

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOHAMED BOUDIAF - M'SILA

FACULTE DES SCIENCES

DEPARTEMENT DE MICROBIOLOGIE & BIOCHIMIE

N°:



DOMAINE : SCINCES DE LA NATURE ET DE LA VIE

FILIERE : SCIENCE BIOLOGIQUE

OPTION : BIOCHIMIE APPLIQUEE

Mémoire présenté pour l'obtention

Du diplôme de Master Académique

Par : Benzahia Anfal ELdjabria

Intitulé

**Prédiction de la Résistance Bactérienne aux
Antibiotiques Basée sur des approches ML et DL**

Soutenu devant le jury composé de :

Dr.Harrar Abdenassar	Université Mohamed Boudiaf M'sila	Président
Dr.Madi Nadia	Université Mohamed Boudiaf M'sila	Rapporteur
Dr.Gadri Said	Université Mohamed Boudiaf M'sila	Rapporteur
Dr.Belbahi Amine	Université Mohamed Boudiaf M'sila	Examineur

Année universitaire : 2022/2023

Dédicace

*Je dédie ce travail à **mes parents** qui ont travaillé si dur pour que j'atteigne ce succès, **mon père, ma mère**, vous avez été le meilleur soutien pour moi en tout, tu étais un symbole de sacrifice, cette réalisation est pour toi autant qu'elle est pour moi, à **mon frère Dr. Ayoub** qui a joué un rôle important dans l'élaboration de mon chemin, à **mon frère Dr. Sohaïb** qui m'a fourni toutes les possibilités et tous les moyens d'atteindre ce que je voulais, à **mon frère Mossaab**, qui regardait mon travail à chaque instant. Pour **mon cousin Hood**, il était le programmeur qui m'a aidé à terminer le travail. A **mes sœurs (Noussaiba, Basmala, Jinan)** et **la femme de mon frère (Amina)** qui m'ont appuyé tout au long de mes études. A **mes amies (Oumaima, Hana, Akila, Hadjer, Chaima)** qui ont été mon support indéfectible tout au long de ce parcours.*

*Au **club ELEC**, en particulier, l'étudiant **Riyad Barka**, qui m'a encouragé à montrer et à clarifier l'idée. Au brillant **programmeur Tarek**, qui a consacré son temps à m'enseigner. Tu étais comme le soldat caché pour ce travail.*

*Aux étudiants du Département **d'informatique (Zakaria, Abdenour, Okba, Kadour)** qui se sont efforcés pour moi d'atteindre ces résultats*

*À toute **ma famille** qui a attendu les moments de ma remise des diplômes avec tous ses détails, merci pour votre soutien et votre confiance en moi pour atteindre ce que je souhaite et ce que vous souhaitez.*

Enfin, à l'avenir qui m'attendra, plein de possibilités infinies et de nouvelles aventures. Je dédie cette réalisation aux opportunités à venir, sachant que cette étape n'est que le début d'un voyage extraordinaire. Je suis prêt à relever les défis et à avoir un impact positif sur le monde.

Remerciement

Si je n'avais pas eu la grâce de Dieu, je n'aurais pas atteint ce succès, alors loué soit Dieu.

*Je tiens à remercier **mon superviseur Dr Madi Nadia**, qui m'a soutenu dans mon idée et a suivi mon travail avec diligence, et qui s'est efforcé de l'atteindre.*

*Tous mes remerciements et **mon appréciation au superviseur, Dr Gadri Said**, qui a cru en mon idée et a travaillé dur pour m'apprendre à atteindre ces résultats. Tous mes remerciements et ma gratitude vont au **chercheur américain, Dr Belgacem Mihi**, qui a eu le grand mérite de clarifier l'idée et de m'aider à la mettre en œuvre.*

*Je remercie mon professeur respecté, le **Dr Harrar Abdelnasser**, qui a été considéré comme le troisième superviseur de ce travail et qui a répondu à toutes mes questions et m'a aidé.*

*Je remercie le directeur de l'incubateur d'entreprises, le professeur **Zawish Reda**, qui nous a offert toutes les possibilités de réussite, et je remercie **tout le personnel de l'incubateur** qui a contribué à notre soutien, en soutenant nos projets et en dispensant les cours nécessaires.*

*Tous mes remerciements au personnel **administratif de la faculté des Sciences**, aux **laboratoires du la faculté et de l'incubateur** qui ont mis à notre disposition tous les moyens et les outils nécessaires pour mener à bien notre travail, en particulier au **Dr Safia**, qui a assuré le suivi des travaux de laboratoire avec **le superviseur**.*

***Je remercie le Dr Arich**, qui a corrigé mon idée avant de l'achever.*

*Tous mes remerciements au **Dr Gaman Moustafa**, économiste, qui a poursuivi le travail d'un point de vue économique.*

Et celui qui ne remercie pas les gens ne remercie pas Dieu, c'est grâce à vous que j'ai pu atteindre cet objectif, alors tous les remerciements vous reviennent.

Sommaire

Introduction	1
Chapitre I. Antibiotique et Antibio-résistance	3
1. Définition de la bio-informatique	3
2. Les bases de données en biologie.....	3
2.1. Définition d'une base de données biologique	3
2.2. Type des bases de données biologiques	3
2.3. Prédiction sur les structures et fonction des protéines	4
3. Présentation des bactéries étudiées	5
3.1. Neisseria gonorrhoeae	5
3.2. Description et classification de N.gonorrhoeae	6
3.3. Description et classification de Escherichia coli.....	8
4. Les antibiotiques	11
4.1. Définition d'un antibiotique.....	11
4.2. Origines des antibiotiques	12
4.3. Effets des antibiotiques	12
4.4. Principales familles d'antibiotiques	13
4.5. Aminosides.....	14
5. Modes d'action des antibiotiques.....	14
5.1. Antibiotiques agissant sur la synthèse du peptidoglycane	14
5.2. Antibiotiques inhibant la synthèse protéique	15
5.3. Antibiotiques agissant sur les acides nucléiques.....	16
6. Antibio-résistances	16
6.1. Définition	16
6.2. Types de résistance aux antibiotiques	16
6.3. Voies de transmission des résistances aux antibiotiques	17
6.4. Biochimie de la résistance.....	18

6.5. Voies biochimiques de la résistance aux antibiotiques	18
6.6. Mécanismes de résistances aux d'antibiotiques	19
7. Apport de l'IA sur la prédiction des bactéries résistante	20
Chapitre II. Apprentissage automatique et Apprentissage profond	21
1. Préface.....	21
2. Définition	21
3. L'apprentissage automatique (Machine Learning).....	21
.3.1 Types d'apprentissage automatique	22
3.2. Algorithmes d'apprentissage supervisé	23
3.3. Domaines d'application de l'apprentissage automatique supervisé	25
3.4. Algorithmes d'apprentissage non supervisé.....	25
3.5. Application de l'apprentissage automatique non supervisé :	26
3.6. Avantages et inconvénients de l'apprentissage automatique supervisé	26
4. L'apprentissage en profondeur (Deep Learning).....	26
4.1. Types de réseaux de neurones profonds	27
4.2. Motivation de l'utilisation du Deep Learning.....	29
4.3. Champs d'application.....	29
4.4. Avantages et l'inconvénients de Deep Learning.....	29
4.5. Deep Learning Vs Machine Learning	30
5. Conclusion	31
Chapitre III. Présentation des outils et des Datasets utilisés	32
1. Introduction.....	32
2. Présentation des outils.....	32
2.1. Matériel	32
.2.2 Environnement de travail.....	32
3. Datasets utilisés.....	36
3.1. Dataset de N. gonorrhoeae	36

3.2. Dataset de E.coli.....	37
4. Conclusion	39
Chapitre IV. Travail réalisé et résultats obtenus	40
1. Introduction.....	40
2. Prétraitement des données utilisées pour E.coli.....	40
3. Modèles d'apprentissage automatique.....	41
3.1. Exemple d'un algorithme ML (K-NN)	41
4. Les résultats obtenus en appliquant les différents algorithmes ML et modèles DL	41
4.1. Critères d'évaluation d'un algorithme/modèle.....	41
4.2. Résultats obtenus en appliquant les algorithmes ML.....	42
4.3. Résultats obtenus par les modèles DL	45
.5 Discussion	47
.5.1 Les résultats obtenus en appliquant les algorithmes ML	47
5.2. Les résultats obtenus en construisant des modèles DL	48
6. Interfaces de l'application développée	49
7. Conclusion	51
Conclusion Générale	52
Références Bibliographiques.....	53

Annexes

ملخص

أصبحت مقاومة المضادات الحيوية مصدر قلق كبير للصحة العالمية بسبب عدم فعالية هذه الأخيرة، وانتشار البكتيريا المقاومة، وخيارات العلاج المحدودة، مما استلزم تطوير أساليب مبتكرة للتنبؤ بهذه الظاهرة ومكافحتها. في دراستنا، تم استخدام خوارزميات التعلم الآلي والتعلم العميق لتطوير نماذج تنبؤية تعتمد على ميزات مختلفة مستمدة من البيانات الجينومية والظاهرية والسريية. في دراستنا قمنا بتطوير نموذج يمكنه التنبؤ بمقاومة المضادات الحيوية في بكتيريا النيسيرية البنية والإشريكية القولونية، مما يتيح تحديداً دقيقاً وسريعاً للسلاسل البكتيرية المقاومة. حيث تم الحصول على دقة 99% بالنسبة للنيسيرية البنية و 89% بالنسبة للإشريكية القولونية. لقد أثبت هذا النموذج فعاليته العالية ودقته في التنبؤ بمقاومة المضادات الحيوية.

الكلمات المفتاحية: البكتيريا المقاومة، المضادات الحيوية، النيسيرية البنية، الإشريكية القولونية، التعلم الآلي، التعلم العميق، الشبكات العصبية الاصطناعية

Abstract

Antibiotic resistance has become a critical global health concern due to the ineffectiveness of antibiotics, the spread of resistant bacteria, and limited treatment options, necessitating the development of innovative approaches to predict and combat this phenomenon. In our study, machine learning and deep learning algorithms have been employed to develop predictive models based on various features derived from genomic, phenotypic, and clinical data. We developed a model that predicts antibiotic resistance in *Neisseria gonorrhoeae* and *Escherichia coli* bacteria, enabling accurate and fast identification of resistant bacterial strains. Accuracy was 99% for *Neisseria.g* and 89% for *E.coli*. This model has proven to be highly effective and accurate in predicting antibiotic resistance.

Keywords: Bacteria resistance, Antibiotics, *Neisseria gonorrhoeae*, *Escherichia coli* bacteria, Machine learning, Deep learning, Artificial Neural networks.

Résumé

La résistance aux antibiotiques est devenue un problème de santé mondial critique en raison de l'inefficacité des antibiotiques, de la propagation de bactéries résistantes et des options de traitement limitées, ce qui nécessite le développement d'approches innovantes pour prédire et combattre ce phénomène. Dans notre étude, des algorithmes d'apprentissage automatique et d'apprentissage profond ont été utilisés pour développer des modèles prédictifs basés sur diverses caractéristiques dérivées de données génomiques, phénotypiques et cliniques. Nous avons développé un modèle qui prédit la résistance aux antibiotiques chez les bactéries *Neisseria gonorrhoeae* et *Escherichia coli*, permettant une identification précise et rapide des souches bactériennes résistantes. Une précision de 99 % a été obtenue pour *Neisseria.g* et de 89 % pour *E.coli*. Ce modèle s'est avéré très efficace et précis dans la prédiction de la résistance aux antibiotiques.

Mots clés : Bactéries résistantes, Antibiotiques, *Neisseria gonorrhoeae*, *Escherichia coli*, Apprentissage automatique ; Apprentissage profond, Réseaux de neurones artificiels.

Liste des abréviations

N.G : Neisseria gonorrhoeae

E. coli: Escherichia Coli

IA: Intelligence Artificielle

NCBI: National Center for Biotechnology Information

PDP: Protein Data Bank

ATB: Antibiotique

ML: Machine Learning

K-NN: k-Nearest Neighbors

LR: Logistic Regression

DT: Decision Tree

RF: Random Forest

DL: Deep Learning

RNA: Réseaux de Neurones Artificiels

CNN: Convolutional Neural Networks

DNN : Deep Neural Networks

AMP: Ampicillin

AMX: Amoxicillin

AMC: Amoxicillin-clavulanate

CTZ: Ceftazidime

CTX: Cefotaxime

CXM: Cefuroxime

CET: Cephalothin

GEN: Gentamicin

TBM: Tobramycin

TMP: Trimethoprim

CIP: Ciprofloxacin

Liste de figures

Figure I.1. Type des bases de données biologiques	4
Figure I.2. Coloration de Gram de <i>Neisseria gonorrhoeae</i>	7
Figure I.3. <i>E.coli</i> , coloration de Gram (Grossissement x 100)	10
Figure I.4. <i>E. coli</i> sur milieu de Mac Conkey	10
Figure I.5. Structure des principales β -lactamines.....	13
Figure II.1. Paradigmes d'intelligence artificielle.....	21
Figure II.2. Différents Types d'apprentissage automatique.....	22
Figure II.3. Illustration schématique du KNN	23
Figure II.4. Régression logistique.....	24
Figure II.5. Illustration schématique d'arbre de décision	24
Figure II.6. Illustration schématique de la forêt aléatoire	25
Figure II.7. Une architecture de Deep Learning	27
Figure II.8. Deep Neural Networks DNN	27
Figure II.9. Convolutional Neural Networks CNN.....	28
Figure II.10. Recurrent Neural Networks RNN.....	29
Figure III.1. Anaconda Navigator.....	33
Figure III.2. Les dimensions de la base de données de N.G.....	36
Figure III.3. Les dimensions de la base de données d' <i>E.coli</i>	37
Figure III.4. Présentation du data-set.....	38
Figure III.5. La distribution des antibiotiques par classe.....	38
Figure III.6. Train et Test modèles	39
Figure IV.1. Supprimer les doublons pour la colonne gène.....	40
Figure IV.2. Transformer et réduire les dimensions de la colonne de gènes	40
Figure IV.3. Convertir les noms d'antibiotiques en nombres.....	40
Figure IV.4. Modèle de prédiction pour un algorithme KNN.....	41
Figure IV.5. L'antibiotique Ciprofloxacin.....	42

Figure IV.6. L'antibiotique Azithromycin.....	43
Figure IV.7. L'antibiotique Penicillin	43
Figure IV.8. Résultats de la prédiction en utilisant l'algorithme Logistic Regression LR.....	44
Figure IV.9. Résultats de la prédiction à l'aide de l'algorithme Random Forest RF.....	44
Figure IV.10. Architecture du modèle DNN appliqué sur Neisseria Gonorrhoeae	45
Figure IV.11. Train loss Vs val loss (Neisseria).....	46
Figure IV.12. Train acc Vs val acc (Neisseria).....	46
Figure IV.13. Architecture du modèle DNN appliqué sur E.coli.....	46
Figure IV.14. Train loss Vs val loss (E.coli).....	47
Figure IV.15. Train acc Vs val acc(E.coli)	47
Figure IV.16. Page de connexion.....	49
Figure IV.17. Page principale	50
Figure IV.18. Page d'analyse des données	50
Figure IV.19. Page de prédiction	51

Liste des tableaux

Tableau I.1. Classification de <i>N.gonorrhoeae</i>	6
Tableau I.2. Principales caractéristiques biochimiques de l'espèce (N.G)	8
Tableau I.3. Classification d'E.Coli	9
Tableau I.4. Principales caractéristique biochimiques de l'espèce E.coli	11
Tableau III.1. Description des attributs utilisés dans la prédiction de la résistance de N.G	36
Tableau III.2. Description des attributs utilisés dans la prédiction de la résistance de E.Coli	37
Tableau IV.1. Résultats de précision des prédictions pour N.G	42
Tableau IV.2. Résultats de précision des prédictions pour E.coli	43
Tableau IV.3. Résultats obtenus pour la bactérie <i>Neisseria genorrhoeae</i>	45
Tableau IV.4. Résultat obtenus pour la bactérie E.coli.....	47

INTRODUCTION

Introduction

Selon l'Organisation Mondiale de la Santé OMS, la résistance aux antibiotiques, conjointement avec les changements climatiques et la pénurie d'eau douce, représente la menace la plus sérieuse pour l'humanité. Alexander Flemming, qui a découvert la pénicilline en 1928, avait prédit que l'abus et l'utilisation excessive des antibiotiques mèneraient à une ère post-antibiotique où les infections mineures ne pourraient plus être traitées et deviendraient mortelles. Cette ère s'approche à grands pas, car il existe déjà des super bactéries pan-résistantes dans la population. Nous avons donc besoin de nouvelles technologies pour mettre fin à cette évolution négative (Carlsson, 2019). Il est essentiel de mettre en œuvre un programme institutionnel de gestion des antibiotiques qui surveille l'utilisation correcte des antibiotiques, contrôle les antibiotiques et génère des antibiogrammes.

L'expansion de l'échelle et la complexité inhérente des données biologiques ont encouragé une utilisation croissante de l'apprentissage automatique et son dérivé l'apprentissage profond en biologie pour construire des modèles informatifs et prédictifs des processus biologiques sous-jacents. (Greener et al., 2022).

Les technologies d'Intelligence Artificielle nous permettront non seulement de collecter, de connecter et d'analyser des données à des échelles sans précédent, mais aussi de construire des modèles prédictifs complets couvrant diverses sous-disciplines (Jakhar & Kaur, 2020). Elles rendront possibles des découvertes ciblées (en testant des hypothèses spécifiques) et non ciblées. L'IA pour la biologie sera la technologie transversale qui améliorera notre capacité à faire de la recherche biologique à toutes les échelles. (Hassoun et al., 2021). En outre, ces types d'outils peuvent faciliter le traitement des patients en cas d'urgence médicale, lorsque le médecin ne peut pas attendre les résultats d'une culture bactérienne. Les applications de l'IA dans le domaine des soins de santé pourraient être illimitées, réduisant le temps nécessaire à la découverte de nouveaux médicaments antimicrobiens, améliorant la précision des diagnostics et des traitements, tout en réduisant les dépenses (Rabaan et al., 2022).

