....	42
5.1. Réseaux de neurones non bouclés «Feed Forward » .....	42
5.1.1. Réseaux de neurones monocouche .....	43
5.1.2. Réseaux de neurones multicouches .....	43
5.1.3. Réseaux de neurones à connectivité locale .....	44
5.2. Réseaux de neurones bouclés « Feed-Back »... .....	44
6. Types de Réseaux de neurones .....	45
6.1. Perceptron .....	45
6.1.1. Perceptron simple .....	45
6.1.2. Perceptron multicouches .....	46
6.2. Modèle ADALINE .....	46
6.3. Modèle de Hopfield .....	47
6.4. Modèle de Kohonen .....	48
6.5. Réseaux de neurones à Fonctions de base Radiales (RFR) .....	49
6.6. Réseaux de neurones compétitifs .....	49
6.7. Réseaux de neurones évolutifs .....	50
7. Choix de réseau de neurones artificiels .....	50
8. Apprentissage .....	51
8.1. Apprentissage automatique .....	51
8.1.1. Définition .....	51
8.1.2. Phases de l'apprentissage automatique .....	52
8.2. Types d'apprentissage automatique .....	52
8.2.1. Apprentissage supervisé .....	52
8.2.2. Apprentissage non-supervisé .....	52
8.2.3. Apprentissage par renforcement .....	53
8.3. Règles d'apprentissage .....	53
8.4. Choix et influence des hyperparamètres .....	54
8.5. Développement de réseaux de neurones artificiels .....	55
8.6. Critères de performance d'un modèle réseau de neurones artificiel .....	57
9. Applications .....	57
10. Limites et avantages .....	58

## PARTIE II. PARTIE PRATIQUE

1. Equipements et logiciels.....	60
1.1. Equipements .....	60
1.2. Logiciels .....	60
2. Description de la base de données .....	61
2.1. Méthode.....	62
2.2. Résultats .....	62
2.3. Discussion.....	69
3. Extraction et transformation des spectres en données Excel.....	71
3.1. Méthode.....	71
3.2. Résultats .....	72
3.3. Discussion .....	77
3.4. Conclusion.....	78
4. Extraction automatique des codes SMILES en utilisant les noms des substances ..	79
4.1. Méthode .....	79
4.2. Résultats .....	79
4.3. Discussion.....	80
4.4. Conclusion.....	81
5. Identification des fonctions chimiques.....	82
5.1. Méthode.....	82
5.2. Résultats .....	96
5.3. Discussion.....	106
5.4. Conclusion.....	106
6. Développement des modèles ANN .....	107
6.1. Méthode .....	107
6.2. Résultats .....	109
6.2.1. Modèle ANN pour l'identification de la fonction « Carboxylic acid or conjugate base ».....	109
6.2.2. Modèle ANN pour l'identification de la fonction « Aldehyde ».....	115
6.2.3. Synthèse des résultats pour l'ensemble des fonctions étudiées.....	121
6.3. Discussion.....	129
6.3.1. Fonction « Carboxylic acid or conjugate base ».....	129
A. Nombre des couches cachées .....	129
B. Fonctions d'activation .....	130
C. Algorithmes d'optimisation .....	130

D.	Série de calibration .....	131
E.	Série de validation .....	132
F.	Comparaison entre les séries de calibration et de validation .....	132
G.	Matrice de confusion .....	133
H.	Courbe ROC .....	135
I.	Importance de variables .....	136
6.3.2.	Fonction « Aldhyde » .....	136
A.	Nombre des couches cachées.....	136
B.	Fonctions d'activation .....	137
C.	Algorithmes d'optimisation .....	137
D.	Matrice de confusion .....	138
E.	Courbe ROC .....	139
F.	Importance de variables .....	139
6.3.3.	Autres fonctions .....	139
A.	Série de calibration .....	139
B.	série de validation .....	140
C.	Couches cachées .....	141
D.	Fonction d'activation .....	141
E.	Algorithmes d'optimisation .....	141
6.4.	Conclusion.....	142
7.	Développement du logiciel InterSpec.Ai .....	144
7.1.	Description du logiciel .....	144
7.2.	Validation du logiciel .....	145
7.2.1.	Méthode .....	145
7.2.2.	Résultat .....	145
A.	Molécule «Alachlor » .....	146
B.	Molécule « Benazolin » .....	147
C.	Molécule Nitrothal-isopropyl .....	148
D.	Molécule « pyridate » .....	149
7.3.	Discussion.....	150
7.3.1.	Molécule « Alachlor » .....	150
7.3.2.	Molécule « Benazolin » .....	152
7.3.3.	Molécule Nitrothal-isopropyl .....	155
7.3.4.	Molécule « Pyridate » .....	157
7.4.	Conclusion.....	158

