



الجمهورية الجزائرية الشعبية الديمقراطية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Faculté des Sciences de la Nature et de la Vie

كلية علوم الطبيعة و الحياة

Département de Biologie appliquée

قسم البيولوجيا التطبيقية

Mémoire présenté en vue de l'obtention du diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : biotechnologies

Spécialité : Bio-informatique

Intitulé :

Méthode d'apprentissage profond pour la prédition de stress abiotique causé par la fusariose. Cas d'étude : Le blé

Présenté par : **BAALI Khalil Sid Ali**

LE 25/06/2025

TALBI Abd El Hamid

Jury d'évaluation:

Président : **Dr. BENHAMDI Asma** (MCA - Université Frères Mentouri, Constantine 1).

Encadreur : **Dr. GHERBOUDJ Amira** (MCA - Université Frères Mentouri, Constantine 1).

Examinateur : **Dr. DAAS Mohamed Skander** (MCA - Université Frères Mentouri, Constantine 1).

Année universitaire

2024 -2025

Remerciement

Nous commençons par exprimer notre profonde gratitude envers **Allah** Tout-Puissant, qui nous a guidés tout au long de ce parcours académique et nous a accordé la force et la motivation nécessaires pour mener à bien ce travail.

Nos sincères remerciements vont également à l'ensemble des enseignants qui nous ont accompagnés durant notre cursus universitaire. Nous tenons tout particulièrement à remercier notre encadrante, **Mme Dr.Amira GHERBOUDJ**, pour son précieux soutien, sa patience, sa rigueur scientifique et sa disponibilité tout au long de ce projet.

Nous exprimons notre profonde reconnaissance envers **Mlle ABDELAZIZ Aya** pour son aide constante et sa grande disponibilité. Son engagement, ses réponses à nos interrogations et ses conseils méthodologiques ont été d'une grande utilité pour l'avancement de nos recherches. Sa passion pour la bio-informatique a été une véritable source d'inspiration pour notre équipe.

Nous souhaitons également témoigner notre gratitude aux membres du jury, notamment son Président, **Dr. BENHAMDI Asma**, ainsi qu'à l'examinateur, **Dr. DAAS Mohamed Skander**, pour l'honneur qu'ils nous font en acceptant d'évaluer ce travail.

Une pensée reconnaissante va à nos familles et à nos amis, dont le soutien indéfectible et les encouragements constants ont été d'un grand réconfort dans l'accomplissement de ce mémoire.

Enfin, nous adressons nos remerciements à toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce projet.

TALBI ET BAALI.

Dédicace

Je souhaite exprimer ma sincère gratitude à toutes les personnes qui ont accompagné et soutenu mon parcours académique et personnel.

À mes chers parents,

Votre amour inconditionnel, votre bienveillance et vos sacrifices ont été les fondations de mon éducation et de mon épanouissement. Votre soutien indéfectible, vos encouragements et votre foi en mes capacités m'ont donné la force de persévérer, même dans les moments les plus exigeants.

À mes frères et sœur, Manar, Sid Ahmed

Votre présence chaleureuse et votre soutien constant ont été pour moi un précieux réconfort.

Vos mots d'encouragement et votre affection ont illuminé mon chemin, faisant de chaque étape un moment partagé.

À mon collègue, TALBI Abdelhamid

Votre sagesse, votre générosité et votre sérieux qu'on a pu à réussir ce projet. Votre amitié et vos conseils avisés ont été une source d'inspiration et de motivation.

À mes amis, fidèles compagnons de route,

Merci d'avoir partagé avec moi les joies comme les défis de cette aventure. Votre amitié, votre écoute et votre solidarité ont transformé les obstacles en opportunités, et les moments de doute en élan renouvelé.

À mon encadrante Dr. Amira GHERBOUDJ et à mes professeurs,

Je tiens à vous témoigner toute ma reconnaissance pour vos enseignements, votre expertise et vos conseils éclairés. Votre passion pour le savoir a éveillé la mienne, et votre guidance a été une lumière tout au long de ce parcours universitaire.

Enfin, je dédie ce travail à toutes ces personnes extraordinaires qui, chacune à leur manière, ont contribué à son aboutissement. Que ces mots traduisent, même modestement, l'immense gratitude que je ressens à votre égard.

BAALI SID ALI.

Dédicace

Je dédie ce modeste travail à :

Mes parents :

Ma mère, honorable, aimable, Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi, A **mes grands-mères** et **mon grand-père**. Et je remercie toute **ma famille** particulièrement, **Mes tantes** et **Mes oncles** du côté maternel, pour leur amour, leur soutien et leurs prières qui m'ont accompagné tout au long de mon parcours.

Mon père, Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de tes sacrifices que tu as consentis pour ma formation.

Un profond respect et un remerciement particulier à celle qui a partagé avec moi les moments les plus beaux et les plus dures de ces 5 années de fac.

Un remerciement particulier et sincère pour **Mon collègue** ‘BAALI SID ALI’ En témoignage de l'amitié qui nous uni et des souvenirs de tous les moments que nous avons passés ensemble.

À **mes amis** les plus proches, (vous savez très bien qui vous êtes),

Ceux qui ont transformé les longues journées en souvenirs inoubliables, ceux qui ont toujours une blague nulle prête au pire moment. **Merci** pour les éclats de rire, les “plans de dernière minute”, et vos phrases culte que je ne pourrai jamais oublier. Vous êtes la preuve vivante que l'amitié rend tout plus léger, même les deadlines.

TALBI ABD ELHAMID

RÉSUMÉ

Cette étude propose une approche innovante combinant l'intelligence artificielle et la bio-informatique pour prédire le stress abiotique causé par la fusariose chez le blé. Nous avons développé deux méthodes distinctes. Une première approche basée sur l'analyse d'images des plantes, et une seconde exploitant des séquences nucléiques, toutes deux utilisant des réseaux de neurones convolutifs (CNN). Les résultats démontrent l'efficacité de ces techniques, avec une précision de 92.4% pour le modèle d'analyse d'images et 99% pour celui utilisant les données nucléiques. Ces performances élevées soulignent le rôle crucial que peuvent jouer l'IA avancée et les outils bio-informatiques dans l'agriculture moderne, notamment pour le diagnostic précoce des pathologies végétales. Nos travaux ouvrent des perspectives prometteuses pour l'intégration de ces technologies dans la gestion optimisée des cultures, et posent les bases pour des recherches futures visant à développer des systèmes d'aide à la décision encore plus performants pour les agriculteurs et toutes les catégories de la société intéressés par ce domaine de la prédiction.