Parmi les bactéries pouvant constituer une menace majeure pour la santé publique figurent *Neisseria gonorrhoeae* et *Escherichia coli*. En raison de leur niveau croissant de résistance aux antibiotiques. *Neisseria gonorrhoeae*, la bactérie responsable de la gonorrhée. La gonorrhée est la deuxième infection sexuellement transmissible (IST) la plus courante en Europe, après la chlamydia. Les taux d'infection par la gonorrhée sont en hausse, avec une augmentation de 26 % signalée entre 2017 et 2018 au Royaume-Uni.

De même, la bactérie *Escherichia coli* a une capacité extraordinaire à acquérir et à propager la résistance par transfert horizontal de gènes avec les espèces environnantes et par des mutations dans son génome existant. *E. coli* présente également une grande redondance des voies métaboliques, ce qui favorise la résistance par le biais de l'adaptabilité métabolique (Arango-Argoty et al., 2018).

Dans ce contexte, nous proposons dans ce travail un nouvel outil bio-informatique pour la cartographie des variantes et la prédiction de la résistance aux antibiotiques. La technologie de séquençage de nouvelle génération (NGS) permettant le séquençage du génome entier (WGS) d'isolats bactériens est désormais à la fois peu coûteuse et largement utilisée.

Notre manuscrit est organisé comme suit :

Introduction : Nous avons présenté brièvement notre problématique, ainsi que le but de cette étude.

Chapitre 1 : nous parlerons de bio-informatique et des bactéries qui feront l'objet d'études pour leur résistance aux antibiotiques.

Chapitre 2 : Ce chapitre porte sur le concept d'intelligence artificielle, ses branches (Machine Learning et Deep Learning) et ses applications.

Chapitre 3 : Dans ce chapitre, nous avons abordé les bibliothèques utilisées, les outils mobilisés lors de développement du programme, ainsi que les bases de données relatives aux bactéries étudiées.

Chapitre 4 : Le dernier chapitre comprend la partie expérimentale, les résultats obtenus, et l'interface de l'application réalisée.

Conclusion : Il comporte un résumé du travail réalisé, les apports et les bénéfices, et les difficultés rencontrées au cours de l'établissement du travail.

PARTIE THÉORIQUE

CHAPITRE I :

ANTIBIOTIQUE ET ANTIBIO-RÉSISTANCE

Chapitre I. Antibiotique et Antibio-résistance

1. Définition de la bio-informatique

La bio-informatique est un domaine interdisciplinaire qui associe la biologie, l'informatique, les statistiques et les mathématiques pour analyser et interpréter des données biologiques à grande échelle. Elle implique le développement et l'application d'outils et de techniques informatiques pour la gestion, le stockage, l'analyse et la visualisation de données biologiques, telles que les séquences d'ADN, les structures protéiques et les profils d'expression génétique (Can, 2014).

2. Les bases de données en biologie

2.1. Définition d'une base de données biologique

Les bases de données biologiques jouent un rôle central en bio-informatique. Elles offrent aux scientifiques la possibilité d'accéder à une grande variété de données biologiquement pertinentes, y compris les séquences génomiques d'un éventail de plus en plus large d'organismes. Les principales bases de données de séquences sont :

- **GenBank** : est une base de données de séquences d'ADN et d'ARN qui permet aux chercheurs d'accéder aux informations génétiques d'un large éventail d'organismes (Baxevanis & Bateman, 2015)..
- **UCSC Genome Browser** : est un outil en ligne largement utilisé qui donne accès à une variété de données génomiques, y compris des séquences génomiques annotées, des données sur l'expression des gènes et des informations de génomique comparative (Baxevanis & Bateman, 2015).
- **Ensembl** : est un projet bio-informatique qui fournit une ressource en ligne complète pour les données génomiques d'un large éventail d'espèces, y compris les humains, les animaux, les plantes et les microbes (Baxevanis & Bateman, 2015).

2.2. Type des bases de données biologiques

Plusieurs bases de données sont construites par des chercheurs, des centres et des équipes de recherche et mises à la disposition des jeunes chercheurs. Figure I.1 présente brièvement la structure ; ainsi que des exemples sur ces bases.

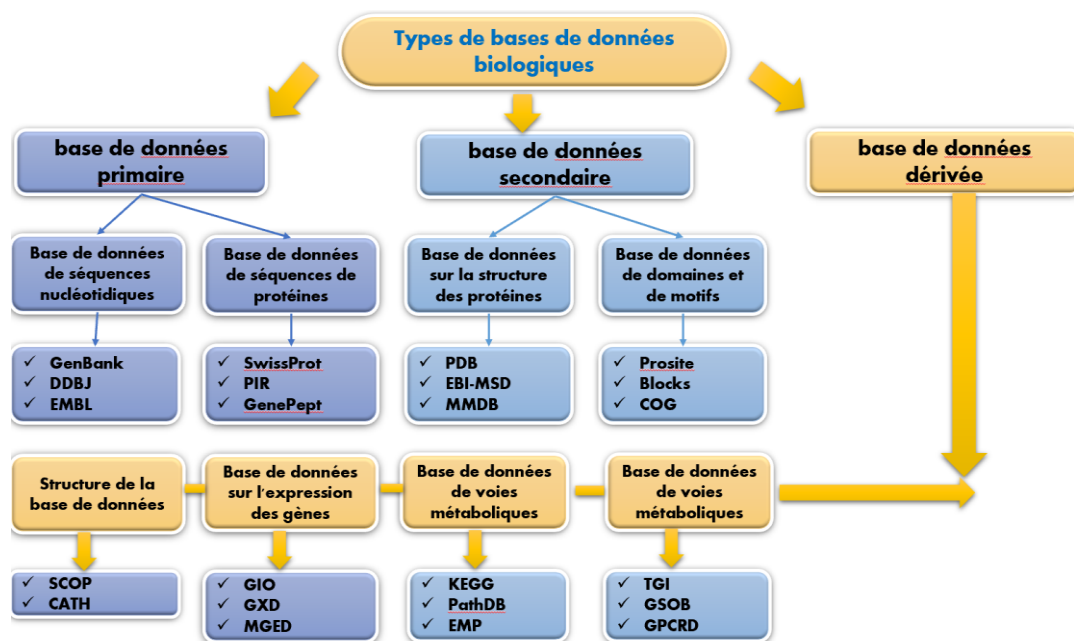


Figure I.1. Type des bases de données biologiques(Sapundzhi & Dzimbova, 2022)

2.3. Prédiction sur les structures et fonction des protéines

Les protéines occupent une place unique dans la bio-informatique structurale. En fait, les origines du domaine lui-même remontent aux travaux pionniers de Max Perutz et John Kendrew visant à déterminer la structure des protéines globulaires (qui ont également donné lieu à l'attribution du prix Nobel de chimie en 1962). Le but ultime de la bio-informatique structurale, lorsqu'il s'agit de protéines, est de découvrir la relation entre les résidus formant une protéine et sa fonction, c'est-à-dire la relation entre le génotype et le phénotype (Torrise et al., 2020).

2.3.1. Méthodes de prédiction de structures 3D

Les méthodes de prédiction de la structure des protéines comprennent la modélisation comparative (modélisation homologique), l'enfilage et l'approche ab initio. Plusieurs outils et logiciels ont été développés pour la modélisation 3D des protéines. Chaque méthode de modélisation des protéines présente certaines limites qu'il convient de garder à l'esprit lors de la sélection d'un outil pour la modélisation de la structure. (Agnihotry et al., 2022).

2.3.2. Prédiction de la structure antibiotique

La bio-informatique peut être utilisée pour prédire la structure des antibiotiques en recourant à des méthodes informatiques telles que la modélisation moléculaire. Cela implique l'utilisation de logiciels et d'algorithmes pour simuler le comportement des atomes et des molécules individuels

et prédire comment ils interagiront les uns avec les autres pour former une structure tridimensionnelle (Ndagi et al., 2020).

L'une des approches consiste à utiliser la modélisation homologique, qui consiste à construire un modèle de la structure de l'antibiotique sur la base de sa similarité avec d'autres structures connues. Pour ce faire, on peut comparer la séquence des acides aminés qui composent l'antibiotique avec d'autres structures connues, puis utiliser ces informations pour prédire la structure tridimensionnelle probable (Ndagi et al., 2020).

2.3.3. Utilisation des données en génomique et protéomique

La génomique et la protéomique impliquent l'étude des informations génétiques et protéiques, respectivement. Dans ces deux domaines, les données jouent un rôle essentiel dans l'analyse et l'interprétation de la grande quantité d'informations générées (DZIUDA, 2010).

- a. La génomique peut être considérée comme l'analyse systématique des données à l'échelle du génome afin d'élargir les connaissances biomédicales. Les études génomiques portent sur la structure et la fonction des gènes et le font simultanément pour tous les gènes d'un génome. L'étude des modèles d'expression des gènes est une partie importante de la génomique fonctionnelle qui se concentre sur l'analyse des gènes et de leurs produits au niveau fonctionnel (DZIUDA, 2010).
- b. La protéomique est l'étude des fonctions et des structures des protéines. La protéomique est souvent considérée comme l'étape suivante de la recherche après la génomique, et comme le domaine qui devrait nous donner des processus biologiques (puisque les protéines sont des acteurs directs de la physiologie cellulaire, alors que les gènes sont essentiellement des entités intermédiaires) (DZIUDA, 2010).

3. Présentation des bactéries étudiées

3.1. *Neisseria gonorrhoeae*

Le gonocoque est un organisme que l'on ne trouve que chez l'homme et qui est très bien adapté à sa niche écologique. Il infecte le plus souvent les surfaces muqueuses, provoquant une urétrite sexuellement transmissible chez l'homme et une endocervicite chez la femme (Tapsall, 2005).

L'histoire de *Neisseria gonorrhoeae*, la bactérie responsable de la gonorrhée, est une histoire de découverte scientifique et de défis de santé publique en cours. En 1879, le médecin allemand Albert Neisser a identifié et nommé la bactérie, ouvrant la voie à un diagnostic amélioré et à des

traitements ciblés. Les tests de diagnostic et les antibiotiques, tels que la pénicilline, sont devenus des outils cruciaux dans la gestion de l'infection. Cependant, l'émergence de souches résistantes aux antibiotiques a compliqué les efforts de traitement, soulignant la nécessité d'une recherche continue sur les thérapies alternatives (Thuener & Clouse, 2013).

3.2. Description et classification de *N.gonorrhoeae*

3.2.1. Description de *N.gonorrhoeae*

Neisseria gonorrhoeae est une espèce de diplocoque à Gram négatif responsable de la gonorrhée, une infection génito-urinaire sexuellement transmissible. Elle est également responsable d'autres formes de maladies gonococciques, notamment la gonococcémie disséminée, l'arthrite septique et l'ophtalmie néonatale gonococcique. *N. gonorrhoeae* est un microbe Gram négatif, diplocoque (de forme ovale), non mobile et aérobie. Il s'agit d'une bactérie non sporulée, non mobile, encapsulée et non acido-alcool-résistante, qui apparaît sous la forme d'un haricot au microscope. Elle n'est jamais présente dans la flore normale, mais a une forte prévalence de maladie avec une faible mortalité (McSheffrey & Gray-Owen, 2015).

3.2.2. Classification de *N.gonorrhoeae*

Neisseria gonorrhoeae est une espèce de bactérie diplocoque Gram négatif appartenant à la famille des Neisseriaceae au sein du phylum des Proteobacteria (sous-groupe β) des Eubacteria. Ils sont classés comme indiqué dans Tableau I.1 ci-après (Limeres Posse et al., 2017).

Tableau I.1. Classification de *N.gonorrhoeae* (L.-K. Ng & I. E. Martin, 2005)

Domaine	Bactéries
Phylum	Pseudomonadota
Classe	Bétaprotéobactéries
Ordre	Neisseriales
Famille	Neisseriaceae
Genre	<i>Neisseria</i>
Espèce	<i>N. gonorrhoeae</i>

3.2.3. Caractères bactériologiques de N.G

➤ Caractères morphologiques

Neisseria gonorrhoeae est un diplocoque gram-négatif qui se présente sous la forme d'une bactérie sphérique ou ovale dont le côté adjacent est concave. Elle ne forme pas de spores, n'est pas mobile et n'est pas capsulée. Le diamètre de la bactérie varie de 0,6 à 1 µm, et elle est généralement observée par paires avec les côtés adjacents aplatis (Maza, 2020).

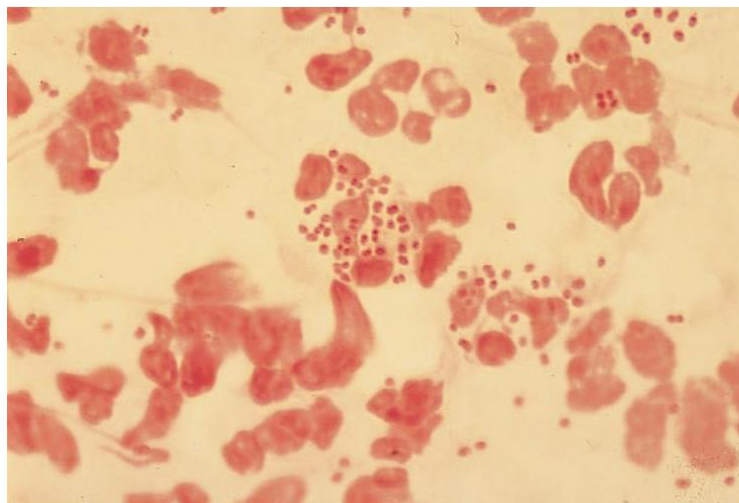


Figure I.2. Coloration de Gram de *Neisseria gonorrhoeae* (Maza, 2020)

➤ Caractères cultureux

Neisseria gonorrhoeae est un organisme très sensible qui nécessite des conditions de culture spécifiques. N.G. a des besoins nutritionnels complexes et se développe facilement dans un milieu contenant du sang ou du sérum, typiquement les milieux gélose au sang, gélose au chocolat et gélose de Mueller-Hinton utilisés pour cultiver *Neisseria gonorrhoeae* en laboratoire. La croissance de ces bactéries est optimale à 35-36°C. Les exigences de N. *gonorrhoeae* en matière de pH vont de 5,8 à 8,0, mais la croissance maximale in vitro est observée entre 7,0 et 7,2. Les N. *gonorrhoeae* sont des bactéries aérobies qui peuvent se développer en présence d'oxygène et des organismes anaérobies qui peuvent se développer en l'absence d'oxygène (Tønjum & van Putten, 2017).

➤ Caractères biochimiques

L'identification de *Neisseria gonorrhoeae* se fait généralement par des tests biochimiques sensibles et précis. Le (Tableau I.2) présente les caractéristiques biochimiques de *Neisseria gonorrhoeae* (L. K. Ng & I. E. Martin, 2005).

Tableau I.2. Principales caractéristiques biochimiques de l'espèce (N.G)(L. K. Ng & I. E. Martin, 2005)

Caractéristique	<i>Neisseria gonorrhoeae</i>
Capsule	-
Catalase	+
Gaz	-
Glucose	+
Maltose	-

3.3. Description et classification de *Escherichia coli*

3.3.1. Description de *Escherichia coli*

Est une bactérie isolée en 1885 par Theodor von Escherich (1857-1911) dans les excréments d'un enfant souffrant de diarrhée et il l'a nommé *Bacterium coli* commune (Mainil, 2013). Elle est couramment appelée "colibacille ". *Escherichia coli* est une entérobactérie mobile capable de fermenter le lactose et de produire de l'indole.

Les *Escherichia coli* font partie de la famille des Enterobacteriaceae. Ce sont des hôtes communs de la microflore intestinale de l'homme et des animaux à sang chaud (mammifères et oiseaux). La majorité des souches d'*E. coli* sont commensales, mais certaines ont toutefois été associées à des pathologies intestinales (Levine, 1987) ou extra-intestinales (Pohl, 1993). Parmi celles-ci, les *Escherichia coli* producteurs de Shiga-toxines (STEC) sont considérés comme des pathogènes émergents importants en santé publique. Ils peuvent être à l'origine de diarrhées souvent sanglantes. Le genre *Escherichia* regroupe cinq espèces : *E. blattae*, *E. coli*, *E. fergusonii*, *E. hermannii* et *E. vulneris*. Chaque espèce d'*Escherichia* possède des caractéristiques biochimiques particulières, permettant de les différencier (Grimont, 1987).

3.3.2. Classification d'*E. Coli*

Il s'agit d'une bactérie Gram-négatif non sporulée faisant partie de la famille des Enterobacteriaceae (voir Tableau I.2). Elle a été caractérisée sur les plans : phénotypique, biochimique et hysiologique. Aujourd'hui, ce sont des techniques basées sur l'utilisation de l'ADN qui permettent une étude génétique des populations et la caractérisation des différentes souches d'*E.coli* (Bouhraoua et al., 2021).

Tableau I.3. Classification d'E.Coli (Bouhraoua et al., 2021)

<i>Règne</i>	<i>Procaryotae</i>
<i>Domaine</i>	Bacteria
<i>Phylum</i>	Proteobacteria
<i>Classe</i>	Gammaproteobacteria
<i>Ordre</i>	Enterobacteriales
<i>Famille</i>	Enterobacteriaceae
<i>Genre</i>	Escherichia
<i>Espèce</i>	Escherichia Coli

Notons que E. coli est une bactérie commune de la microflore commensale intestinale de l'homme et de la plupart des animaux à sang chaud (mammifères et oiseaux). Elle colonise de manière asymptomatique le tractus digestif de l'homme dans les premières heures qui suivent la naissance et constitue l'espèce bactérienne dominante de la flore anaérobie facultative du colon humain. Sa niche écologique se trouve dans la couche de mucus sécrétée par l'épithélium du colon où elle assure, avec les autres composants de la microflore, une barrière de protection de la muqueuse (Pantel et al., 2015)

3.3.3. Caractères bactériologiques de E.Coli

➤ Caractères morphologiques

Bacille à bout arrondi, mesurant approximativement 2 à 4µm de longueur sur 0.6 µm de largeur, ne possédant ni capsule ni spores, il se présente isolé ou en courtes chaînettes, et en quelques cas, sous forme de très long filaments (Figure 1). Pourvu de cils, E. coli est généralement mobile grâce à une ciliature péritriche mais cette mobilité est très variable selon le milieu où la souche a étéensemencée (Achi, 2018). La figure 1 représente une observation microscopique de coloration de Gram d'E. coli.