# Résumé

Ce projet vise à développer un outil d'interprétation automatique des spectres infrarouges pour diverses molécules organiques, traditionnellement analysées via des méthodes qui requièrent une expertise significative. Une base de données comprenant les spectres infrarouges de 1540 molécules différentes a été utilisée. Ces données sont traitées via un script Python qui extrait les informations pertinentes dans des fichiers Excel. Utilisant les notations SMILES et SMART, le script génère un ensemble de données constitué de variables X, représentant les longueurs d'onde et leur transmittances correspondants, et de variables Y, indiquant l'absence ou la présence de fonctions chimiques spécifiques pour chaque molécule.

Un second script Python, exécuté dans Jupyter, a servi à développer un réseau de neurones artificiels. Ce réseau traite les données préalablement collectées afin de créer un modèle capable d'apprendre à partir de ces données et de généraliser à de nouvelles données non observées auparavant. Les résultats obtenus incluent des tableaux de performances basés sur cinq critères, le "rappel" étant jugé le plus crucial, ainsi que divers graphiques, tels que des courbes ROC et des matrices de confusion.

L'analyse des résultats a permis d'évaluer les performances de 111 fonctions au total. Parmi celles-ci, 49 fonctions ont démontré des performances très satisfaisantes. Plus précisément, 83,6% de ces fonctions ont obtenu une précision supérieure à 0,8 ce qui est considéré comme excellent. Notamment, 31,7% d'entre elles ont même dépassé 0,9. De plus, 16,32% des fonctions ont atteint une précision se situant entre 0,7 et 0,8, ce qui est jugé comme bon. En basant sur ces derniers, l'interface du logiciel, ainsi que sa structure, ont été élaborées avec Visual Studio 2019.

Le taux des fonctions bien apprises par réseaux de neurones artificiels est considéré comme très encourageant, les échecs restants étant principalement attribués au faible nombre de modalités disponibles pour certaines fonctions. Aussi le logiciel InterSpec.Ai a été testé avec succès sur des molécules non incluses dans le jeu de données initial.

En conclusion, ce système, confirmant l'efficacité de l'outil développé pour faciliter l'interprétation des spectres infrarouges, répondant ainsi pleinement à l'objectif initial du projet.

**Mots clés :** interprétation des spectres infrarouges, intelligence artificielle, logiciel InterSpec.Ai, réseaux de neurones artificiels, Python, visual studio 2019.

# Abstract

This project aims to develop an automated tool for interpreting infrared spectra for various organic molecules, traditionally analyzed using methods that require significant expertise. A database containing the infrared spectra of 1540 different molecules was used. This data is processed via a Python script that extracts the relevant information into Excel files. Using SMILES and SMART notations, the script generates a dataset consisting of X variables, representing the wavelengths and their corresponding transmittances, and Y variables, indicating the absence or presence of specific chemical functions for each molecule.

A second Python script, executed in Jupyter, was used to develop an artificial neural network. This network processes the previously collected data to create a model capable of learning from this data and generalizing to new, previously unobserved data. The results obtained include performance tables based on five criteria, with "recall" being deemed the most crucial, as well as various graphs, such as ROC curves and confusion matrices.

The analysis of the results made it possible to evaluate the performances of 111 functions in total. Among these, 49 functions demonstrated very satisfactory performances. More precisely, 83.6% of these functions obtained a precision greater than 0.8, which is considered excellent. Notably, 31.7% of them even exceeded 0.9. Additionally, 16.32% of the functions achieved a precision between 0.7 and 0.8, which is considered good. Based on these results, the software interface and structure were developed using Visual Studio 2019.

The rate of functions well-learned by the ANN is considered very encouraging, with the remaining failures mainly attributed to the low number of modalities available for certain functions. The InterSpec.Ai software was also successfully tested on molecules not included in the initial dataset.

In conclusion, this system, confirming the effectiveness of the developed tool for facilitating the interpretation of infrared spectra, fully meets the initial objective of the project.

**Key words:** Interpretation of infrared spectra, artificial intelligence, InterSpec.Ai, artificial neural networks, Python, Visual Studio 2019.