Mots clés : Intelligence Artificiel, prédiction, méthode Convolutional Neural Network (CNN), images, bio-informatique, la fusariose, blé, séquences nucléiques, agriculture moderne.

ABSTRACT

This study proposes an innovative approach combining artificial intelligence and bioinformatics to predict abiotic stress caused by fusarium in wheat. We have developed two distinct methods. The first is based on the analysis of plant images and the second exploits nuclear sequences, both using convolutional neural networks (**CNN**). The results show the effectiveness of these techniques, with an accuracy of **92.4%** for the image analysis model and **99%** for the model using nuclear data. These high performances underline the crucial role that advanced artificial intelligence and bioinformatics tools can play in modern agriculture, particularly in the early diagnosis of plant diseases. Our work opens up promising avenues for integrating these technologies into optimal crop management and lays the foundations for future research aimed at developing more effective decision support systems for farmers.

Key words : artificial intelligence , prediction , method Convolutional Neural Network (CNN) , pictures , bioinformatics , fusarium , wheat , nucleic sequences , modern agriculture.

ملخص

تهدف هذه الدراسة إلى تقديم نهج مبتكر يجمع بين تقنيات الذكاء الاصطناعي والمعلوماتية الحيوية للتنبؤ بالإجهاد الاحيوي الناتج عن الإصابة بآفة رأس الفيوزاريوم في محصول القمح. اعتمدنا في هذا العمل على منهجية ثنائية المحاور، حيث تم تطوير نموذجين تحليليين متكاملين. يعتمد النموذج الأول على معالجة الصور النباتية باستخدام خوارزميات التعلم العميق، بينما يرتكز النموذج الثاني على تحليل التسلسلات النووية باستخدام الشبكات العصبية الالتفافية، أسفرت التجارب عن نتائج دقيقة ومشجعة، حيث حقق نموذج تحليل الصور دقة تنبؤية بلغت 92.4%， في حين تحقق نموذج تحليل البيانات النووية بدقة ملحوظة وصلت إلى 99% . تؤكد هذه المؤشرات الأدائية العالية على الجدوى العملية لدمج تقنيات الذكاء الاصطناعي في المجال الزراعي، لا سيما في مجال التشخيص المبكر للأمراض النباتية، تمثل هذه النتائج إضافة علمية مهمة في مجال الزراعة الذكية، حيث تفتح آفاقاً جديدة لتحسين إدارة المحاصيل ورفع كفاءتها الإنتاجية. كما تضع الأسس لمبادرات بحثية مستقبلية تهدف إلى تطوير أنظمة دعم القرار الزراعي، مما يسهم في تعزيز الأمن الغذائي وترشيد الموارد الزراعية.

الكلمات المفتاحية : الذكاء الاصطناعي ، التنبؤ ، القمح ، طريقة الشبكات العصبية الالتفافية ، الصور ، البيانات النووية ، الامراض النباتية ، رأس الفوزاريوم ، الكفاءة الإنتاجية ، النموذج .

Liste des figures

Figure 1 : L'origine géographique des blés cultivés. Adapté de Dubcovsky and Dvorak (2007).....	4
Figure 2 : Blé tendre et blé dur Source	5
Figure 3 : Hiérarchie taxonomique des plantes : exemple du triticum.....	6
Figure 4 : exemple d'un effet de la rouille sur le blé.....	9
Figure 5 : Exemple d'un effet de l'Oïdium.(Bayer, 2021)	9
Figure 6 : Exemple d'un effet du Septoriose.(Bayer, 2021)	10
Figure 7 : Exemple d'un effet du Xanthomonas translucens.(Ephytia, 2018)....	11
Figure 8 : Exemple d'un effet du Mosaïques du blé.(Ephytia, 2017)	12
Figure 9 : Fusariose de l'épi du blé ; épillets décolorés	13
Figure 10 : Fusariose de l'épi du blé.....	13
Figure 11 : Fusariose de l'épi du blé ; la teinte orangée dénote la présence du champignon pathogène.....	14
Figure 12 : Zoom sur la Fusariose du blé.....	15
Figure 13 : des zones comportant des épis décolorés et vides.(DZ, 2021)	16
Figure 14 : schéma qui présente les principes de base d'IA	19
Figure15 : Nombre approximatif d'articles publiés sur l'apprentissage profond et l'apprentissage profond dans la bio-informatique par année.	22
Figure 16 : La similitude entre les neurones et les réseaux neuronaux	22
Figure 17 : un schéma représente le fonctionnement du deep Learning	23
Figure 18 : Un ensemble schématique du flux de travail de l'apprentissage profond.(Yu, Qi and Ding, 2022)	24
Figure 19 : les bibliothèques nécessaires	31
Figure 20 : structure du fichier csv	31
Figure 21 : modèle séquentiel.....	31

Figure 22 : la fonction principale d’entraînement du modèle	32
Figure 23 : Script de l’entraînement	32
Figure 24 : Bibliothèques et modules importés.....	33
Figure 25 : Code de détection des plantes.....	34
Figure 26 : Structure du fichier.....	34
Figure 27 : Le chemin du dossier	35
Figure 28 : Modèle de construction	35
Figure 29 : Séquentiel du modèle	36
Figure 30 : Code d’entraînement	36
Figure 31 : Le script de l’entraînement	37
Figure 32 : Structure HTML.....	37
Figure 33 : Partie d’une structure CSS	38
Figure 34 : Partie d’un code JAVASCRIPT	38
Figure 35 : L’interface en cas d’image	39
Figure 36 : L’interface en cas de la séquence	39
Figure 37 : Matrice de confusion	40
Figure 38 : Résultat de la prédiction cas de la séquence	41
Figure 39 : Diagramme de précision (Training Accuracy et Validation Accuracy).....	42
Figure 40 : Courbe de perte (Training Loss et Validation Loss)	42
Figure 41 : Résultat de la prédiction d’image	43
Figure 42 : Résultat d’interface qui distingue entre une plante et un objet.....	44