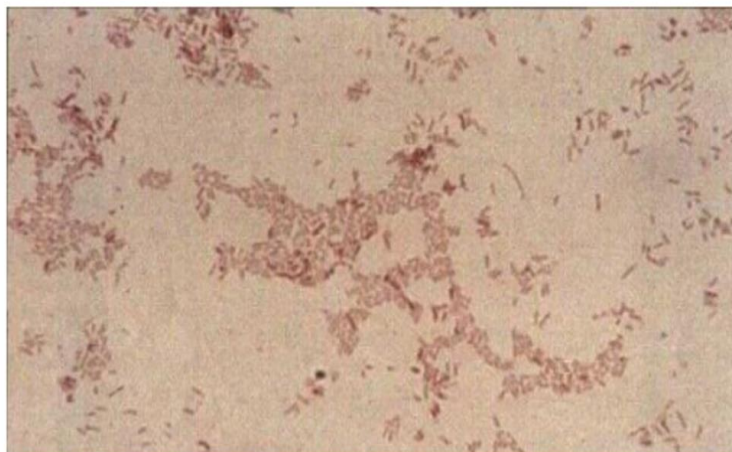


Figure I.3. E.coli, coloration de Gram (Grossissement x 100) (Huff et al., 1999).

➤ **Caractères cultureux**

Elle se développe en 24 heures à 37 °C sur les milieux gélosés en donnant des colonies rondes, lisses, à bords réguliers, de 2 à 3 mm de diamètre, non pigmentées. Sur gélose au sang, les colonies d'E. coli sont lisses, gris terne, et de 2 à 3 mm de diamètre. Sur milieu Mac Conkey, les colonies d'E coli sont lactose-positives, de couleur rose à rouge, plates, sèches et de 2 à 3 mm de diamètre. Elles sont généralement entourées d'une zone rose plus foncée des sels biliaires précipités. Les souches d'E. coli lactose-négatives produisent des colonies incolores sur Mac qui sont de 2 à 3 mm de diamètre (Huff et al., 1999) (Benabdallah, 2016).

La figure 2 représente une culture d'E. coli lactose-positive sur milieu Mac Conkey.



Figure I.4. E. coli sur milieu de Mac Conkey (Huff et al., 1999)

➤ **Caractères biochimiques**

Les tests biochimiques les plus communément utilisés dans l'identification sont : le type de fermentation, l'utilisation du lactose et du citrate, la production d'indole à partir de tryptophane,

l'hydrolyse de l'urée et la production de sulfure d'hydrogène(LADJAL & BAGHDADI). Le tableau ci-après les caractères biochimique de E.Coli.

Tableau I.4. Principales caractéristique biochimiques de l'espèce E.coli (LADJAL & BAGHDADI).

Caractéristique	Escherichia coli
ONPG	+
Oxydase	-
Rouge de méthyle	+
Voges -Proskauer	-
Production d'indole	(Généralement present)
Utilisation de citrate	-
Production de H₂S	-

4. Les antibiotiques

- En 1877, Pasteur et Joubert ont observé qu'un micro-organisme se multiplie mal dans un liquide envahi de moisissures. Ernest Duchesne, en 1897, a remarqué que les palefreniers enduisent de moisissures, recouvrant ainsi, les cuirs placés dans des endroits chauds, humides et sombres des écuries, pour éviter que les plaies de leurs chevaux ne s'infectent. Il décrit ainsi l'inhibition de la croissance des micro-organismes par une moisissure, un Pénicillium(Avril & Rabel, 1980).
- En 1929, Fleming a découvert un Pénicillium sur une boîte de Pétri. Il a mis en évidence l'inhibition du staphylocoque doré par cette culture de Pénicillium. En 1942, Chain a obtenu une forme stable et utilisable in vivo (essais sur souris) de la pénicilline, qui permettrait l'élaboration du premier antibiotique. En 1942, la pénicilline a été produite à l'échelle industrielle et a été utilisée pendant la 2ème guerre mondiale (Avril & Rabel, 1980).

4.1. Définition d'un antibiotique

Un Antibiotique (ATB) est une substance antibactérienne d'origine biologique (champignon ou bactérie), synthétique ou semi synthétique, qui agit sélectivement à faible concentration et en quelques heures sur une étape du métabolisme bactérien(Avril & Rabel, 1980).

4.2. Origines des antibiotiques

Un antibiotique peut être élaboré par un organisme (naturel) ou produit par synthèse (synthétique ou semi- synthétique).

- Les antibiotiques naturels ou produits par les micro-organismes : champignons (pénicillines, céphalosporines) ou bactéries (streptomycine, chloramphénicol, polypeptides).
- Les antibiotiques synthétiques ou produits obtenus entièrement par voie chimique : sulfamides, acide nalidixique.
- Les antibiotiques semi-synthétiques : ces antibiotiques sont obtenus à partir d'une fraction moléculaire naturelle sur laquelle a été greffé un radical chimique.

4.3. Effets des antibiotiques

➤ **Antibiotiques bactériostatiques**

Quand il inhibe seulement la croissance et la multiplication des bactéries : l'infection ne sera finalement vaincue que par les défenses naturelles de l'organisme(Yoon et al., 1992).

➤ **Antibiotique bactéricide**

Quand il détruit les germes. L'étendue de l'activité antibactérienne d'un antibiotique définit son spectre : plus un antibiotique détruit des types différents de bactéries, plus son spectre est large (Yoon et al., 1992).

➤ **Antibiotiques à spectre large**

Il s'agit d'un antibiotique efficace sur un grand nombre de types de germes. Ainsi, l'antibiotique sera actif sur une grande partie de tous les cocci et de tous les bacilles. Les bacilles Gram-négatif sont les plus nombreux. D'une manière indirecte, un spectre large est forcément actif sur les bacilles Gram-négatif. Les antibiotiques à spectre large sont utilisés lorsque le germe n'est pas identifié et que la pathologie peut être dûe à différents types de germes (Róžańska & Osek, 2013).

➤ **Antibiotiques à spectre étroit**

Il s'agit d'un antibiotique efficace sur un nombre limité de types de germes. Cette spécificité lui permet ainsi de cibler un germe et une pathologie (Róžańska & Osek, 2013).

➤ **Antibiotiques à spectre moyen ou limité**

Il s'agit d'un antibiotique dont l'efficacité est réduite ou partielle sur un groupe de germes. Il

concerne aussi un spectre anciennement large et réduit par l'apparition de résistance bactérienne (Róžańska & Osek, 2013).

4.4. Principales familles d'antibiotiques

4.4.1. β -lactamines

Les β -lactamines constituent la famille d'antibiotiques la plus importante aussi bien par le nombre et la diversité des molécules utilisables que par leur indication en thérapeutique (Livermore, 1995). La grande variété de leur mode d'administration, leur large spectre d'activité antibactérien associé à une action bactéricide, une bonne diffusion tissulaire, une bonne tolérance et un faible nombre d'interactions médicamenteuses expliquent l'importance de leur utilisation seule ou en association (Cavallo et al., 2004).

4.4.2. Classification des β -lactamines

Comme illustré dans Figure I.5, les β -lactamines sont classées en quatre groupes :

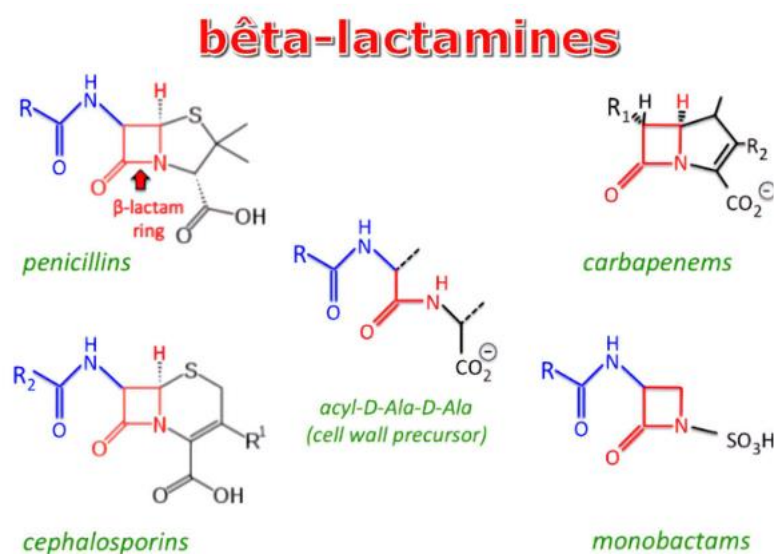


Figure I.5. Structure des principales β -lactamines (Fares & Djedjai, 2018).

A : pénicillines ; B : céphalosporines ; C : céphamycines ; D : acide clavulanique ; E : imipénème ; F : monobactames.

- Les pénicillines également appelées les pénames : les molécules de ce groupe possèdent un cycle thiazolidine accolé au noyau β -lactame. Elles diffèrent par la nature de leur chaîne latérale (Khellafi et al., 2009).

- Les céphalosporines : constituées d'un noyau β -lactame associé à un noyau dihydrothiazine(Moghrani et al., 2018).
- Les carbapénèmes : se distinguent des pénicillines par la présence d'un atome de carbone au lieu d'un soufre en position 1 et une liaison insaturée en C2-C3, également présenté sur les céphalosporinases (Wolff et al., 2009).
- Les monobactames : le noyau est limité au cycle β -lactame. Seul l'aztréonam est à l'heure actuelle utilisé en clinique humaine (Chaabane et al., 2009).

4.5. Aminosides

Les aminosides, également appelés aminoglycosides sont des molécules naturelles produites par des actinomycètes ou obtenues par hémisynthèse. Elles sont constituées de deux ou plusieurs sucres aminés liés par une fonction glycosidique à un noyau hexose. Ce sont des oligosaccharides basiques, polaires et hydrophiles qui sont, d'un point de vue structural, organisés autour d'un cycle central caractéristique de type aminocyclitol sur lequel sont greffés divers sucres aminés (Forge & Schacht, 2000). Les aminosides ont un large spectre antibactérien particulièrement contre les bactéries Gram négatif, peu sensibles aux pénicillines et aux sulfamides. Elles sont majoritairement bactéricides avec une action rapide et dose dépendante qui constitue l'un de leurs grands avantages thérapeutiques (Vakulenko et al., 2003).

5. Modes d'action des antibiotiques

Les quatre principaux modes d'action des antibiotiques sont(Mammeri & Amiens, 2013) :

- Inhibition de la synthèse de la paroi bactérienne, c'est-à-dire l'inhibition la synthèse de peptidoglycane.
- Inhibition de la synthèse de la membrane cytoplasmique.
- Inhibition de la synthèse de l'ADN bactérien.
- Inhibition de la synthèse de protéines bactériennes.

5.1. Antibiotiques agissant sur la synthèse du peptidoglycane

Certaines bactéries sont protégées de l'environnement extérieur par une paroi qui contient en particulier une couche de peptidoglycane plus ou moins épaisse ; un polymère spécifique comportant des acides aminés et des sucres. C'est ce peptidoglycane présent dans la paroi qui contribue à la solidité mécanique de la bactérie. Les antibiotiques inhibant la synthèse du peptidoglycane (notamment les β -lactamines) bloquent les différentes étapes de cette synthèse, ayant ainsi un effet létal pour la bactérie (bactéricide)(Lavollay et al., 2008).

Ces PLP (Protéines Liant la Pénicilline) sont un groupe d'enzymes regroupant les transpeptidases, transglycosylases et carboxypeptidases, impliquées dans la synthèse et le remodelage du peptidoglycane, constituant principal de la paroi bactérienne. La fixation des β -lactamines sur les PLP est facilitée parce que les β -lactamines présentent une analogie structurale entre le noyau β lactame et le dipeptide terminal D-alanyl-D-alanine, du pentapeptide constitutif du peptidoglycane (Stratton et al., 2000). En se fixant sur les PLP, les β -lactamines vont subir l'ouverture de leur cycle et bloquer le fonctionnement de ces enzymes. L'inactivation principalement des transpeptidases déclenche un ensemble de réactions qui aboutit à l'inactivation des inhibiteurs des autolysines bactériennes. Ce sont ces autolysines qui vont alors dégrader le peptidoglycane, entraînant finalement la lyse bactérienne et l'effet d'auto-suicide (Stratton et al., 2000).

5.2. Antibiotiques inhibant la synthèse protéique

La synthèse des protéines est un processus essentiel des cellules vivantes. L'élément central de ce processus dans lequel l'ARN messager est traduit en protéine est le ribosome, l'organite cellulaire qui est responsable de cette étape. Il existe un grand nombre de molécules d'antibiotiques capables de bloquer sélectivement la traduction des protéines chez les bactéries. De ce fait, approximativement la moitié des antibiotiques utilisés en thérapeutique ont pour cible le ribosome bactérien. La grande majorité de ces antibiotiques est bactériostatique, à l'exception des aminosides qui sont bactéricides (Chaussade et al., 2013).

5.2.1. Antibiotiques se fixant sur la sous-unité 30S du ribosome

➤ Mécanisme d'action des aminosides

Le mode d'action des aminosides consiste en leur fixation sur la sous-unité 30S : à concentration subthérapeutique, ils provoquent des erreurs de lecture du code génétique lors de la traduction des protéines, et à concentration thérapeutique, ils inhibent l'élongation de la chaîne peptidique en bloquant le complexe d'initiation (Lavigne, 2007).

➤ Mécanisme d'action des tétracyclines

Ils inhibent la synthèse protéique en se fixant de façon réversible à la sous-unité 30S des ribosomes et en empêchant l'attachement des aminoacyl-ARNt au site A du ribosome, et en inhibant la phase d'élongation de la chaîne polypeptidique (Lavigne, 2007).

5.3. Antibiotiques agissant sur les acides nucléiques

➤ **Mécanisme d'action des quinolones**

Ce sont des antibiotiques bactéricides. Leur mode d'action consiste en l'inhibition sélective de la synthèse de l'ADN bactérien en agissant sur deux enzymes impliquées dans cette synthèse : l'ADN gyrase et l'ADN topo- isomérase IV. Les quinolones altèrent ainsi rapidement la réplication de l'ADN, induisant la mort de la bactérie (Siedner et al., 2007).

➤ **Mécanisme d'action des fluoroquinolones**

Les enzymes bactériennes : l'ADN gyrase ou topoisomérase II et la topo-isomérase IV sont les cibles principales des fluoroquinolones (Muylaert et Mainil, 2013). Les fluoroquinolones interagissent avec le complexe ADN-enzyme c'est-à-dire avec l'ADN gyrase liée à l'ADN bactérien ou avec la topoisomérase IV liée à l'ADN bactérien pour créer des changements de conformation qui aboutissent à l'inhibition de l'activité enzymatique. Le nouveau complexe fluoroquinolone- enzyme-ADN bloque la progression de la fourche de réplication inhibant ainsi la synthèse normale de l'ADN (de Lastours et al., 2012).

6. Antibio-résistances

6.1. Définition

Les définitions les plus fréquemment employés se fondent sur les critères microbiologiques (résistance *in vitro*) et sur les critères cliniques (résistance *in vivo*). Selon la définition microbiologique de terme, une souche est dite résistante lorsqu'elle se cultive en présence de concentration plus élevée en antibiotique comparativement à d'autres souches qui lui sont phylogénétiquement liées (Muylaert & Mainil, 2012).

Par conséquent, la résistance est une propriété qui ne peut être étudiée que par comparaison d'au moins deux souches, dont l'une de référence souvent appelée souche sauvage et développée en laboratoire à partir d'individus prélevés dans la nature, d'une même espèce ou d'un même genre, cultivés dans les mêmes conditions. Selon la définition clinique, une souche est qualifiée de résistante lorsqu'elle survit à la thérapie antibiotique mise en place (Muylaert & Mainil, 2012).

6.2. Types de résistance aux antibiotiques

➤ **Résistance naturelle**

Pour chaque classe d'antibiotiques, il existe des espèces bactériennes sur lesquelles l'antibiotique est inactif par défaut de cible ou d'accès à la cible. On parle d'espèces bactériennes

naturellement résistantes et de mécanismes de résistance intrinsèques. Ceci peut être dû à l'absence de la cible (comme l'absence de paroi chez les mycoplasmes les rendant insensibles aux β -lactamines) ou encore à l'absence de pénétration de l'antibiotique (rôle de la membrane externe par exemple chez les bactéries Gram négatif avec la vancomycine).

➤ **Résistance acquise**

Les bactéries peuvent également acquérir la résistance à un antibiotique. Cette acquisition peut être liée à une (des) mutation(s) modifiant la cible de l'antibiotique, ou un schéma métabolique. Cette acquisition peut être la conséquence d'un transfert horizontal, y compris entre espèces éloignées phylogéniquement. Les gènes de résistance aux antibiotiques chez les micro-organismes producteurs (et par lesquels ils résistent à leurs propres produits) sont généralement localisés sur le chromosome. Le transfert de ces gènes sera rendu plus efficace après leur intégration sur des éléments mobiles tels que plasmides, transposons, intégrons ou encore sur des phages. Ces mécanismes de résistance peuvent alors diffuser très rapidement dans une population (Afssa, 2006).

6.3. Voies de transmission des résistances aux antibiotiques

Différentes voies peuvent être mises en œuvre :

➤ **Transmission verticale**

La diffusion de résistance peut se faire par la voie verticale. En effet, le génome bactérien est soumis à des mutations chromosomiques. Ce phénomène est rare et spontané. Dans le cas d'une mutation codant une résistance à un antibiotique, ce dernier joue le rôle de révélateur, cette résistance représente 10 à 20% de la résistance clinique rencontrée (Lalis, 2008).