Liste des tables

Tableau 1 : Composition chimiques du blé	6
Tableau 2 : les données des images utilisées.....	27
Tableau 3 : données des séquences utilisées	28
Tableau 4 : Les caractéristiques de l'ordinateur	28
Tableau 5 : Valeurs obtenues du modèle séquence	41
Tableau 6 : les valeurs obtenues du modèle d'image	43
Tableau 7 : Comparaison notre travaille avec une autre méthode.....	45

Liste des abréviations

- **ADN** : Acide désoxyribonucléique
- **DL**: Deep Learning
- **CNN**: Réseaux neurones convolutions
- **ITS**: Internal Transcribed Spacer
- **CSV**: Comma separated values
- **IA**: Intelligence artificielle
- **RF**: Random Forest
- **ML**: Machine Learning
- **HTML**: HyperText Markup Language
- **CSS**: Cascading Style Sheets
- **FASTA**: Fast All Sequences Technique.
- **FN** : Faux négatifs
- **JS** : JavaScript.
- **RAM** : Random Access Memory.
- RNN** : Réseaux de neurones.
- **ROM**: Read-Only Memory.
- **VP**: Vrais positifs

TABLE DES MATIÈRES

Remerciement	1
RÉSUMÉ	2
Liste des figures.....	1
Liste des tables	3
Liste des abréviations	1
TABLE DES MATIÈRES	1
INTRODUCTION GÉNÉRALE.....	1
Chapitre 01 : Un regard approfondi sur le blé	1
1. INTRODUCTION	3
2. Origine et histoire de la culture du blé.....	3
3. Classification taxonomique du blé	5
4. Composition nutritionnelle moyenne du grain de blé	6
5. L'importance génétique du blé	7
6. Les maladies du blé	8
6.1. Rouilles (<i>Puccinia</i> spp.) :.....	8
6.2 Oïdium (<i>Blumeria graminis</i> f. sp. <i>Triticici</i>) :	9
6.3 Septoriose (<i>Zymoseptoria tritici</i>) :	10
6.4. <i>Xanthomonas translucens</i> (Tache bactérienne) :	11
6.5. Mosaïque du blé (Wheat Streak Mosaic Virus - WSMV) :	12
6.6. La fusariose chez le blé	12
6.6.1. État général.....	12
6.6.2. Symptômes	13
6.6.3 Développement et infection de la maladie	14
6.6.4 Les types de la fusariose qui affecte le blé en Algérie	15
Chapitre 2 : Techniques de l'intelligence artificielle.....	18
1. Introduction	18
2. L'histoire de l'intelligence artificielle.....	18
3. Les techniques d'intelligence artificielle.....	19
3.1. Apprentissage automatique (Machine Learning)	19
3.1.1. Étapes de la construction d'un modèle de machine Learning.....	20

3.2. L'apprentissage profond	20
3.2.1. Historique d'apprentissage profond.....	20
3.2.2. L'apprentissage profond dans la bio-informatique.....	21
3.2.3. Fonctionnement du l'apprentissage profond	22
3.2.4. Techniques de l'apprentissage profond	23
3.2.4.1. Réseaux de neurones convolutionnels (CNN)	23
3.2.4.2. Réseaux de neurones récurrents (RNN).....	24
3.2.4.3. Réseaux de neurones génératifs (GAN).....	24
3.2.4.4. Réseaux de neurones siamois.....	24
4. Contributions de l'IA en Bio-informatique	25
4.1. Les apports de l'IA dans l'analyse des séquences.....	25
4.2. Applications de l'IA pour la détection des variants génétiques	26
4.3. IA et annotation génomique	26
Chapitre 03 : Contribution et étude expérimentale	27
1.Introduction	27
2.Matériel.....	27
2.1 Données utilisés	27
2.1.1 Données images.....	27
2.1.2 Données séquences.....	28
2.2. Configuration de la machine.....	28
2.3 Langages de programmation et environnement de travail.....	28
2.4 Bibliothèques	29
3. Méthodes	30
3.1 Modèle de classification par séquence.....	30
3.1.1 Préparation des données	30
3.1.2 Construction du modèle	31
3.1.3 Entraînement et évaluation du modèle	32
3.2. Modèle de classification par image	33
3.2.1 Préparation des données	33
3.2.2 Construction du Modèle	35
3.2.3. Entraînement et évaluation du modèle	36
3.3. L'interface	37

3.3.1 Construction de l'interface	37
3.3.2. Structure de l'interface	38
4. Résultat	40
4.1 Modèle de classification par séquence.....	40
4.1.1 Matrice de confusion.....	40
4.1.2 Valeurs obtenues du modèle de séquence	40
4.1.3 Plateforme du modèle de séquence	41
4.2 Modèle de classification par image	41
4.2.1 Diagramme de précision (Training Accuracy et Validation Accuracy).....	41
4.2.2 Diagramme de perte (Training Loss et Validation Loss)	42
4.2.3. Valeurs obtenues du modèle d'image	43
4.2.4 Plateforme du modèle d'image	43
5. Discussions.....	44
Conclusion et Perspectives.....	46
Références Bibliographiques	47

INTRODUCTION GÉNÉRALE

INTRODUCTION GÉNÉRALE

Le blé (**Triticum spp.**) est l'une des céréales les plus importantes au monde, jouant un rôle central dans l'alimentation humaine et l'économie agricole. Cultivé depuis des millénaires, il est à la base de nombreux produits alimentaires comme le pain, les pâtes et les biscuits. Grâce à sa richesse en glucides, en fibres et en protéines, il constitue une source essentielle d'énergie pour des milliards de personnes. Son importance génétique qui réside dans la diversité de son génome, permettant une amélioration continue des variétés pour accroître les rendements et la résistance aux maladies. (**Shewry, P. R., & Hey, S. J. (2015)**)

Le blé est sensible à plusieurs maladies qui peuvent affecter son rendement et sa qualité. La classification des espèces représente un enjeu scientifique fondamental pour l'identification et l'organisation des entités biologiques. Traditionnellement basée sur des critères morphologiques puis moléculaires, cette approche présente cependant des limites importantes lorsqu'il s'agit de céréales. Les méthodes classiques souffrent en effet de subjectivité, de coûts élevés et de complexité technique, ce qui peut conduire à des résultats peu fiables et à des confusions dans l'identification des espèces. (**Figueroa, M., Hammond-Kosack, K. E., & Solomon, P. S. (2018)**)

Face à ces défis, les nouvelles techniques d'intelligence artificielle offrent des solutions prometteuses. Elles permettent une classification plus objective, plus précise et plus efficace des espèces céréaliers, ouvrant ainsi la voie à des avancées significatives dans ce domaine de recherche. L'intelligence artificielle apporte des solutions innovantes en biologie et en agriculture. Avec des ordinateurs plus performants et davantage de données disponibles, cette technologie aide concrètement les chercheurs et les agriculteurs. Elle améliore la gestion des cultures, permet de prévoir les rendements, identifie les maladies des plantes et classe les espèces végétales. En analysant automatiquement les données, l'intelligence artificielle fournit des informations précieuses pour guider les décisions. Ces avancées technologiques contribuent à développer une agriculture plus efficace et plus durable.