➤ **Transmission horizontale**

Les bactéries ont aussi la possibilité d'effectuer un transfert de résistance horizontal, y compris entre des espèces éloignées phylogéniquement, cette transmission peut donc se faire des bactéries pathogènes vers des bactéries commensales ou inversement, ce type de transfert de résistance concerne souvent plusieurs familles d'antibiotiques simultanément (Davison et al., 2001; Lalis, 2008). Dans cette transmission, le transfert des gènes porteurs de résistances est rendu plus efficace après leur intégration sur des petits éléments mobiles : les plasmides qui sont des petites molécules d'ADN circulaires indépendantes du chromosome bactérien et autonomes ; elles sont présentes dans le cytoplasme bactérien de manière facultative (Afssa, 2006) (Lalis, 2008). Trois principaux

mécanismes de transfert horizontal d'éléments génétiques sont connus entre bactéries donneuses et réceptrice d'une même espèce ou d'espèces de genres différents (Cordaux et al., 2012) :

- Transformation : intégration par une bactérie réceptrice, d'un fragment d'ADN nu étranger suite à la lyse d'une autre bactérie.
- Transduction : transfert d'un fragment d'ADN étranger à une bactérie réceptrice par l'intermédiaire d'un vecteur viral (bactériophage).
- Conjugaison : transfert d'un fragment d'ADN issu d'une bactérie donneuse à une bactérie réceptrice, sous la forme d'un plasmide dans la grande majorité des cas. Il s'agit du mécanisme le plus efficace (transfert le plus rapide d'importantes quantités d'information génétique) et donc le plus souvent impliqué dans la dissémination de la résistance aux antibiotiques (Raleigh & Low, 2013).

6.4. Biochimie de la résistance

➤ **Résistance croisée**

La résistance croisée correspond à la résistance à tous les membres d'une classe d'antibiotiques due à un seul mécanisme de résistance (Courvalin, 2008). On peut citer les mutations dans l'ADN gyrase (topoisomérases de type II) et topoisomérase IV au niveau de la région appelée Quinolone Resistance Determining Region (QRDR) des gènes *gyrA* et par C ciles des quinolones et qui sont composées de 2 fois 2 sous-unités codées respectivement dits gènes *gyrA*, *gyrB* et par E qui confèrent la résistance aux fluoroquinolones ou la résistance aux 4-6-desoxystreptamines par méthylation de l'ARN16S (Galimand et al., 2005)

➤ **Co-résistance**

Dans la co-résistance, plusieurs mécanismes de résistance sont associés chez la même bactérie, parfois stabilisés par intégration dans le chromosome. Chacun confère (par résistance croisée) la résistance à une classe d'antibiotiques, ce qui entraîne un large phénotype résistant de la bactérie hôte (Courvalin, 2008).

6.5. Voies biochimiques de la résistance aux antibiotiques

On peut classer les mécanismes de résistance aux antibiotiques en 4 groupes :

➤ **Inactivation enzymatique de l'antibiotique**

La bactérie résistante produit une enzyme capable d'induire une modification de la molécule d'ATB par l'ajout de groupements acétyle, adéninyle ou phosphorique, aboutissant ainsi à son in

activation ou à sa destruction. C'est le mécanisme le plus important quantitativement et qualitativement. Il est utilisé par les bactéries Gram-négatif contre les β -lactamines (Ferrahi, 2021).

➤ **Modification de la cible**

La liaison de l'antibiotique à sa cible est inhibée par une reprogrammation ou camouflage de cette dernière. La molécule ne la reconnaît plus et devient inactive. Ce phénomène est dû à des bactéries qui ont la capacité de mutation d'un gène responsable de la biosynthèse de la protéine sur laquelle agit l'antibactérien (Ferrahi, 2021). Chez *Staphylococcus aureus*, la résistance par modification de PLP, par exemple, est due à la présence d'une PLP ayant une très faible affinité pour les β -lactamines.

➤ **Diminution de la perméabilité**

Les porines sont des protéines formant des pores dans la membrane externe des bactéries à Gram négatif et permettant le passage de certaines molécules hydrophiles (Nauciel et Vildé, 2008). Des mutations peuvent entraîner la perte de certaines porines ou les altérer et de ce fait entraver la pénétration de l'ATB (Bouhraoua et al., 2021)

➤ **Excrétion de l'antibiotique par efflux**

Il existe chez les bactéries des systèmes permettant la non-accumulation de l'antibiotique: c'est l'excrétion ou efflux (Bouhraoua et al., 2021) L'efflux actif est un mécanisme de transport membranaire nécessitant de l'énergie qui pompe l'antibiotique de l'intérieur vers l'extérieur plus vite qu'il ne rentre. Les antibiotiques exerçant leur action sur des cibles cytoplasmiques seront les plus touchés (Bouhraoua et al., 2021).

6.6. Mécanismes de résistances aux d'antibiotiques

➤ **Mécanismes de résistance aux β -lactamines**

Les mécanismes de résistance aux β -lactamines peuvent être divisés en deux grandes catégories, les mécanismes de résistance : enzymatiques et non enzymatiques.

Les mécanismes de résistance enzymatiques sont liés à la production de β -lactamases qui hydrolysent les β -lactamines. On peut schématiquement classer les β -lactamases en deux catégories : les pénicillinases et les céphalosporinases. Les premières ont un support plasmidique ou chromosomique et sont présentes chez les bactéries à Gram positif et à Gram négatif. Les secondes sont d'origine chromosomique et uniquement présentes chez les bactéries à Gram négatif.

Les mécanismes dits « non enzymatiques » sont liés soit à des modifications des PLP, surtout chez les bactéries à Gram positif, soit à des modifications de la membrane externe chez les bactéries à

Gram négatif. L'autre mécanisme de résistance non enzymatique est la conséquence d'une diminution de la perméabilité par modification de différentes structures de la membrane externe. Une diminution quantitative des porines semble être le mécanisme majeur de ce type de résistance chez les entérobactéries, alors que ce serait des modifications du LPS qui expliqueraient la résistance par « imperméabilité »(Mansour et al., 2008).

7. Apport de l'IA sur la prédiction des bactéries résistante

L'intelligence artificielle contribue à prédire la résistance bactérienne et à la traiter de plusieurs façons. Lorsque les dernières techniques d'apprentissage automatique et d'apprentissage profond ont été utilisées pour construire des modèles prédictifs afin de faciliter les recherches futures sur la résistance bactérienne et l'analyse des big data afin d'améliorer le traitement et de réduire le problème de la résistance bactérienne dans la communauté médicale, et nous aborderons ce point dans le chapitre suivant.

CHAPITRE II :

APPRENTISSAGE AUTOMATIQUE ET

APPRENTISSAGE PROFOND

Chapitre II. Apprentissage automatique et Apprentissage profond

1. Préface

L'intelligence artificielle (IA) a fait des progrès considérables ces dernières années. L'IA a été appliquée à un large éventail de secteurs, dont les soins de santé, la finance, les transports et les loisirs, entre autres.

2. Définition

L'intelligence artificielle désigne une machine conçue pour simuler la conscience humaine. L'IA est une vaste branche de l'informatique qui comprend de nombreux termes différents. Pour que les ordinateurs puissent atteindre ces capacités, ils ont besoin d'un grand nombre de données. De grands ensembles de données permettent aux algorithmes de l'IA d'identifier des modèles, de faire des prédictions (Shinde & Shah, 2018).

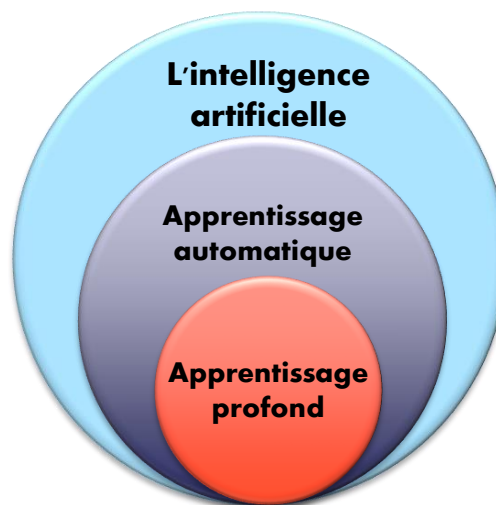


Figure II.1. Paradigmes d'intelligence artificielle.

3. L'apprentissage automatique (Machine Learning)

L'apprentissage automatique est une sous-catégorie de l'intelligence artificielle qui permet aux ordinateurs de reconnaître des formes, ou d'apprendre en permanence, de faire des prédictions basées sur des données et de prendre des décisions sans avoir été spécifiquement programmés pour le faire (Möller, 2023).

3.1. Types d'apprentissage automatique

Généralement, il existe quatre types fondamentaux: de l'apprentissage automatique, notamment :

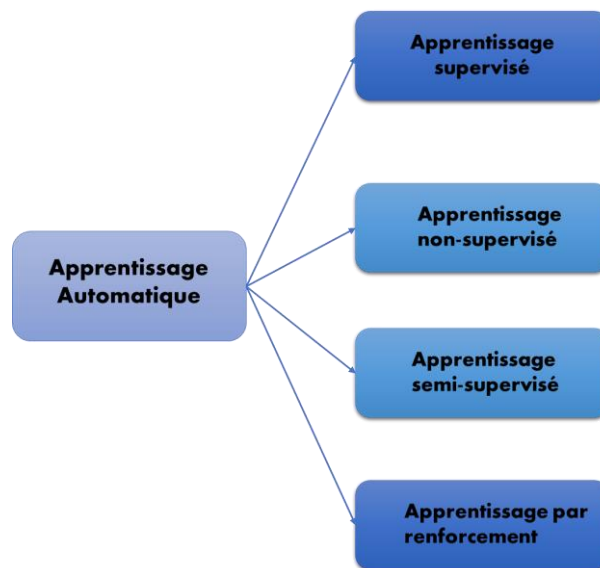


Figure II.2. Différents Types d'apprentissage automatique.

3.1.1. Apprentissage supervisé

Le modèle apprend à généraliser et à faire des prédictions sur de nouvelles données en se basant sur les modèles qu'il a appris à partir des données d'apprentissage (S. GADRI, 2022).

3.1.2. Apprentissage non-supervisé

L'apprentissage non supervisé est très différent. Plutôt que de présélectionner une sortie préférée pour l'algorithme, La classification est faite sur des classes non prédéfinies ce dernier est alimenté par un ensemble de données et reçoit les outils nécessaires pour comprendre ses propriétés(D. S. GADRI, 2022)

3.1.3. Apprentissage semi-supervisé

L'apprentissage automatique semi-supervisé est une combinaison de méthodes d'apprentissage automatique supervisé et non supervisé. L'idée qui sous-tend l'apprentissage semi-supervisé est qu'un modèle peut apprendre à mieux se généraliser en tirant parti des informations supplémentaires fournies par les données non étiquetées(S. GADRI, 2022).

3.1.4. Apprentissage par renforcement

La méthode d'apprentissage par renforcement est une approche par essai et erreur qui permet à un modèle d'apprendre en utilisant le retour d'information de ses propres actions. L'ordinateur reçoit un "retour positif" lorsqu'il comprend ou classe correctement des données et un "retour négatif" lorsqu'il échoue (S. GADRI, 2022).

3.2. Algorithmes d'apprentissage supervisé

3.2.1. Algorithme K-NN (k plus proches voisins / K Nearest Neighbors)

Dans KNN, les objets sont classés en fonction de leur plus proche voisin dans l'ensemble de données. Il utilise l'ensemble des données en tant qu'ensemble d'apprentissage au lieu de les diviser en ensembles de données d'apprentissage et de test (Nabipour et al., 2020).

- Lorsqu'un résultat est requis pour un nouvel ensemble de données, l'algorithme parcourt l'ensemble des données et trouve le k- plus proche voisin, k étant donné par l'utilisateur (Nabipour et al., 2020).

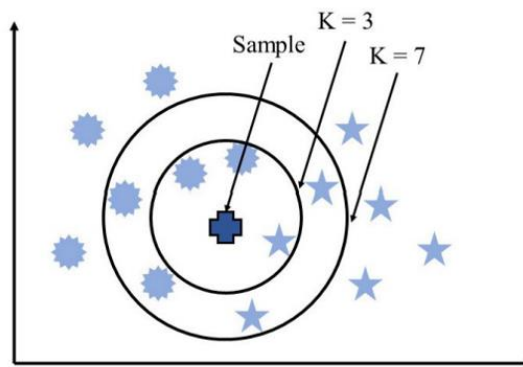


Figure II.3. Illustration schématique du KNN (Nabipour et al., 2020).

3.2.2. Régression logistique (Logistic Regression)

Dans la régression logistique, nous classons les données d'entrée en deux catégories, Vrai ou Faux, en traçant une ligne entre les deux à l'aide d'une fonction de transformation. Elle est principalement utilisée pour les problèmes de classification binaire où nous pouvons séparer les données par une seule ligne et comme base de référence pour l'évaluation de problèmes complexes, dont la probabilité est soit 0, soit 1 (Nasteski, 2017).

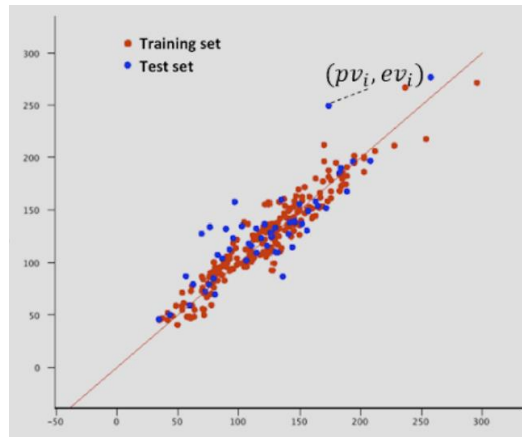


Figure II.4. Régression logistique (Walsh *et al.*, 2021)

3.2.3. Arbre de décision (Decision Tree)

Algorithme d'arbre de décision couramment utilisé pour les tâches de classification et de régression. Il s'agit d'un type d'arbre décisionnel binaire, où chaque nœud interne représente une décision binaire et possède deux branches correspondant aux deux résultats possibles de la décision (Ray, 2019).

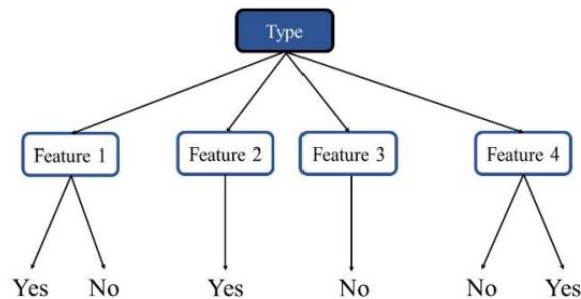


Figure II.5. Illustration schématique d'arbre de décision (Nabipour *et al.*, 2020).

3.2.4. Forêt aléatoire (Random Forest)

Random Forest : est un algorithme d'apprentissage d'ensemble qui utilise plusieurs arbres de décision pour faire des prédictions. Les forêts aléatoires peuvent être utilisées pour résoudre des problèmes de régression (variable cible numérique) et de classification (Nabipour *et al.*, 2020).

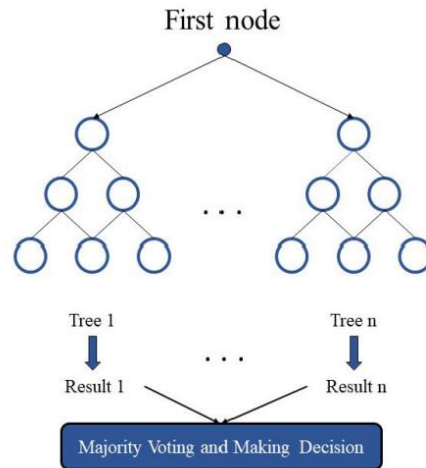


Figure II.6. Illustration schématique de la forêt aléatoire (Nabipour et al., 2020)

3.3. Domaines d'application de l'apprentissage automatique supervisé

L'apprentissage automatique a envahi presque tous les domaines, parmi lesquels, on peut citer les suivants :

1. Bio-informatique.
2. Diagnostic médical.
3. Classification des images.
4. Traitement du langage naturel NLP.
5. Systèmes de recommandation RS.
6. Prévisions météorologiques.
7. Marketing des bases de données.
8. Extraction d'informations.
9. Détection de spams.
10. Reconnaissance d'objets dans le domaine de la vision par ordinateur.
11. Robotique
12. Autres

3.4. Algorithmes d'apprentissage non supervisé

L'apprentissage non supervisé est basé sur une variété d'algorithmes, tels que :

- Algorithme des k-moyennes (K-means)
- Hierarchical clustering
- Principal Component Analysis (PCA)

3.5. Application de l'apprentissage automatique non supervisé :

L'apprentissage non supervisé est aussi utilisé dans plusieurs domaines ; à noter (D. S. GADRI, 2022) :

1. Economie (segmentation de marchés).
2. Médecine (localisation de tumeurs dans le cerveau humain par images Scanner et IRM), etc.
3. Marketing direct (anticipation des commandes).
4. Bio-informatique (séquences d'ADN).
5. Bourse (prédiction des valeurs des actions)
6. Traitement d'images.
7. Reconnaissance de formes

3.6. Avantages et inconvénients de l'apprentissage automatique supervisé

➤ **Avantages :**

- Les expériences des utilisateurs peuvent optimiser les performances des modèles.
- Ils produisent des résultats en utilisant l'expérience antérieure et vous permettent également de collecter des données.
- Les algorithmes d'apprentissage automatique supervisé peuvent être utilisés pour mettre en œuvre un certain nombre de problèmes du monde réel (online).

➤ **Inconvénients :**

- La formation de modèles d'apprentissage automatique supervisé peut prendre beaucoup de temps si l'ensemble de données est plus important.
- La classification des données volumineuses pose parfois un défi plus important.
- On peut être confronté à des problèmes de surajustement.
- Nous avons besoin d'un grand nombre de bons exemples si nous voulons que le modèle soit performant pendant la formation du classificateur.

4. L'apprentissage en profondeur (Deep Learning)

L'apprentissage profond représente un sous-ensemble de l'apprentissage automatique et de l'intelligence artificielle. La technologie DL est issue des réseaux neuronaux artificiels (RNA) pour apprendre à prédire et à prendre des décisions. Il s'inspire de la structure et du fonctionnement du cerveau humain, en particulier du réseau interconnecté de neurones.

- "Profond" est un terme technique qui fait référence au nombre de couches d'un RNA. Il existe trois types de couches : la couche d'entrée (reçoit les données d'entrée), la couche de sortie (produit le résultat du traitement des données) et la couche cachée (produit le résultat du traitement des données) et la couche cachée (extrait les modèles dans les données) (Kaur, 2019) .

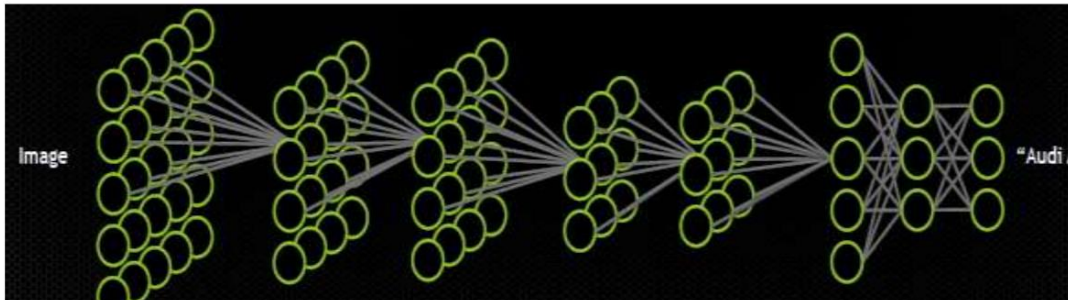


Figure II.7. Une architecture de Deep Learning (Gadri, 2020)

4.1. Types de réseaux de neurones profonds

4.1.1. Un simple réseau de neurones profond (Deep Neural Networks DNN)

Un réseau de neurones profonds (DNN) est un réseau de neurones qui a un certain degré de complexité. Il peut être considéré comme un réseau neuronal empilé, ou comme un réseau composé de plusieurs couches, généralement deux ou plus, comprenant l'entrée, la sortie et au moins une couche cachée entre les deux (Chien, 2019).

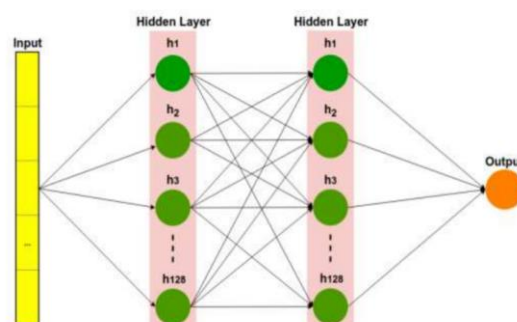


Figure II.8. Deep Neural Networks DNN (Gadri, 2020).

4.1.2. Réseaux de neurones convolutifs (Convolutional Neural Networks CNN)

Les réseaux neuronaux convolutifs (CNN) sont un type de modèle d'apprentissage profond spécifiquement conçu pour le traitement des données de grille structurées, telles que les images et les vidéos. (Gadri, 2020).

Un modèle CNN est constitué de plusieurs types de couches, notamment : des couches de convolution (Conv Layers), des couches d'échantillonnage (Pooling Layers), une couche pleinement connectée (Fully Connected Layer). Dans les couches convolutives, le réseau applique un ensemble de filtres à l'image d'entrée pour en extraire les caractéristiques (Gadri, 2020).

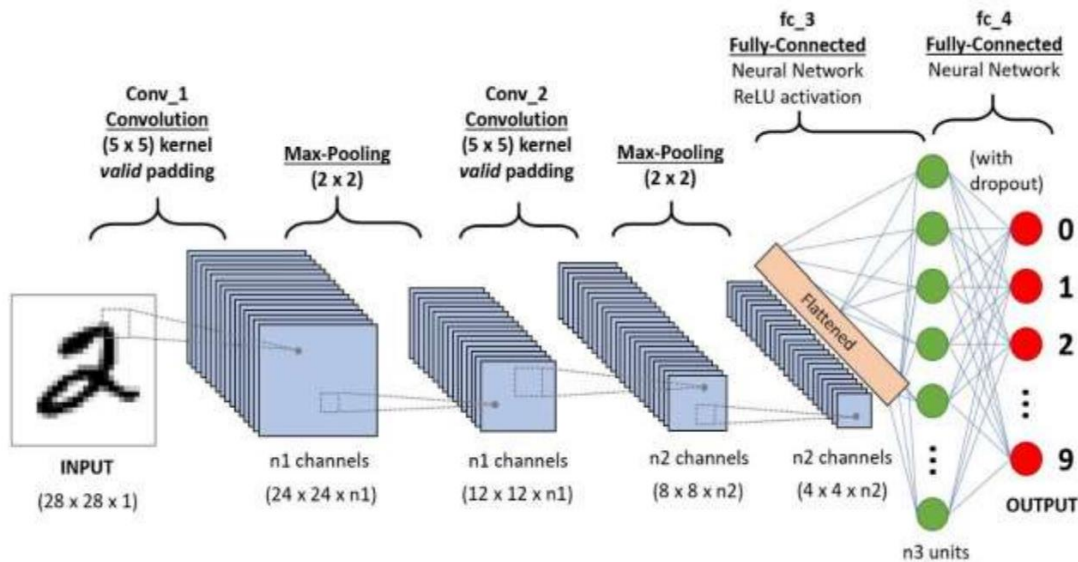


Figure II.9. Convolutional Neural Networks CNN (Gadri, 2020).

4.1.3. Réseaux de neurones récurrents (Recurrent Neural Networks RNN)

Un réseau neuronal récurrent (RNN) est un modèle de réseau neuronal artificiel destiné à traiter des données séquentielles ou temporelles. À la différence des réseaux feedforward traditionnels, les RNN peuvent maintenir des données anciennes ou historiques afin de prévoir les valeurs futures. Les RNN sont une classe de réseaux neuronaux qui permettent d'utiliser les sorties précédentes comme entrées tout en ayant des états cachés. La caractéristique principale des RNN est l'état caché, qui se souvient de certaines informations sur une séquence. Il existe plusieurs variantes d'architectures RNN qui ont été développées au fil des ans pour remédier aux limites de l'architecture RNN standard, l'une de ces variantes étant le réseau à mémoire à long terme (LSTM), qui est capable d'apprendre des dépendances à long terme en se souvenant d'informations pendant de longues périodes.(Gadri, 2020).

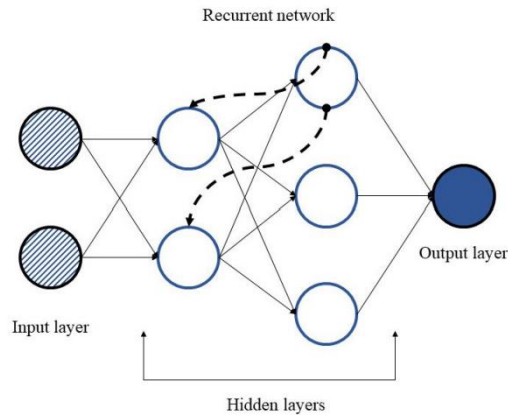


Figure II.10. Recurrent Neural Networks RNN (Nabipour et al., 2020).

4.2. Motivation de l'utilisation du Deep Learning

- Il existe une grande variété de Datasets disponibles.
- Les équipements graphiques (cartes GPU) connaissent une évolution constante.
- Le matériel est relativement abordable en termes de prix.
- Les techniques et les méthodes de traitement évoluent constamment (Gadri, 2020).

4.3. Champs d'application

1. Le Deep Learning au service de la rédaction de textes.
2. (Cyber)sécurité.
3. La traduction.
4. Pour la conduite autonome.
5. Les robots industriels.
6. En médecine.
7. Le Deep Learning dans le secteur du marketing et de la vente.
8. Dans le domaine de l'agriculture.
9. Classification des images.

4.4. Avantages et l'inconvénients de Deep Learning

➤ Avantages de l'apprentissage profond

- Précision accrue : Les algorithmes d'apprentissage profond peuvent atteindre des niveaux de précision plus élevés que les algorithmes d'apprentissage automatique traditionnels.

- Extraction automatisée des caractéristiques : Les algorithmes d'apprentissage profond peuvent extraire automatiquement des caractéristiques pertinentes à partir de données brutes, réduisant ainsi la nécessité d'une ingénierie manuelle des caractéristiques.
- Évolutivité : Les algorithmes d'apprentissage profond peuvent s'adapter à de grands ensembles de données et à des problèmes complexes, ce qui les rend adaptés aux applications de big data.

➤ Inconvénients de l'apprentissage profond

- Besoin de grandes quantités de données : Les algorithmes d'apprentissage profond requièrent de considérables quantités de données labellisées pour s'entraîner avec précision.
- Ressources informatiques : Les modèles d'apprentissage profond nécessitent d'importantes ressources informatiques, y compris des GPU avancés, pour s'entraîner et fonctionner efficacement. Cela peut être onéreux et long.
- Interprétabilité : Les modèles d'apprentissage profond peuvent être difficiles à interpréter, car les modèles sont souvent une boîte noire, ce qui rend difficile la compréhension de la façon dont ils prennent des décisions.

4.5. Deep Learning Vs Machine Learning

- L'apprentissage automatique est une approche visant à former les ordinateurs et les machines à apprendre à partir de données antérieures afin qu'ils puissent déterminer les données ou les comportements futurs.
- L'apprentissage profond est une branche de l'apprentissage automatique dans laquelle les techniques et algorithmes de réseaux neuronaux artificiels sont utilisés pour former et apprendre les données non supervisées.
- Machine learning nécessite des données structurées, tandis que deep learning peut analyser des images, des vidéos et des données non structurées d'une manière que l'apprentissage automatique ne peut pas facilement faire.
- Dans le passé, il n'était possible de connecter que quelques couches de nœuds en raison de simples limitations informatiques. Cependant, au fur et à mesure que la technologie s'est améliorée, il est devenu possible de construire des réseaux neuronaux profonds avec davantage de couches cachées.
- ML est constitué de modèles, de processus et d'outils qui soutiennent la technologie.

- DL est ML plus grande échelle avec des réseaux neuronaux plus sophistiqués, des ensembles de données exponentiellement plus grands et une plus grande puissance de calcul.

5. Conclusion

Ce chapitre a présenté les principaux concepts de l'apprentissage automatique et de l'apprentissage profond, expliquer la relation qui existe entre les deux, ainsi que les algorithmes les plus utilisés dans le domaine de prédiction, et que nous avons adopté dans notre étude.

PARTIE PRATIQUE

CHAPITRE III :

PRÉSENTATION DES OUTILS ET DES

DATASETS UTILISÉS

Chapitre III. Présentation des outils et des Datasets utilisés

1. Introduction

Après avoir étudié en deuxième partie l'IA, le ML et le DL, je présente dans cette partie une étude approfondie sur l'environnement de programmation et les outils utilisés.

2. Présentation des outils

2.1. Matériel

La programmation est souvent un processus itératif, et certaines tâches prennent naturellement plus de temps en raison de leur complexité. Les expériences ont toutes été réalisées sur une machine offrant des performances acceptables et dont les caractéristiques sont les suivantes :

- Un ordinateur portable Lenovo i7-10610U CPU @ 1.80GHz 2.30 GHz
- Taille de la RAM 16.0 GB
- Taille du disque dur 256GB SSD
- Intel(R) Core (TM)
- 64-bit Windows 11 Pro système opérationnel

2.2. Environnement de travail

2.2.1. Python

Python est actuellement l'un des langages les plus populaires pour les applications scientifiques. Il possède une nature interactive de haut niveau et une riche collection de bibliothèques scientifiques qui en font un bon choix pour le développement algorithmique et l'analyse exploratoire des données. Il est de plus en plus utilisé dans les établissements universitaires et dans l'industrie. Il contient un module célèbre appelé scikit-learn tool intégrant un grand nombre d'algorithmes de ML pour des problèmes supervisés et non supervisés tels que : arbres de décision, régression logistique, Naïve Bayes, KNN, ANN, etc. Cet ensemble d'algorithmes permet de simplifier le ML pour les non-spécialistes travaillant sur un usage général (S. Gadri, 2020).

2.2.2. Anaconda

Anaconda est une plateforme open-source qui fournit des kits d'outils de science des données et des paquets intégrés pour effectuer diverses tâches de science des données. Il s'agit de la

plateforme de science des données la plus populaire au monde et elle est utilisée par les scientifiques des données, les entreprises axées sur les données et les communautés open source. Anaconda est le berceau de la science des données Python et comprend plus de 6 000 bibliothèques Python et plus de 2 000 paquets R interopérables. Anaconda est géré par le système de gestion de paquets conda, qui analyse l'environnement actuel avant d'exécuter une installation afin d'éviter de perturber d'autres cadres et paquets.

- Anaconda Navigator offre un moyen pratique de créer et de gérer des environnements Python, d'installer et de mettre à jour des paquets, et de lancer des applications telles que Jupyter Notebook, JupyterLab, Spyder et VS Code, qui peuvent être considérées comme des IDE. Dans notre travail expérimental, nous utilisons Anaconda 22.9.0 qui est compatible avec python 3.9.13.

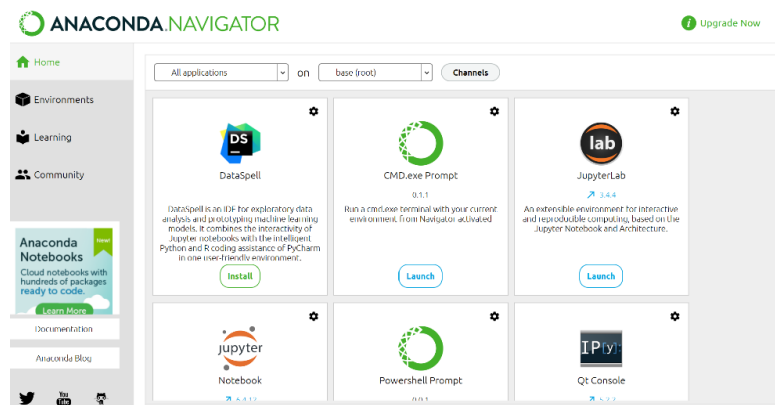


Figure III.1. Anaconda Navigator

2.2.3. Installation Anaconda

A. Téléchargement

Pour télécharger Anaconda pour Windows 11 Pro

- Cliquez sur ce lien <https://www.anaconda.com/download>
- Choisissez l'anaconda

B. Installation

- Chercher dans Google « Anaconda Python » ➡ Python V pour Windows 11 pro.
- Installer le fichier d'installation d'Anaconda 64-bit pour Windows Installer pour ➡ l'option : Juste pour moi.
- Vous Pouvez ajouter le chemin d'installation à l'environnement des variables (Rubrique : PATH) pour faciliter l'accès.

- Après la fin de l'installation ouvrir Anaconda Prompt (Menu Démarrer → Anaconda → Anaconda Prompt).
- Ouvrez Anaconda Prompt en mode ADMINISTRATEUR. Ceci vous évitera des erreurs de droits en écritures par la suite lors de l'installation des bibliothèques / paquets. (Anaconda prompt → bouton droit → menu → mode administrateur → C:\WINDOWS\system32>.
- Tapez la commande d'installation des bibliothèques.

C. Bibliothèques Python et Formats Utilisés

- Sklearn

Scikit-learn est une bibliothèque Python associée à NumPy et SciPy. Elle est considérée comme l'une des meilleures bibliothèques pour travailler avec des données complexes.

Il constitue une bibliothèque de Machine Learning pour le langage de programmation Python. Après avoir nettoyé et manipulé vos données avec Panda ou NumPy, Scikit-learn est utilisé pour construire des modèles d'apprentissage automatique, car il dispose de milliers d'outils utilisés pour la modélisation et l'analyse prédictive. Il dispose de divers algorithmes de classification, de régression et de regroupement, notamment les machines à vecteurs de support, les forêts aléatoires, le gradient boosting, 3 k-means et DBSCAN(Hao & Ho, 2019). Dans cette étude, nous avons installé sklearn : 1.0.2.

- SciPy

SciPy (Scientific Python) est une autre bibliothèque Python libre et gratuite souvent utilisée dans la recherche de données pour des calculs complexes (Saabith et al., 2021). Dans cette étude, nous avons installé SciPy1.9.1.

- Pandas

Pandas est une bibliothèque Python d'apprentissage automatique qui offre des structures de données de haut niveau et une variété d'outils analytiques. La capacité de cette bibliothèque à traduire des actions complexes avec des données en utilisant une ou deux commandes est l'une de ses meilleures caractéristiques(Saabith et al., 2021). Pandas 1.4.4 a été installé pour ce projet.

- NumPy

NumPy est une bibliothèque open source qui contient des tableaux multidimensionnels. Le ndarray NumPy peut être utilisé pour stocker des données dans un objet tableau homogène à "n" dimensions. NumPy est utilisé dans l'industrie pour calculer des tableaux, NumPy est également

utilisé par des bibliothèques Python avancées comme Pandas et SciPy. NumPy est plus efficace que List de Python en termes de : Vitesse, Mémoire (Gholizadeh, 2022). Dans cette étude, nous avons installé 1.21.5.

- **Matplotlib**

La bibliothèque la plus utilisée pour l'exploration et la visualisation des données est Matplotlib, qui permet de créer des graphiques de base, notamment des graphiques linéaires, des diagrammes de dispersion, des histogrammes, des diagrammes à barres et des diagrammes circulaires (Gholizadeh, 2022). Dans cette étude, nous avons installé 3.5.2.

- **Tensorflow**

TensorFlow est une bibliothèque libre à usages multiples pour le calcul numérique à l'aide d'organigrammes de données. Elle offre des API pour les débutants et les experts afin de développer des applications pour les ordinateurs de bureau, les téléphones portables, le web et le cloud. TensorFlow peut être utilisé à partir de nombreux langages de programmation tels que Python, C++, Java, Scala, R, et Runs sur une variété de plateformes incluant Unix, Windows, iOS, Android. Nous notons également que Tensorflow peut être exécuté sur des machines simples (CPU, GPU, TPU) ou des machines distribuées de plusieurs centaines de cartes GPU (Gadri & Adouane, 2022).