Dans notre mémoire, nous avons développé un modèle d'intelligence artificielle basé sur la classification en utilisant l'algorithme réseaux neurones convolutions (**CNN**) que nous avons appliqué sur les séquences **ITS (Internal Transcribed Spacer)** et des photos permet aux utilisateurs de prédire les maladies du blé. En intégrant les techniques d'intelligence artificielle, des langages de programmation ainsi que des techniques d'analyse de données basées sur l'expertise agricole, le modèle vise à relever les défis actuels dans la détection du stress abiotique chez le blé. Notre objectif est d'aider les agriculteurs, les acteurs de la filière

blé et les organisations agricoles à détecter précocelement les maladies du blé, comme la fusariose. Nous nous concentrerons principalement sur l'entraînement et l'évaluation de ce modèle afin d'améliorer la précision de la classification des maladies du blé.

Notre manuscrit fournit une présentation détaillée du thème abordé. Il est organisé en trois chapitres :

Chapitre 1 : Le premier chapitre propose une analyse approfondie du blé, en commençant par une présentation générale de cette céréale fondamentale. Nous retracions l'origine et l'histoire de sa culture, mettant en lumière son évolution et son rôle clé dans le développement des sociétés agricoles. La classification taxonomique du blé est explicitée, précisant sa place dans le règne végétal. Une analyse détaillée de la composition du grain de blé est ensuite et le présentée, avec un focus sur ses valeurs nutritionnelles et ses utilisations. Les besoins minéraux du blé sont examinés, identifiant les éléments essentiels à sa croissance optimale. Nous abordons également l'importance génétique du blé, en soulignant les progrès réalisés en sélection variétale et amélioration génétique. Enfin, le chapitre se concentre sur les maladies du blé, en accordant une attention particulière à la fusariose qui affecte le blé en Algérie a l'aide des anciens littératures algériennes en mentionne (**Dr.BOUANAKA Hamza, Dr.BELLIL Ines, Et Pr.KHELIFI Douadi**).

Chapitre 02 : Ce chapitre présente les bases fondamentales de l'intelligence artificielle (**IA**), en mettant l'accent sur les techniques de l'apprentissage profond, telles que les réseaux de neurones convolutifs (**CNN**), qui permettent une analyse automatique et précise des données complexes. Nous expliquons de manière claire le fonctionnement de l'apprentissage profond, en détaillant les mécanismes d'entraînement des modèles, depuis l'extraction des caractéristiques jusqu'à la classification des données. Ensuite, nous illustrons les contributions de l'**IA** en bio-informatique, notamment dans l'étude du génome du blé, où ces méthodes facilitent l'identification de maladies comme la fusariose et l'annotation automatique des séquences **ADN**. Enfin, nous décrivons les outils spécialisés utilisés pour détecter les variations génétiques et analyser les segments clés du génome, démontrant ainsi l'impact des approches d'intelligence artificiel dans la recherche agronomique.

Chapitre 3 : Dans Le troisième chapitre nous avons détaillé notre contribution et compris les donnes utilisées, les modèles propose et les résultats obtenus.

Chapitre 01 : Un regard approfondi sur le blé

1. INTRODUCTION

Le blé (**Triticum spp.**), est une plante herbacée annuelle de la famille des Poacées, joue un rôle vital dans l'alimentation mondiale (**pain, pâtes, semoule**) grâce à sa richesse en glucides, protéines et fibres. Cultivé dans divers climats, il reste dominant en zones tempérées, soulignant son importance agricole et économique.

Le blé est une céréale monocotylédone autogame. Ses feuilles alternes poussent sur une tige cylindrique (chaume), et ses fleurs forment des épis composés d'épillets aplatis. Les grains, protégés par une enveloppe riche en fibres (**son**), se développent après fécondation.(Guellier, 2017) Il est cultivé dans des climats variés, des régions sèches (**États-Unis, Australie, Afrique du Nord**) aux zones humides (**Europe**). La moitié de la production mondiale provient de régions irriguées, notamment en **Inde, Pakistan et Chine.**(Baloch, 1999)

Cependant, sa production est confrontée à plusieurs défis, notamment les maladies (comme la fusariose), les changements climatiques et la nécessité d'améliorer les rendements tout en préservant l'environnement.

Dans ce chapitre, nous présentons un aperçu sur l'origine et l'histoire du blé, sa classification taxonomique, la composition nutritionnelle moyenne du grain de la culture, son importance génétique, les maladies du blé y compris la fusariose et ses variantes qui affectent le blé en Algérie.

2. Origine et histoire de la culture du blé

Les trois principales céréales, le blé, le riz et le maïs, représentent l'alimentation de base pour les populations du monde entier. Au cours de l'évolution de la civilisation indo-européenne, le blé s'est imposé comme la céréale prédominante des peuples occidentaux vivant sous un climat tempéré. (Henry et De Buyser, 2001)

Le blé figure parmi les principales denrées alimentaires de l'humanité. La saga du blé est intrinsèquement liée à celle de l'humanité et de l'agriculture ; sa culture remonte à avant l'histoire et constitue une caractéristique emblématique de l'agriculture néolithique, qui a vu le jour en Europe il y a **8000** ans. Il semblerait que la plus ancienne culture soit. Le blé tendre est apparu entre **5000** et **6000** ans avant notre ère dans le croissant fertile, puis s'est répandu en Europe à partir de la Grèce.(Doussinault et al., 1992)