- **Keras**

Keras est l'API de haut niveau officielle de TensorFlow qui se caractérise par de nombreuses caractéristiques importantes : Bibliothèque de réseaux neuronaux minimaliste et hautement modulaire écrite en Python, Capable de fonctionner au-dessus de TensorFlow ou Theano, Large adoption dans l'industrie et la communauté de recherche, Production facile de modèles, Supporte à la fois les réseaux convolutifs et les réseaux récurrents et les combinaisons des deux, Supporte des schémas de connectivité arbitraires (y compris l'entraînement multi-entrées et multi-sorties), Fonctionne de manière transparente sur le CPU et le GPU (Gadri, 2021).

- **Le format CSV**

Un fichier CSV est un simple fichier texte que vous pouvez ouvrir dans un grand nombre de programmes, y compris tout programme fonctionnant avec du texte brut, comme l'application Bloc-notes.

3. Datasets utilisés

3.1. Dataset de *N. gonorrhoeae*

Nous avons utilisé un ensemble de données composé de Unitigs, qui sont de courtes séquences d'ADN. Unitigs constituent un moyen efficace et flexible de représenter la variation de l'ADN des bactéries. Elle contient 3786 séquences génétiques.

```
In [1]: runfile('D:/Datasets/untitled2.py', wdir='D:/Datasets')
##### Dimension of metadata #####
(3786, 31)
```

Figure III.2. Les dimensions de la base de données de N.G

- Le fichier CSV utilisé dans notre étude contient 31 colonnes. Tableau III.1 Décrit les attributs les plus importants de la base de données de *N.gonorrhoeae* utilisée dans la partie expérimentale.

Tableau III.1. Description des attributs utilisés dans la prédiction de la résistance de N.G

Variable	Description
ID_échantillon	Une base de données composée de Unitigs (parties d'ADN de la bactérie <i>N.gonorrhoeae</i>)
Pays	UK, Canada, Italy, Austria, ect...
Continent	Europe, Africa, America, Oceania, Asia
Beta lactamase	Il est représenté par deux symboles(S,R) : Résistance(R), Sensible(S)
Antibiotiques	Azithromycin,Ciprofloxacin, Ceftriaxone, Cefixime, Tetracycline, Penicillin.
NG_MAST(<i>N.gonorrhoeae</i> Multi-Antigen Sequence Typing)	NG-MAST est utilisé dans les études épidémiologiques et la surveillance pour comprendre la diversité génétique des souches de <i>N. gonorrhoeae</i> et suivre la propagation de différents génotypes.
ATB_MIC (e.g. azm_mic)	Concentration Minimum Inhibitrice : Il représente la concentration la plus faible de l'antibiotique qui inhibe la croissance visible du micro-organisme (il a été utilisé avec six antibiotiques).
log2_ATB_mic(e.g. log2_azm_mic)	Cette transformation peut être utilisée à diverses fins d'analyse ou de recherche, telles que rendre les données MIC plus interprétables, normaliser les données ou faciliter l'analyse statistique.
Phynotype	Nous appelons les différents modèles de résistance « phénotypes » (e.g. azm_sr, il a été utilisé avec six antibiotiques)

- Nous disposons de trois fichiers de données : Metadata.csv pour les données réservées à l'apprentissage, Valid.csv pour valider le modèle, Test.csv pour tester le modèle sur des données externes. Nous avons utilisé un ensemble de données d'environ 2720 exemples pour l'apprentissage des classificateurs, tandis que 681 exemples pour le test.
- Neisseria gonorrhoeae est considéré résistante (cible=1) à un traitement particulier, soit non résistante (cible=0) à un traitement particulier.

3.2. Dataset de E.coli

Nous avons utilisé un ensemble de données d'article: Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data que nous avons obtenus de NCBI (Moradigaravand et al., 2018), elle contient 98756 séquences génétiques.

```
▶ #Display database dimensions
print(bec.shape)

(98756, 6)
```

Figure III.3. Les dimensions de la base de données d'E.coli

- Le fichier CSV contient six colonnes. Tableau III.2 Décrit les différentes colonnes de la base de données de E.coli.

Tableau III.2. Description des attributs utilisés dans la prédiction de la résistance de E.Coli

Variable	Description
Gene (Sequences)	Données de séquence du génome de la bactérie Escherichia coli + gène accessoire
Fonction	Protéine non caractérisée ; protéine hypothétique ; Bêta-lactamase
Moyenne Importance 50 répliques	Une mesure de la valeur centrale dans la population.
Rank moyen dans 50 répliques	Il est souvent utilisé pour évaluer l'importance relative ou la performance de différentes variables ou entités.
Fréquence d'apparition dans 50 répliques	Désigne le nombre de fois qu'un élément, un événement. Il s'agit d'une mesure de la fréquence à laquelle une chose se produit ou est observée.
Antibiotique	AMP, AMX, AMC, CTZ, CTX, CXM, CET, GEN, TBM, TMP, CIP.

➤ **Présentation dataset (E.coli)**

quences_available_in_Github)	Function	Average_Importance_50_replica	Average_Rank_in_50_replica	Frequency_of_appearance_in_50_replica	Antibiotic
menA	Uncharacterized protein	0.018865	1888.812500	16.0	AMP
group_42361	Uncharacterized protein	0.002059	2085.000000	2.0	AMP
menA	Uncharacterized protein	0.002319	691.333333	3.0	TBM
group_17190	hypothetical protein	100.000000	1.000000	50.0	CTX
wbuC	hypothetical protein	28.467702	3.340000	50.0	CTX

Figure III.4. Présentation du data-set

➤ **Distribution des antibiotiques par classe**

```

#Display the distribution of elements by class
print(bec.groupby('Antibiotic').size())

```

```

Antibiotic
AMC      6730
AMP     16982
AMX     14859
CET      4977
CIP      7393
CTX      5430
CTZ     14170
CXM      8832
GEN      7857
TBM      8372
TMP      3149
dtype: int64

```

Figure III.5. La distribution des antibiotiques par classe

- Nous disposons de trois fichiers de données : Dataset.csv pour les données d'apprentissage, Valid.csv pour la validation du modèle, Test.csv pour le test. Nous avons utilisé un ensemble de données d'environ 79004 exemples pour l'apprentissage des classificateurs, tandis que 19752 avis pour le modèle de test.

```
▶ print(X_train.shape)
print(Y_train.shape)
print(X_test.shape)
print(Y_test.shape)

(79004, 5)
(79004,)
(19752, 5)
(19752,)
```

Figure III.6. Train et Test modèles

4. Conclusion

Dans ce chapitre, nous avons présenté tous les outils de l'environnement de programmation utilisés au cours de notre étude (la partie expérimentale). Nous avons également décrit les bases de données utilisée dans la prédiction de la résistance pour les deux bactéries N.G et E.Coli.

**CHAPITRE IV : TRAVAIL
RÉALISÉ ET RÉSULTATS
OBTENUS**

Chapitre IV. Travail réalisé et résultats obtenus

1. Introduction

Ce chapitre présente la partie expérimentale et contient donc les éléments suivants :

- Le prétraitement de l'ensemble des données utilisées (E.Coli)
- Les algorithmes de ML et les modèles de DL proposés pour la prédiction
- L'interface de notre application.

2. Prétraitement des données utilisées pour E.coli

- Suppression des doublons : Identification et suppression des doublons dans l'ensemble de données (gène, fonction, antibiotique).

```
#duplicated Gene
bec.drop_duplicates("Gene(Sequences_available_in_Github)")
```

Figure IV.1. Supprimer les doublons pour la colonne gène

- Transformation des données : Feature scaling : mise à l'échelle des caractéristiques numériques dans un intervalle cohérent :
 - ✓ La colonne Gène est convertie en **speciesNorm**
 - ✓ La colonne Fonction est convertie en **typeNorm**.
- Réduction de la dimensionnalité : Réduction du nombre de caractéristiques dans l'ensemble de données tout en préservant les informations les plus pertinentes.

```
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
#Encode labels in column 'species'.
bec['species'] = label_encoder.fit_transform(bec['Gene(Sequences_available_in_Github)'])
bec['species'].unique()
bec['speciesNorm'] = list(np.array(bec['species'].tolist())/26699)
```

Figure IV.2. Transformer et réduire les dimensions de la colonne de gènes

- Codage des variables catégorielles : Conversion des variables catégorielles pour chacune des trois colonnes (Gène, Fonction, Antibiotique) en représentations numériques.

```
bec.replace({"Antibiotic":{"AMP":1,'AMX':2,'CTZ':3,'CMX':4,'TBM':5,'GEN':6,'CIP':7,'AMC':8,'CTX':9,'CET':10,'TMP':11}},inplace=True)
bec.drop_duplicates("Antibiotic")
```

Figure IV.3. Convertir les noms d'antibiotiques en nombres

- Fusion d'ensembles de données : Combinaison de plusieurs ensembles de données contenant des informations connexes en un seul ensemble de données à des fins d'analyse.

3. Modèles d'apprentissage automatique

Dans notre étude, nous avons utilisé quatre (04) algorithmes d'apprentissage: algorithme de K-Nearest Neighbors (K-NN), Logistic Regression (LR), CART Decision Tree et algorithme de Random Forest (RF).

3.1. Exemple d'un algorithme ML (K-NN)

L'un des algorithmes ML fréquemment utilisés est l'algorithme KNN (pour E.coli)

```
print("Prediction Results Using KNN Algorithm")
# Call the classifier KNeighborsClassifier with k=11
knn = KNeighborsClassifier(n_neighbors=11)
# Start classifier training
knn.fit(X_train,Y_train)
# Predict on dataset which model has not seen before
y_pred = knn.predict(X_test)
# Calculate Model Accuracy

print("Model Accuracy:",knn.score(X_test, Y_test))

print(confusion_matrix(Y_test, y_pred))
print(classification_report(Y_test, y_pred))
```

Figure IV.4. Modèle de prédiction pour un algorithme KNN

4. Les résultats obtenus en appliquant les différents algorithmes ML et modèles DL

4.1. Critères d'évaluation d'un algorithme/modèle

Il existe plusieurs critères pour évaluer la performance d'un algorithme ML ou un modèle DL, parmi ces critères on peut citer les suivants :

- Accuracy : mesure la proportion d'instances correctement classées.
- Precision : Calcule le rapport entre les vrais positifs et la somme des vrais positifs et des faux positifs, indiquant la capacité du modèle à identifier correctement les instances positives.
- Recall : calcule le rapport entre les vrais positifs et la somme des vrais positifs et des faux négatifs, ce qui représente la capacité du modèle à trouver toutes les instances positives.
- F1 Score : Moyenne harmonique de la précision et du rappel, fournissant une mesure équilibrée des performances du modèle.

4.2. Résultats obtenus en appliquant les algorithmes ML

4.2.1. Résultats de *N. gonorrhoeae*

En appliquant les algorithmes ML et des modèles DL, nous avons abouti à des résultats très précis. Dans cette section nous présentons les valeurs de précision obtenue pour chaque algorithme.

Tableau IV.1. Résultats de précision des prédictions pour N.G

Algorithme	Précision (Accuracy)
K-NN	99.4%
Logistic Regression	99.3%
CART Decision Tree	99.8%
Random Forest	99.9%

Ces valeurs représentent la capacité des algorithmes à effectuer des prédictions précises sur l'ensemble de données.

4.2.2. Quelques statistiques complémentaires

- Diagramme circulaire de l'antibiotique ciprofloxacine est soit résistante (cible=1) à un traitement particulier, soit non résistante (cible=0) à un traitement particulier.

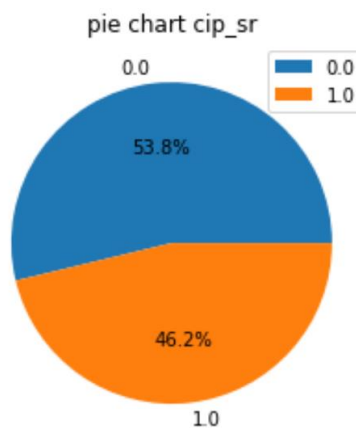


Figure IV.5. L'antibiotique Ciprofloxacine

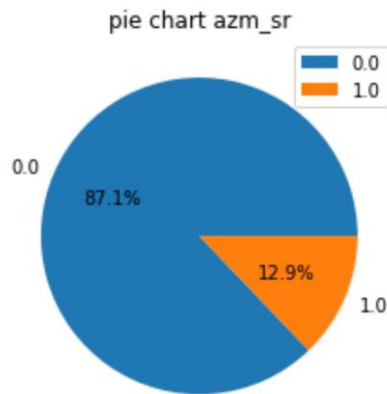


Figure IV.6. L'antibiotique Azithromycin

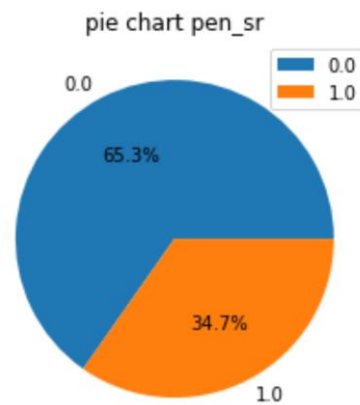


Figure IV.7. L'antibiotique Penicillin

4.2.3. Résultats Obtenus pour E.coli

Pour le cas de la bactérie, E.Coli, nous avons obtenus des résultats de précision moyenne très satisfaisants en appliquant les algorithmes de ML. Voici quelques illustrations:

Tableau IV.2. Résultats de précision des prédictions pour E.coli

Algorithme	Précision (Accuracy)
K-NN	53%
Logistic Regression	39%
CART Decision Tree	86%
Random Forest	89%

Ci-après un histogramme représentant les résultats de prédiction et la précision obtenu pour chacun des algorithmes LR et RF.

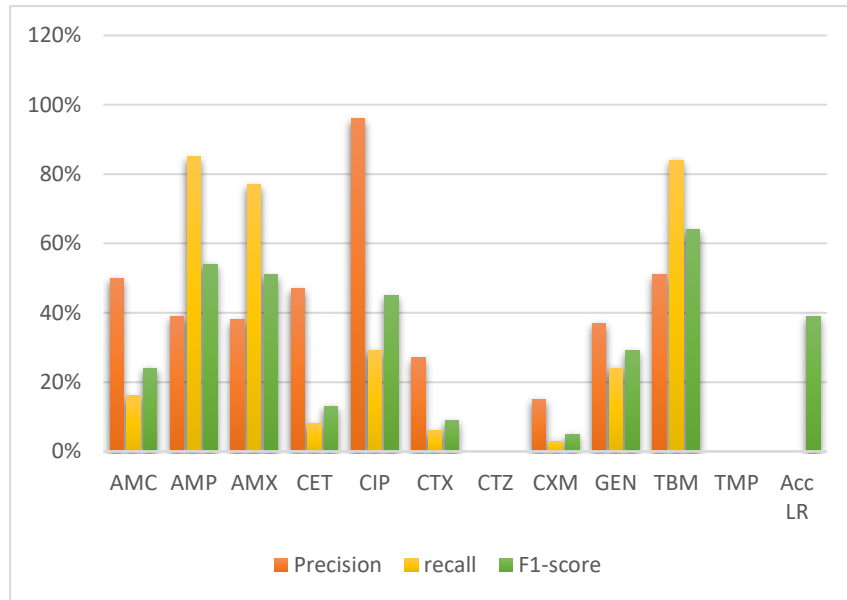


Figure IV.8. Résultats de la prédiction en utilisant l'algorithme Logistic Regression LR



Figure IV.9. Résultats de la prédiction à l'aide de l'algorithme Random Forest RF

4.3. Résultats obtenus par les modèles DL

4.3.1. Modèles DNN

Nous avons construit deux modèles DNN (Deep Neural Network), un pour la prédiction de la résistance pour la bactérie N.G, et l'autre pour la prédiction sur E.Coli. Ci-après les résultats obtenus après les phases d'apprentissage et de test.

➤ **Modèle DNN appliqué sur Neisseria gonorrhoeae**

```

Layer (type)                Output Shape                Param #
-----
dense (Dense)                (None, 16)                  6160
dense_1 (Dense)              (None, 8)                   136
dense_2 (Dense)              (None, 1)                    9
-----
Total params: 6,305
Trainable params: 6,305
Non-trainable params: 0
    
```

Figure IV.10. Architecture du modèle DNN appliqué sur Neisseria Gonorrhoeae

De même, on a exécuté le modèle sous les conditions suivantes : *Epochs = 7, Batch_size = 32*

➤ **Résultats obtenus**

Les résultats obtenus pour la bactérie Neisseria gonorrhoeae sont très satisfaisants, comme les montre le tableau ci-après :

Tableau IV.3. Résultats obtenus pour la bactérie Neisseria gonorrhoeae

Train_loss	0.007
Train_accuracy	99.91%
Test_loss	0.012
Test_accuracy	99.85%

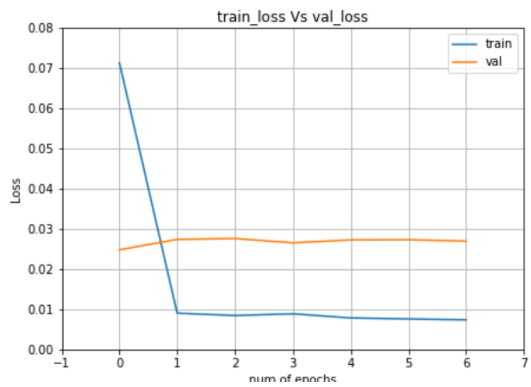


Figure IV.11. Train loss Vs val loss (Neisseria)

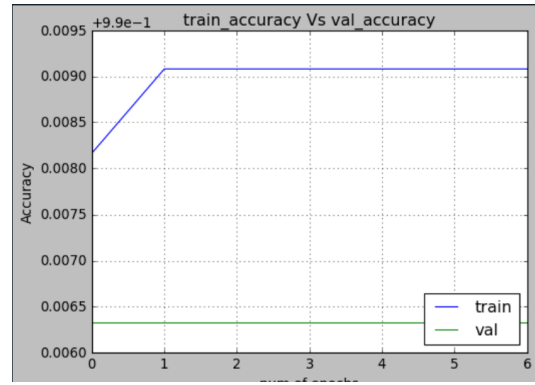


Figure IV.12. Train acc Vs val acc (Neisseria)

➤ Modèle DNN appliqué sur E.coli

On souligne ici qu'on a exécuté le modèle sous les conditions suivantes : *Epochs = 10*, *Batch_size = 32*

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 24)	96
batch_normalization (Batch Normalization)	(None, 24)	96
dropout (Dropout)	(None, 24)	0
dense_1 (Dense)	(None, 18)	450
batch_normalization_1 (Batch Normalization)	(None, 18)	72
dropout_1 (Dropout)	(None, 18)	0
dense_2 (Dense)	(None, 12)	228
batch_normalization_2 (Batch Normalization)	(None, 12)	48
dropout_2 (Dropout)	(None, 12)	0
dense_3 (Dense)	(None, 11)	143
=====		
Total params: 1,133		
Trainable params: 1,025		
Non-trainable params: 108		

Figure IV.13. Architecture du modèle DNN appliqué sur E.coli

➤ Résultats obtenus

Après application du modèle DNN sur la bactérie E.Coli, les résultats obtenus n'étaient pas parfaits mais promoteurs et peut être améliorés au cours du temps si on aura les données suffisantes.

Tableau IV.4. Résultat obtenu pour la bactérie E.coli

Train_loss	1.3988
Train_accuracy	42.64%
Test_loss	1.3143
Test_accuracy	45.55%

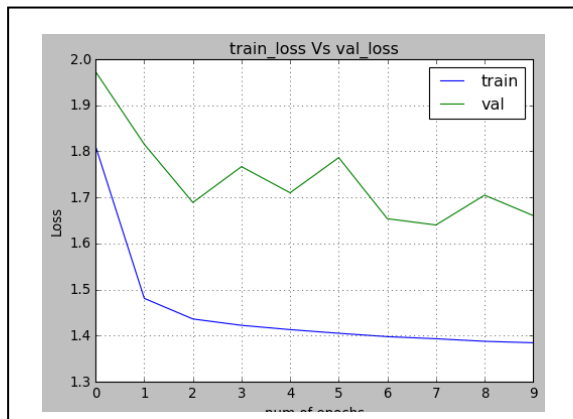


Figure IV.14. Train loss Vs val loss (E.coli)

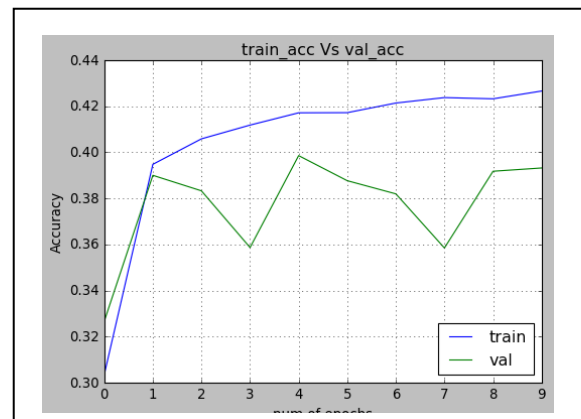


Figure IV.15. Train acc Vs val acc (E.coli)

5. Discussion

5.1. Les résultats obtenus en appliquant les algorithmes ML

L'étude fournit des résultats encourageants, montrant la précision moyenne élevée de différents algorithmes d'apprentissage automatique pour *Neisseria gonorrhoeae* (KNN, régression logistique, arbre de décision CART et forêt aléatoire) sur un ensemble de données de taille moyenne. Ces scores de précision, qui varient entre 99,3 % et 99,9 %, indiquent l'efficacité des algorithmes à faire des prédictions pour cet ensemble spécifique de données. L'algorithme Random Forest a obtenu la plus grande précision prédictive en raison de la qualité de l'ensemble de données (Unitigs qui ont contribué à la production d'assemblages génomiques précis) d'une part, et la combinaison de plusieurs arbres de décision à la fois d'une autre part. Cela peut être interprétée par le fait que les données ont été bien traitées et nettoyées. En outre, les caractéristiques et les variables utilisées représentent des données numériques, ce qui a facilité le processus de prévision

car elles reflètent les performances de chaque algorithme en matière de prévision sur l'ensemble de données.

De même, la précision moyenne des quatre (04) algorithmes d'apprentissage automatique appliqués sur un ensemble de données lié à la bactérie *Escherichia coli* sont plus ou moins acceptables. KNN a atteint une précision de 53 %, ce qui indique qu'il n'est peut-être pas en mesure de capturer efficacement les modèles sous-jacents de l'ensemble de données, c'est-à-dire qu'il a mal prédit la résistance. La régression logistique a atteint une précision de 39 %, ce qui indique que l'ensemble de données ne correspond pas à l'algorithme car il prend en charge les variables numériques et n'a pas pu résoudre les problèmes de régression ou il ne peut pas trouver une relation linéaire entre elles, ou aussi les caractéristiques peuvent ne pas être suffisamment utiles pour prédire la résistance avec une précision élevée. L'arbre de décision CART a bien fonctionné, avec une précision de 86 %, ce qui indique son efficacité dans la classification binaire des modèles. La grande précision de l'algorithme Random Forest indique qu'il exploite efficacement la connaissance collective de plusieurs arbres de décision, C.à.d., il prend en charge un vaste ensemble de données pour la classification d'où, il permet de faire des prédictions précises sur l'ensemble de données utilisées.

En résumé, l'algorithme RF a obtenu les meilleurs résultats, suivi par l'arbre de décision CART. En revanche, les algorithmes LR et KNN ont montré une précision faible, indiquant qu'ils ne sont peut-être pas les meilleurs algorithmes adaptés à ce jeu de données. Il est important de noter que le choix de l'algorithme peut dépendre de plusieurs facteurs, tels que la nature des données, le problème spécifique à résoudre et les ressources disponibles pour l'apprentissage.

5.2. Les résultats obtenus en construisant des modèles DL

Les résultats obtenus par l'application des modèles DL indiquent clairement qu'il existe une variation significative de la performance des réseaux neuronaux profonds (DNN) entre les ensembles de données *Neisseria gonorrhoeae* et *E. coli*. Pour *Neisseria gonorrhoeae*, le réseau neuronal profond a atteint des précisions d'apprentissage et de test très élevées qui varient entre 99,91 % et 99,85 %, respectivement, avec de faibles valeurs de perte (loss function). En revanche, pour *E. coli*, le DNN a obtenu une précision d'apprentissage et de test beaucoup plus faibles (42,64 % et 45,55 %, respectivement), accompagnées de valeurs de perte beaucoup plus élevées.

La principale raison de la plus grande précision obtenue dans la prédiction des résultats pour *Neisseria gonorrhoeae* par rapport à *E. coli* réside dans la nature du dataset utilisé et la complexité des modèles sous-jacents. Les données relatives à *Neisseria gonorrhoeae* présentaient des schémas

plus clairs et plus facilement reconnaissables, et le DNN a pu les apprendre efficacement, ce qui s'est traduit par une plus grande précision. En revanche, les données relatives à E. coli étaient plus complexes, contenant du bruit ou des caractéristiques moins distinctes qui posent des problèmes pour des prédictions précises.

6. Interfaces de l'application développée

Notre application est composée de quatre pages : page de connexion, page principale, page d'analyse des données et page de prédiction.

Page de connexion : pour se connecter au site Antibiotic prediction

Page principale : elle présente les outils utilisés : Section des prédictions, Exploration des données, Performance du modèle, Prédictions enregistrées, Paramètres, Aide et Documentation.

Page d'analyse des données : formulaire de saisie permettant aux utilisateurs d'entrer des informations pertinentes pour la prédiction, Tels que le nom de la bactérie, sa fonction et certaines caractéristiques de chaque bactérie.

Page de prédiction : pour prédire la résistance aux antibiotiques.

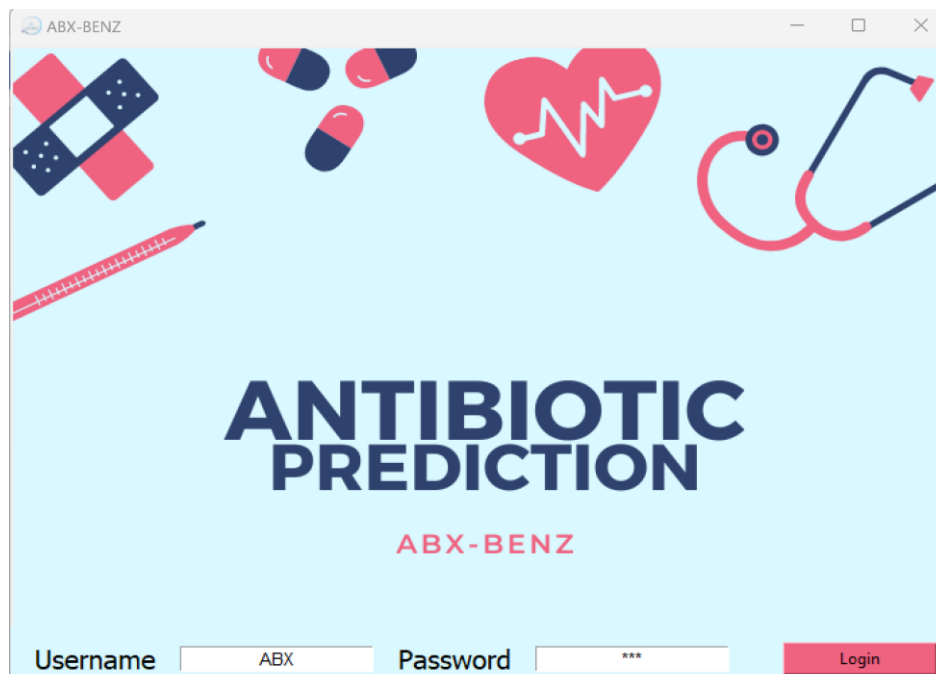


Figure IV.16. Page de connexion

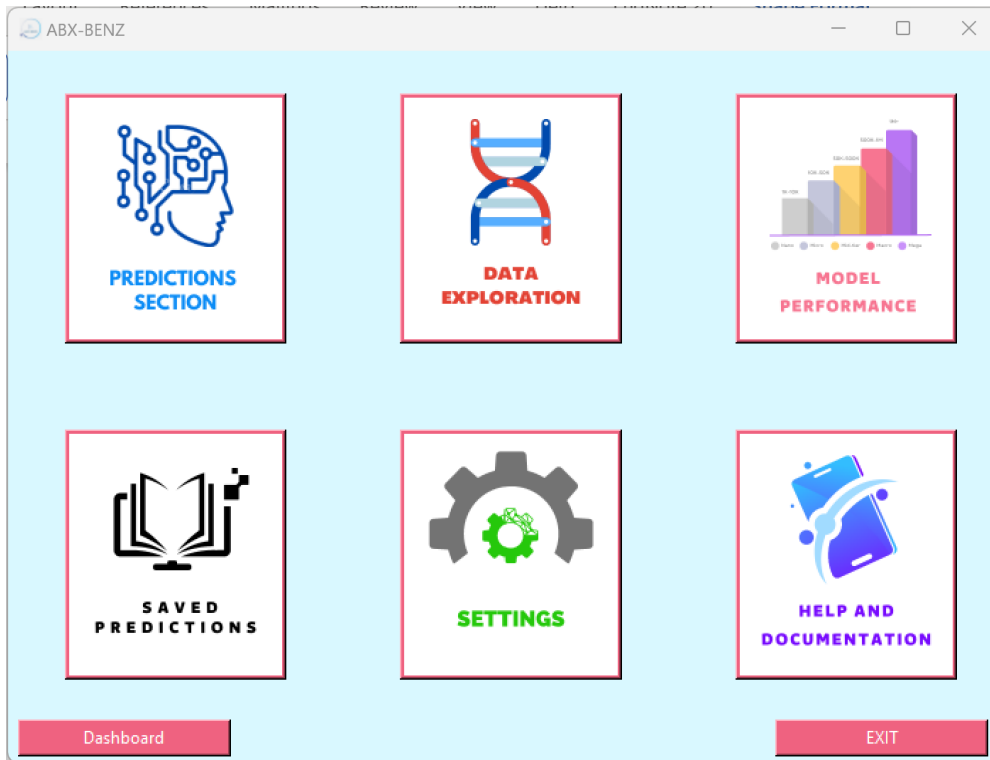


Figure IV.17. Page principale

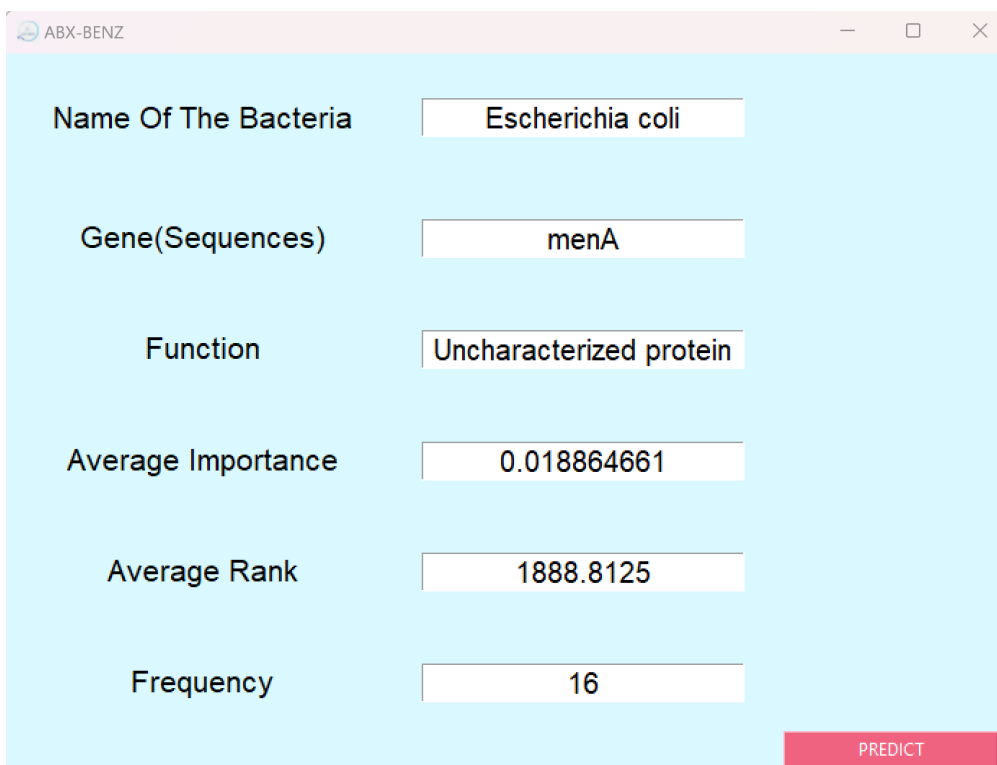


Figure IV.18. Page d'analyse des données

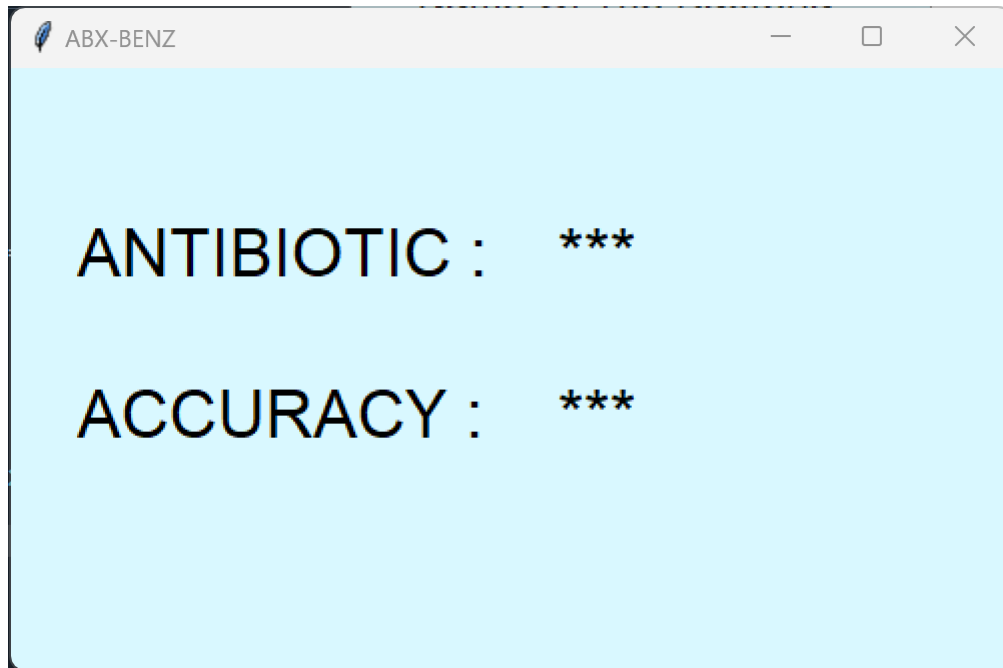


Figure IV.19. Page de prédiction

7. Conclusion

Dans notre travail, nous avons essayé de relier les modèle développés et l'interface de l'application. Nous avons déterminé que les modèles proposés étaient faciles à utiliser et efficaces pour prédire la résistance aux antibiotiques en particulier pour le cas de la bactérie N.G.

CONCLUSION GÉNÉRALE

Conclusion Générale

En conclusion, l'utilisation des algorithmes d'apprentissage automatique et d'apprentissage profond pour prédire la résistance des bactéries aux antibiotiques s'est révélée très prometteuse pour relever l'un des défis les plus urgents dans le domaine des soins de santé. En s'appuyant sur de vastes ensembles de données et des modèles informatiques robustes, l'apprentissage automatique et l'apprentissage profond ont fourni des informations précieuses sur le potentiel de résistance aux antibiotiques, permettant une prise de décision claire et une intervention proactive. Cette combinaison de la biologie et de l'intelligence artificielle nous a donné plusieurs avantages, tels que l'augmentation de la précision en analysant de grandes quantités de données et en identifiant des modèles qui peuvent ne pas être apparents pour les humains. L'apprentissage automatique peut contribuer à une prédiction rapide, ce qui permet de limiter la propagation des infections résistantes.

Cependant, nous avons rencontré quelques difficultés et défis au long de cette étude, car le succès des algorithmes ML et DL dépend fortement de la disponibilité des données pertinentes et de haute qualité, ainsi que du développement de mécanismes de résistance, qui est un phénomène complexe et dynamique en constante évolution.