Les données historiques (**figure 1**) indiquent que la dissémination du blé depuis cette région s'est opérée à la fois vers l'Afrique, l'Asie et l'Europe, avec une voie commerciale

privilégiée passant par l'Italie et la Sicile pour approvisionner le Maghreb.(Boulal et al., 2007)

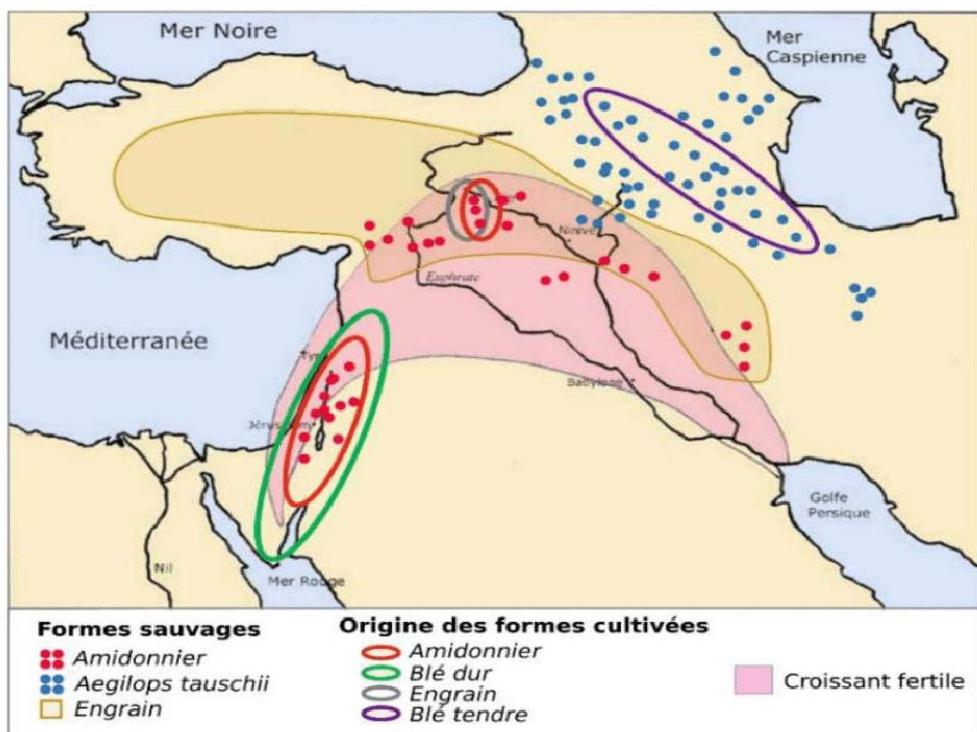


Figure 1 : L'origine géographique des blés cultivés. Adapté de Dubcovsky and Dvorak (2007).

En Algérie, Léon Ducellier (1878-1937) en particulier, parcourant le blé, fit au début du siècle le recensement d'une flore mal connue. Il a identifié et étudié les nombreuses sortes qui habitaient les champs cultivés, collecté les spécimens les plus distinctifs, les plus prolifiques, les plus résistants à la sécheresse ou à certaines maladies. Avant l'arrivée des Français, le blé tendre était méconnu en Afrique du Nord. (Lery, 1982)

Initialement, les blés ont progressé sans l'intervention de l'homme, puis sous l'influence des premières pratiques agricoles de sélection. (Henry et De Buyser, 2001)

D'après Sears (1954) et Okamoto (1962) in Auriauet al., (1992), Belaid (1996), Feillet (2000) et Henry et De Buyser (2001), Les deux variétés de céréales les plus répandues sont : le blé dur (*Triticum durum* Desf) : **AABB** ($2n = 4 \times 28$). Tétraploïde et le blé tendre (*Triticum aestivum* L) : **AABB DD** ($2n = 6 \times 28 + 2 \times 6 = 42$). Hexaploïdie, (Figure 2)

Blé tendre vs blé dur



Figure 2 : Blé tendre et blé dur Source

La parenté génétique des blés est compliquée et pas entièrement comprise. Le croisement naturel entre le *Triticum monococcum* (**génome A**) et l'*Aegilops* (*bicornis*, *speltoides*, *longissima* ou *searsii*) (**génome B**) a conduit à la création d'un blé dur sauvage de type **AABB** (*Triticum turgidum* ssp. *dicoccoïdes*), qui a par la suite évolué graduellement vers *Triticum turgidum* ssp. *Dicoccum*, puis vers *Triticum durum* [blé dur cultivé] (**Figure 02**). Les blés tendres cultivés (**AABBDD**) proviendraient d'un croisement, aussi naturel, entre *Triticum turgidum* sp. *dicoccum* (**AABB**) et *Aegilops squarrosa* (**DD**). (Feillet, 2000 ; Henry et De Buyser, 2001)

3. Classification taxonomique du blé

Le schéma ci-dessous résume la position taxonomique du **triticum aestivum** dans le règne végétal. (Bonneuil et al., 2006)

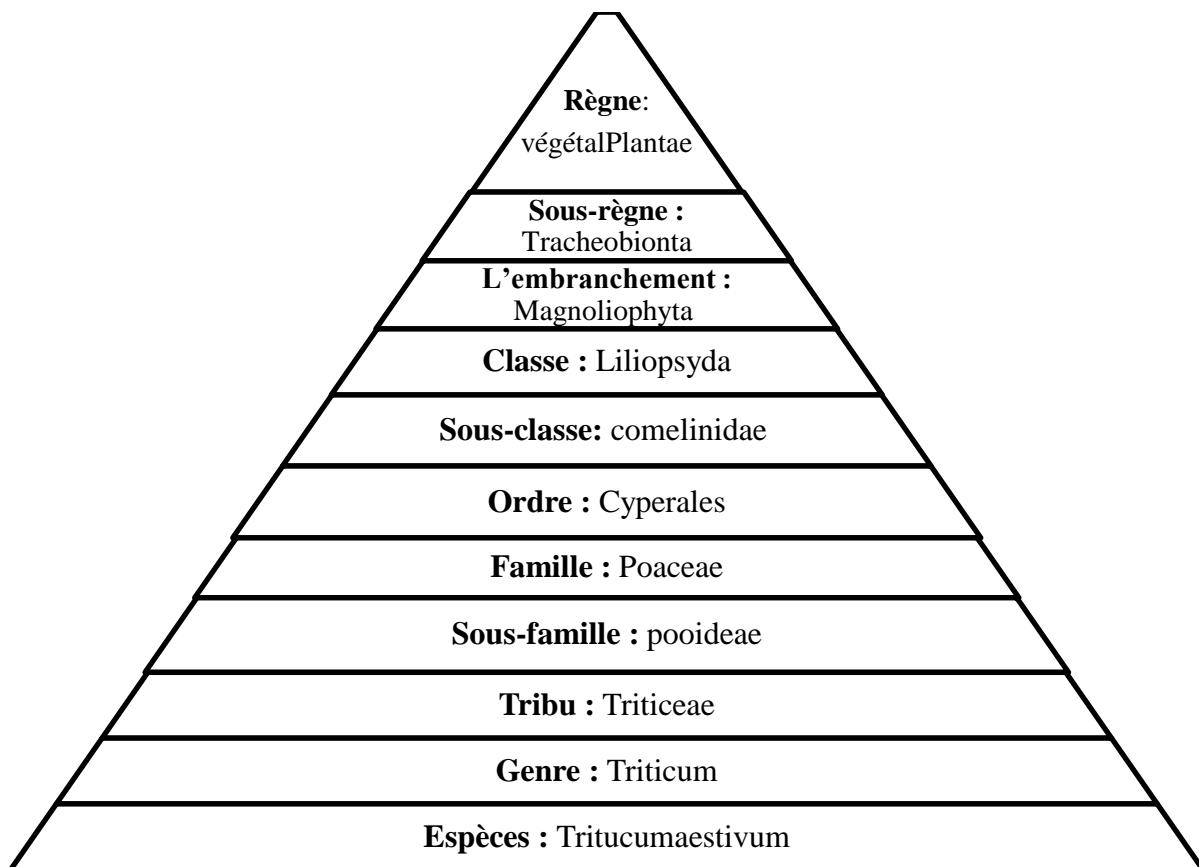


Figure 3 : Hiérarchie taxonomique des plantes : exemple du triticum

4. Composition nutritionnelle moyenne du grain de blé

Le grain est principalement constitué d'amidon (environ 70%), de protéines (10 à 15%) selon les variétés et les conditions de culture, les autres constituants, pondéralement mineurs (quelques % seulement), sont les lipides, la cellulose, les sucres libres, les minéraux et les vitamines (**tableau 1**). (Feillet, 2000)

Tableau 1 : Composition chimiques du blé

Composant	Pourcentage moyen
Glucides (amidon)	60-70%
Protéines (dont gluten)	10-15%
Lipides	1-3%
Minéraux (fer, magnésium, zinc, phosphore)	1-2%
Vitamines (B1, B2, B3, E)	Traces importantes
Fibres alimentaires	10-15%

5. L'importance génétique du blé

Le blé (**Triticum spp.**) revêt une importance génétique considérable grâce à la complexité de son génome, la variété de ses espèces et sa contribution essentielle à la sécurité alimentaire internationale. Il est crucial d'améliorer génétiquement cette plante pour booster les rendements, améliorer sa résistance aux maladies et aux conditions météorologiques extrêmes.(IWGSC *et al.*, 2018)

En août 2018, l'International Wheat Genome Sequencing Consortium (IWGSC), dont l'INRA joue un rôle prépondérant, a dévoilé dans la revue Science la première séquence de référence du génome du blé. Cette réalisation remarquable en science, d'autant plus impressionnante vu la taille et la complexité de ce génome – qui est cinq fois plus grand que celui de l'homme et quarante fois plus que celui du riz –, a été rendue possible par la coopération d'équipes françaises, dont le CEA, le CNRS, ainsi que les universités de Clermont-Auvergne, d'Évry-Val-d'Essonne, Paris-Sud et Paris-Saclay.

Cela facilitera notamment l'identification des gènes essentiels pour l'agronomie, proposant de nouvelles opportunités pour la sélection variétale du blé et l'amélioration de sa culture en présence à des défis mondiaux. Initialement, il représente une percée cruciale pour l'apprehension des processus de fonctionnement et d'évolution de ce génome d'une complexité hors du commun.(INRAE, 2018)

Le blé tendre (*Triticum aestivum*) est l'espèce la plus cultivée, générant pour environ **95%** de la production globale de blé. Cette espèce hexaploïde (AABBDD - ayant six jeux de chromosomes), issu d'une hybridation naturelle entre différentes espèces ancestrales de graminées, se différencie par sa forte capacité d'adaptation et sa productivité élevée. Il contient environ **17 milliards** de paires de bases, soit 5 fois plus que le génome humain.

En ce qui concerne le blé dur (**Triticum durum**), il s'agit d'une espèce tétraploïde (possédant quatre ensembles de chromosomes) connue pour sa riche teneur en protéines et sa capacité à résister à la sécheresse. On le valorise particulièrement pour la production de pâtes alimentaires.(AgrifoodMarket, no date)

Ce génome polyploïde, bien qu'il le rende difficile à analyser, lui accorde également une diversité génétique considérable, cruciale pour l'adaptabilité à des environnements divers.

Depuis des plusieurs siècles, l'homme choisit des variétés pour améliorer :

- Le rendement (ex : variétés à haut potentiel productif).
- La résistance aux maladies (ex : rouille du blé, fusariose).

- La tolérance aux stress climatiques (sécheresse, froid).
- La qualité nutritionnelle (protéines, gluten).