Malgré ces défis, l'application de Machine Learning et de Deep Learning dans le domaine de prédiction de la résistance aux antibiotiques offre un énorme potentiel pour révolutionner les pratiques de soins dans le secteur de la santé. En tenant compte des avantages et des défis, les chercheurs et les professionnels de la santé peuvent travailler ensemble pour surmonter les obstacles et maximiser l'impact de cette technologie dans la lutte contre la résistance aux antibiotiques.

Comme perspective, nous proposons aux futurs chercheurs (biologistes, médecins, informaticiens, ...) de travailler ensemble dans le but d'améliorer les résultats prédictifs, et même étendre l'étude pour prendre en considération d'autres bactéries et pourquoi pas des virus très connus qui menacent la vie des personnes comme corona-virus et ces différentes variantes.

RÉFÉRENCES

BIBLIOGRAPHIQUES

Références Bibliographiques

- Achi, L. (2018). <Etude phénotypique des souches Escherichia Coli multi-résistantes.pdf>.
- Afssa. (2006). Usages vétérinaires des antibiotiques, résistance bactérienne et conséquences pour la santé humaine.
- Agnihotry, S., Pathak, R. K., Singh, D. B., Tiwari, A., & Hussain, I. (2022). Protein structure prediction. In *Bioinformatics* (pp. 177-188). Elsevier.
- Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., & Zhang, L. (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, 6, 1-15.
- Avril, F., & Rabel, C. (1980). *Manuscrits enluminés d'origine italienne*. Bibliothèque nationale.
- Baxevanis, A. D., & Bateman, A. (2015). The importance of biological databases in biological discovery. *Current protocols in bioinformatics*, 50(1), 1.1. 1-1.1. 8.
- Benabdallah, H. (2016). <Etude phénotypique de quelques souches d'Escherichia coli productrices des carbapénèmes>.
- Bouhraoua, N., AMRAOUI, A., IRKI, S., & MESSAI Chafik, R. (2021). *Isolement et antibioresistance des souches d'Escherichia coli isolées chez la volaille*
- Carlsson, E. (2019). Machine Learning for Prediction of Antibiotic Resistance.
- Cavallo, J. D., Fabre, R., Jehl, F., Rapp, C., & Garrabé, E. (2004). Bêtalactamines. *EMC - Maladies Infectieuses*, 1(3), 129-202. <https://doi.org/https://doi.org/10.1016/j.emcmi.2004.03.003>
- Chaabane, A., Aouam, K., Boughattas, N., & Chakroun, M. (2009). Allergie aux bêtalactamines: mythe et réalités. *Médecine et maladies infectieuses*, 39(5), 278-287.
- Chaussade, H., Sunder, S., Bernard, L., Coloby, P., Guy, L., Karsenty, G., Bastide, C., & Bruyère, F. (2013). Les médicaments antibiotiques en urologie. *Progrès en urologie*, 23(15), 1327-1341.
- Chien, J.-T. (2019). Chapter 7 - Deep Neural Network. In J.-T. Chien (Ed.), *Source Separation and Machine Learning* (pp. 259-320). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-804566-4.00019-X>
- Cordaux, R., Pichon, S., Hatira, H. B. A., Doublet, V., Grève, P., Marcadé, I., Braquart-Varnier, C., Souty-Grosset, C., Charfi-Cheikhrouha, F., & Bouchon, D. (2012). Widespread Wolbachia infection in terrestrial isopods and other crustaceans. *Zookeys*(176), 123.
- Courvalin, P. (2008). La résistance des bactéries aux antibiotiques : combinaisons de mécanismes biochimiques et génétiques. *Bulletin de l'Académie Vétérinaire de France*, 161(1), 7-12. <https://doi.org/10.4267/2042/47917>
- Davison, H., Guy, C., McGarry, J., Guy, F., Williams, D., Kelly, D., & Trees, A. (2001). Experimental studies on the transmission of Neospora caninum between cattle. *Research in Veterinary Science*, 70(2), 163-168.
- de Lastours, V., Cambau, E., Guillard, T., Marcade, G., Chau, F., & Fantin, B. (2012). Diversity of individual dynamic patterns of emergence of resistance to quinolones in Escherichia coli from the fecal flora of healthy volunteers exposed to ciprofloxacin. *The Journal of infectious diseases*, 206(9), 1399-1406.
- DZIUDA, D. M. (2010). DATA MINING FOR GENOMICS AND PROTEOMICS In *Analysis of Gene and Protein Expression Data*.

- Fares, N., & Djedjai, S. (2018). *Caractérisation phénotypique de la résistance aux β -lactamines à large spectre des bacilles à Gram négatifs isolés à partir des produits pathologiques* [Université laarbi tebessi tebessa].
- Ferrahi, H. (2021). <L'antibiorésistance des souches de Salmonella pullorum gallinarum isolées des élevages de poulet de chair dans la wilaya de Médéa (1).pdf>.
- Forge, A., & Schacht, J. (2000). Aminoglycoside antibiotics. *Audiology and Neurotology*, 5(1), 3-22.
- Gadri, D. S. (2020). <Chap 3. Apprentissage Profond.pdf>.
- GADRI, D. S. (2022). <Ch 2. Modèles d'apprentissage automatique non supervisé.pdf>.
- Gadri, S. (2021). Developing an efficient predictive model based on ml and dl approaches to detect diabetes. *Informatica*, 45(3).
- GADRI, S. (2022). <Chap 1. Modèles d'apprentissage automatique.pdf>.
- Gadri, S., & Adouane, N. E. (2022). Efficient traffic signs recognition based on cnn model for self-driving cars. *Intelligent Computing & Optimization: Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021 (ICO2021)* 3,
- Galimand, M., Lambert, T., & Courvalin, P. (2005). Emergence and dissemination of a new mechanism of resistance to aminoglycosides in Gram-negative bacteria: 16S rRNA methylation. *Weekly releases (1997–2007)*, 10(4), 2626.
- Gholizadeh, S. (2022). Top Popular Python Libraries in Research. *Authorea Preprints*.
- Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology*, 23(1), 40-55.
- Grimont, P. (1987). Taxonomie des Escherichia. *Médecine et maladies infectieuses*, 17, 6-10.
- Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics*, 44(3), 348-361.
- Hassoun, S., Jefferson, F., Shi, X., Stucky, B., Wang, J., & Rosa Jr, E. (2021). Artificial intelligence for biology. *Integrative and Comparative Biology*, 61(6), 2267-2275.
- Huff, G., Huff, W., Balog, J., & Rath, N. (1999). Sex differences in the resistance of turkeys to Escherichia coli challenge after immunosuppression with dexamethasone. *Poultry science*, 78(1), 38-44.
- Jakhar, D., & Kaur, I. (2020). Artificial intelligence, machine learning and deep learning: definitions and differences. *Clinical and experimental dermatology*, 45(1), 131-132.
- Kaur, D. J. a. I. (2019). Artificial intelligence, machine learning and deep learning: definitions and differences. <https://doi.org/10.1111/ced.14029>
- Khellafi, A., Samer, A., Aboura, L., & Boussouf, L. E. (2009). *La résistance bactérienne aux bêta-lactamines* [Université de Jijel].
- LADJAL, S., & BAGHDADI, A. *Mécanismes de résistance aux β -lactamines chez Escherichia coli au niveau des hôpitaux algériens* UNIVERSITE MOHAMED BOUDIAF-M'SILA].
- Lalis, A. (2008). *Variabilité phénotypique et génétique des Mastomys (Rodentia, Muridae) de Guinée: évolution, environnement et infection virale* Paris 6].
- Lavigne, C. (2007). À qui appartient l'objet de recherche? Penser l'implication du chercheur dans son objet: le handicap (surdit ). *Nouvelle revue de psychosociologie*(2), 23-39.

- Lavollay, M., Arthur, M., Fourgeaud, M., Dubost, L., Marie, A., Veziris, N., Blanot, D., Gutmann, L., & Mainardi, J.-L. (2008). The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L, D-transpeptidation. *Journal of bacteriology*, *190*(12), 4360-4366.
- Limeres Posse, J., Diz Dios, P., & Scully, C. (2017). Chapter 3 - Systemic Bacteria Transmissible by Kissing. In J. Limeres Posse, P. Diz Dios, & C. Scully (Eds.), *Saliva Protection and Transmissible Diseases* (pp. 29-51). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-813681-2.00003-2>
- Mainil, J. (2013). *Escherichia coli* virulence factors. *Veterinary immunology and immunopathology*, *152*(1-2), 2-12.
- Mammeri, H., & Amiens, C. (2013). Mode d'action des antibiotiques. *Service de bactériologie, CHU Amiens. P, 2*.
- Mansour, W., Bouallegue, O., Dahmen, S., & Boujaafar, N. (2008). Caractérisation des mécanismes enzymatiques de résistance aux β -lactamines chez des souches de *Acinetobacter baumannii* isolées à l'hôpital universitaire Sahloul, Sousse en Tunisie (2005). *Pathologie Biologie*, *56*(3), 116-120. <https://doi.org/https://doi.org/10.1016/j.patbio.2007.10.004>
- Maza, L. d. I. (2020). *Neisseria*. In *Color Atlas of Medical Bacteriology* (pp. 184-190). <https://doi.org/10.1128/9781683671077.ch21>
- McSheffrey, G. G., & Gray-Owen, S. D. (2015). *Neisseria gonorrhoeae*. In *Molecular medical microbiology* (pp. 1471-1485). Elsevier.
- Moghrani, H., Aouissi, M., & Achek, R. (2018). *Etude de la résistance aux antibiotiques des entérobactéries productrices de β -lactamases à spectre étendu (BLSE) isolées à l'hôpital de Médéa*
- Möller, D. P. (2023). Machine Learning and Deep Learning. In *Guide to Cybersecurity in Digital Transformation: Trends, Methods, Technologies, Applications and Best Practices* (pp. 347-384). Springer.
- Moradigaravand, D., Palm, M., Farewell, A., Mustonen, V., Warringer, J., & Parts, L. (2018). Prediction of antibiotic resistance in *Escherichia coli* from large-scale pan-genome data. *PLoS computational biology*, *14*(12), e1006258.
- Muylaert, A., & Mainil, J. (2012). Bacterial antimicrobial resistances: the mechanisms and their contagiousness. *Annales de Médecine Vétérinaire*,
- Nabipour, M., Nayyeri, P., Jabani, H., S, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE access*, *8*, 150199-150212. <https://doi.org/10.1109/access.2020.3015966>
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons.B*, *4*, 51-62. <https://doi.org/10.20544/horizons.B.04.1.17.P05>
- Ndagi, U., Falaki, A. A., Abdullahi, M., Lawal, M. M., & Soliman, M. E. (2020). Antibiotic resistance: bioinformatics-based understanding as a functional strategy for drug design. *RSC Adv*, *10*(31), 18451-18468. <https://doi.org/10.1039/d0ra01484b>
- Ng, L.-K., & Martin, I. E. (2005). The laboratory diagnosis of *Neisseria gonorrhoeae*. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *16*(1), 15-25.
- Ng, L. K., & Martin, I. E. (2005). The laboratory diagnosis of *Neisseria gonorrhoeae*. *Can J Infect Dis Med Microbiol*, *16*(1), 15-25. <https://doi.org/10.1155/2005/323082>

- Pantel, A., Souzy, D., Sotto, A., & Lavigne, J.-P. (2015). Evaluation of two phenotypic screening tests for carbapenemase-producing Enterobacteriaceae. *Journal of Clinical Microbiology*, 53(10), 3359-3362.
- Rabaan, A. A., Alhumaid, S., Mutair, A. A., Garout, M., Abulhamayel, Y., Halwani, M. A., Alestad, J. H., Bshabshe, A. A., Sulaiman, T., & AlFonaisan, M. K. (2022). Application of Artificial Intelligence in Combating High Antimicrobial Resistance Rates. *Antibiotics*, 11(6), 784.
- Raleigh, E. A., & Low, K. B. (2013). Conjugation. In S. Maloy & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (pp. 144-151). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-374984-0.00321-1>
- Ray, S. (2019). A quick review of machine learning algorithms. 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon),
- Róžańska, H., & Osek, J. (2013). Stability of antibiotics in milk samples during storage. *Journal of Veterinary Research*, 57(3), 347-349.
- Saabith, S., Vinothraj, T., & Fareez, M. (2021). A review on Python libraries and Ides for Data Science. *Int. J. Res. Eng. Sci.*, 9(11), 36-53.
- Sapundzhi, F., & Dzimbova, T. (2022). The importance of biological databases in modeling of structure-activity relationship. *interpretation*, 7, 11.
- Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. 2018 Fourth international conference on computing communication control and automation (ICCUBEA),
- Siedner, M. J., Pandori, M., Castro, L., Barry, P., Whittington, W. L., Liska, S., & Klausner, J. D. (2007). Real-time PCR assay for detection of quinolone-resistant *Neisseria gonorrhoeae* in urine samples. *Journal of Clinical Microbiology*, 45(4), 1250-1254.
- Stratton, I. M., Adler, A. I., Neil, H. A. W., Matthews, D. R., Manley, S. E., Cull, C. A., Hadden, D., Turner, R. C., & Holman, R. R. (2000). Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *Bmj*, 321(7258), 405-412.
- Tapsall, J. W. (2005). Antibiotic resistance in *Neisseria gonorrhoeae*. *Clinical Infectious Diseases*, 41(Supplement_4), S263-S268.
- Thuener, J. E., & Clouse, A. L. (2013). Gonorrhea. In *Sexually Transmitted Diseases: A Practical Guide for Primary Care* (pp. 61-69). Springer.
- Tønjum, T., & van Putten, J. (2017). 179 - *Neisseria*. In J. Cohen, W. G. Powderly, & S. M. Opal (Eds.), *Infectious Diseases (Fourth Edition)* (pp. 1553-1564.e1551). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-7020-6285-8.00179-9>
- Torrisi, M., Pollastri, G., & Le, Q. (2020). Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal*, 18, 1301-1310.
- Vakulenko, S. B., Donabedian, S. M., Voskresenskiy, A. M., Zervos, M. J., Lerner, S. A., & Chow, J. W. (2003). Multiplex PCR for detection of aminoglycoside resistance genes in enterococci. *Antimicrobial agents and chemotherapy*, 47(4), 1423-1426.
- Walsh, I., Fishman, D., Garcia-Gasulla, D., Titma, T., Pollastri, G., Harrow, J., Psomopoulos, F. E., & Tosatto, S. C. (2021). DOME: recommendations for supervised machine learning validation in biology. *Nature methods*, 18(10), 1122-1127.
- Wolff, M., Joly-Guillou, M.-L., & Pajot, O. (2009). Les carbapénèmes. *Réanimation*, 18, S199-S208.

Yoon, I. J., Joo, H. S., Christianson, W. T., Kim, H. S., Collins, J. E., Carlson, J. H., & Dee, S. A. (1992). Isolation of a cytopathic virus from weak pigs on farms with a history of swine infertility and respiratory syndrome. *Journal of Veterinary Diagnostic Investigation*, 4(2), 139-143.

ANNEXES

Annexes

أصبحت مقاومة المضادات الحيوية مصدر قلق كبير للصحة العالمية بسبب عدم فعالية هذه الأخيرة، وانتشار البكتيريا المقاومة، وخيارات العلاج المحدودة، مما استلزم تطوير أساليب مبتكرة للتنبؤ بهذه الظاهرة ومكافحتها. في دراستنا، تم استخدام خوارزميات التعلم الآلي والتعلم العميق لتطوير نماذج تنبؤية تعتمد على ميزات مختلفة مستمدة من البيانات الجينومية والظاهرية والسريية. في دراستنا قمنا بتطوير نموذج يمكنه التنبؤ بمقاومة المضادات الحيوية في بكتيريا النيسيرية البنية والإشريكية القولونية، مما يتيح تحديداً دقيقاً وسريعاً للسلاسل البكتيرية المقاومة. حيث تم الحصول على دقة 99% بالنسبة للنيسيرية البنية و 89% بالنسبة للإشريكية القولونية. لقد أثبت هذا النموذج فعاليته العالية ودقته في التنبؤ بمقاومة المضادات الحيوية.

الكلمات المفتاحية: البكتيريا المقاومة، المضادات الحيوية، النيسيرية البنية، الإشريكية القولونية، التعلم الآلي، التعلم العميق، الشبكات العصبية الاصطناعية

Abstract

Antibiotic resistance has become a critical global health concern due to the ineffectiveness of antibiotics, the spread of resistant bacteria, and limited treatment options, necessitating the development of innovative approaches to predict and combat this phenomenon. In our study, machine learning and deep learning algorithms have been employed to develop predictive models based on various features derived from genomic, phenotypic, and clinical data. We developed a model that predicts antibiotic resistance in *Neisseria gonorrhoeae* and *Escherichia coli* bacteria, enabling accurate and fast identification of resistant bacterial strains. Accuracy was 99% for *Neisseria.g* and 89% for *E.coli*. This model has proven to be highly effective and accurate in predicting antibiotic resistance.

Keywords: Bacteria resistance, Antibiotics, *Neisseria gonorrhoeae*, *Escherichia coli* bacteria, Machine learning, Deep learning, Artificial Neural networks.

Résumé

La résistance aux antibiotiques est devenue un problème de santé mondial critique en raison de l'inefficacité des antibiotiques, de la propagation de bactéries résistantes et des options de traitement limitées, ce qui nécessite le développement d'approches innovantes pour prédire et combattre ce phénomène. Dans notre étude, des algorithmes d'apprentissage automatique et d'apprentissage profond ont été utilisés pour développer des modèles prédictifs basés sur diverses caractéristiques dérivées de données génomiques, phénotypiques et cliniques. Nous avons développé un modèle qui prédit la résistance aux antibiotiques chez les bactéries *Neisseria gonorrhoeae* et *Escherichia coli*, permettant une identification précise et rapide des souches bactériennes résistantes. Une précision de 99 % a été obtenue pour *Neisseria.g* et de 89 % pour *E.coli*. Ce modèle s'est avéré très efficace et précis dans la prédiction de la résistance aux antibiotiques.

Mots clés : Bactéries résistantes, Antibiotiques, *Neisseria gonorrhoeae*, *Escherichia coli*, Apprentissage automatique ; Apprentissage profond, Réseaux de neurones artificiels.