L'hybridation et le sélection variétal ont participé à améliorer la productivité, en particulier durant la Révolution Verte du XXe siècle, grâce aux variétés semi-naines mises au point par Norman Borlaug.**(Borlaug, 1968)**

Avec le séquençage du génome du blé (**publié en 2018**), de nouvelles techniques améliorent encore sa sélection :

- Indicateurs moléculaires (MAS) : identification des gènes pertinents pour faciliter la sélection.**(Battenfield et al., 2016)**
- CRISPR-Cas9 : Modification génétique focalisée à renforcer la résistance aux maladies sans passer par aux OGM.**(Zhou et al., 2023)**
- Blé génétiquement modifié (OGM) : Testé pour sa résistance à la sécheresse et aux insectes, il est peu cultivé en raison de réglementations précises.**(Tyczewska, 2023)**

Dans le contexte de la croissance démographique et du changement climatique, l'amélioration génétique du blé est essentielle pour préserver la sécurité alimentaire à sur le plan mondial, et diminuer l'usage d'eau et de produits de graine, minimiser l'effet des maladies et nocives. L'étude de la génétique du blé est donc essentielle pour assurer des cultures plus résilientes et durables.**(Ray et al., 2019)**

6. Les maladies du blé

6.1. Rouilles (Puccinia spp.) :

La maladie du blé rouillé est due à diverses espèces fongiques intégrant au genre Puccinia, notamment Puccinia triticina et Puccinia graminis qui sont les plus répandues. Ces champignons sont capables d'infecter toutes les parties visibles de la plante de blé, telles que les feuilles, les tiges, les épis et les glumes. Les signes de la rouille du blé peuvent varier selon le type de champignon et le stade de croissance de la plante, mais incluent généralement :

- **Affaiblissement de la plante :** La rouille peut diminuer la force du blé en limitant sa capacité à réaliser la photosynthèse et à déplacer des nutriments et de l'eau. Cela peut provoquer à une croissance incomplète, une chute de la production de céréales et, dans des situations extrêmes, à la mort de la plante.
- **Dommages aux épis et aux grains :** Dans certaines situations, la rouille du blé peut attaquer les épis et les glumes (**figure 4**), ce qui pourrait avoir un impact défavorable sur la croissance et la qualité des grains. Les infections sévères peuvent conduire à la

production de grains plus petits, déformés ou ridés, ainsi qu'à une diminution du nombre de grains par épi.(AgronoBlog, 2025)



Figure 4 : exemple d'un effet de la rouille sur le blé

6.2 Oïdium (*Blumeria graminis f. sp. Tritici*):

Une maladie cryptogamique capable d'affecter le blé tout au long de sa période de croissance. On la remarque sur les feuilles, les tiges et également sur les épis. Il s'agit d'une des maladies les plus simples à détecter sur le blé. L'oïdium se développe sous forme de mycélium sur les repousses de céréales et sur les plantes adventices présentes dans les champs. Les cleistothécies, qui représentent une forme de conservation des spores de la maladie, sont capables de résister à la température et à la sécheresse, ce qui permet au champignon de survivre.

Dans des conditions climatiques favorables (où les températures varient de 2 à 30°C avec un pic de contamination et de développement entre 15 et 20°C, accompagné d'une humidité élevée), les cleistothécies relâchent les ascospores capables de causer une infection en automne.



Figure 5 : Exemple d'un effet de l'Oïdium.(Bayer, 2021)

Au printemps, le mycélium qui était en dormance se met à se développer et produit des spores. Ils se multiplient de manière optimale à **15°C** dans un environnement où l'humidité relative dépasse **95%**. La germination des spores est inhibée par la présence d'eau libre. Sous des conditions de sécheresse, la formation de spores fraîches peut se produire en l'espace d'une semaine (voir la **Figure 5**).**(Bayer, 2021)**

6.3 Septoriose (*Zymoseptoria tritici*) :

Une maladie foliaire du blé peut se développer dès la phase de plantule. Les symptômes sont généralement remarqués en premier lieu sur les feuilles inférieures. Lorsque la feuille étandard et les deux dernières feuilles sont affectées par cette maladie pendant la période de floraison (ce qui conduira à l'échaudage du grain), il peut y avoir d'importantes pertes de rendement.

Les contaminations primaires ont lieu durant l'automne. Les ascospores et pycnidiospores qui se développent sur les restes de paille seront transportées par le vent et vont s'installer sur les nouvelles pousses de blé, facilitant ainsi leur croissance. Durant l'hiver, les conditions atmosphériques défavorables empêchent la progression. Au printemps, l'épidémie risque de se renouveler avec la montée des températures plus douces. La maladie se multiplie grâce aux pycnides qui provoquent un effet de « *splashing* » (projection des spores par les gouttes de pluie) des niveaux inférieurs vers les niveaux supérieurs, et aussi vers les plantes adjacentes. Le risque de contamination se trouve de la phase 2 nœuds jusqu'à la phase floraison comme illustré ci-dessous dans la **(Figure 6)**.**(Bayer, 2021)**



Figure 6 : Exemple d'un effet du Septoriose.**(Bayer, 2021)**

6.4. *Xanthomonas translucens* (Tache bactérienne) :

Cette bactérie a pour hôtes principaux l'orge, le seigle, le blé et le triticale. On peut la trouver dans certaines plantes herbacées telles que le brome, le Phalaris spp. Et le chiendent. Essentiellement, la bactérie se maintient dans les graines. Pour sa propagation et l'infection de la plante, un taux d'humidité élevé est nécessaire, accompagné de la présence d'eau libre sur les parties végétales. Elle s'introduit dans les végétaux via les stomates et les fractures. La transmission de la plante d'une plante à l'autre se produit par les éclaboussures causées par la pluie ou par le contact entre les plantes et les insectes, et est facilitée par les dommages causés par le gel. Son développement se fait idéalement à 22°C, cependant la bactérie tolère une vaste gamme de températures (15-30°C). Selon le type de bactérie, les conditions environnementales et la variété, les symptômes peuvent s'apparaître sur les feuilles ou les glumes.

Sur les feuilles : des stries imbibées d'eau, étroites, de teinte jaunâtre pour l'orge et le triticale, séparées au centre avec des bords d'aspect rouillé pour le blé. Dans des conditions d'humidité, les stries diffusent en parallèle avec les nervures. Un exsudat bactérien se manifeste par des gouttes de couleur brun sombre ou translucide lorsqu'elles sont exposées à la lumière. Ces gouttes peuvent s'écouler et sécher, formant une mince couche de dépôt qui peut s'écailler. Les stries peuvent converger et engendrer d'importantes lésions qui entraînent la dégradation de la feuille.

Sur les épis : des stries sombres entre les vaisseaux, qui démarrent à l'extrémité des glumes et des glumelles. Elles sont capables de fusionner et d'apporter une couleur noire aux épis. Les grains peuvent être détachés et de couleur brune, avec un poids minime (voir la figure 7). (Ephytia, 2018)

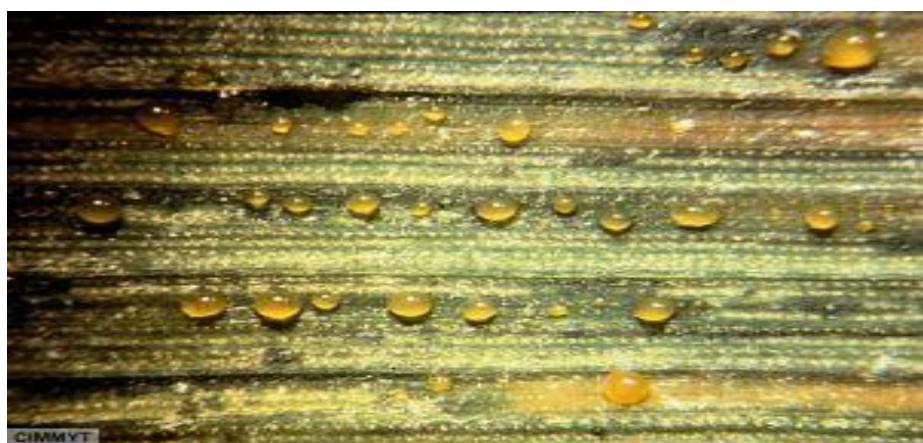


Figure 7 : Exemple d'un effet du *Xanthomonas translucens*. (Ephytia, 2018)

6.5. Mosaïque du blé (Wheat Streak Mosaic Virus - WSMV) :

Le WSMV et le WMV se diffusent sur diverses graminées cultivées, principalement sur le blé et l'orge, mais aussi sur des plantes adventices. L'acarien *Aceria tosicella*, à ses stades adulte et nymphal, est responsable de la transmission des deux virus. Toutefois, seules les nymphes ont la capacité de contracter le virus. Elles peuvent par la suite le transférer aux plantes hôtes pendant une période minimale de 9 jours. La transmission des virus à la descendance n'est pas possible. Concernant les symptômes se repose :

- Sur les plantes : décoloration des végétaux situés en bordure du champ, monoculture et retard dans leur développement.
- Sur les feuilles : des motifs de mosaïque ou de marbrures vertes accompagnées de stries jaunes, en parallèle mais fréquemment séparées, susceptibles d'évoluer vers une chlorose généralisée. On peut aussi observer des nécroses de couleur brun clair à beige dans le cas du WMV. Lors d'une invasion par les acariens, les feuilles se dressent avec les bords repliés. (Voir la **Figure 8**)
- Sur les épis : stériles ou sans grains, grains de petite taille et ratatinés (en ce qui concerne le WSMV). (Ephytia, 2017)

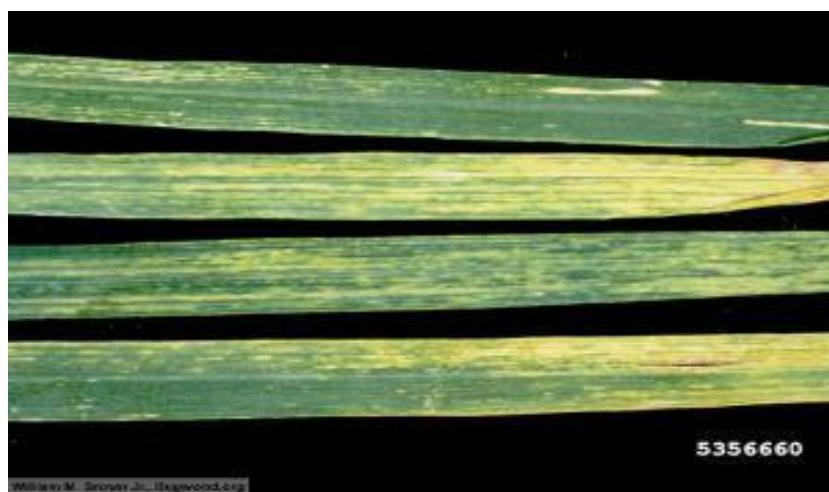


Figure 8 : Exemple d'un effet du Mosaïques du blé.(Ephytia, 2017)

6.6. La fusariose chez le blé

6.6.1. État général

L'épi fusarium est une maladie commune dans nos zones, elle impacte toutes les céréales de paille, mais principalement les productions de blé et d'orge en raison de leur importance dans notre environnement agricole. Dans ces deux cultures, la fusariose de l'épi provoque des chutes de rendement, cependant, ce sont principalement les toxines générées

dans le grain par les champignons nuisibles qui posent le plus de soucis à l'ensemble du secteur des céréales. L'avoine est bien moins capable aux infections et les dégâts sont rarement considérables. La fusariose ne néglige pas le seigle et le triticale. Actuellement, les variétés disponibles présentent une sensibilité relativement élevée, cependant elles sont peu fréquemment cultivées. Au Québec, la fusariose pose un enjeu permanent à la fois en termes de production et de vente de céréales. (**'RAP Grandes cultures 2017'**)

6.6.2. Symptômes

Dans le cas du blé, les grains infectés par le fusarium sont de petite taille et présentent une surface ridée. Ils présentent souvent une apparence semblable à celle de la craie. La détection de grains fusariés n'indique pas obligatoirement la présence de toxines ; à l'inverse, l'absence de grains clairement fusariés ne garantit pas l'absence de toxines.

La présence d'un ou plusieurs épillets décolorés sur les épis de couleur verte dans un champ de blé signale l'existence de la maladie (voir **Figures 9 et 10**).



Figure 9 : Fusariose de l'épi du blé ; épillets décolorés



Figure 10 : Fusariose de l'épi du blé

L'infection des épillets et leur répartition peuvent varier ; ils peuvent être concentrés sur une partie de l'épi, presque l'ensemble de l'épi peut être touché, ou encore, on peut noter des symptômes assez isolés à quelques épillets décolorés. Il est parfois possible d'observer une couleur rose ou orangée sur ces épillets, qui est liée à la fructification du champignon (**Figure 11**).('RAP Grandes cultures 2017')



Figure 11 : Fusariose de l'épi du blé ; la teinte orangée dénote la présence du champignon pathogène

6.6.3 Développement et infection de la maladie

Près de 17 espèces de *Fusarium* sont liées à cette maladie. C'est *Fusarium graminearum* qui provoque le plus de dégât. Les *Fusarium*, qui provoquent une infection au niveau des fleurs et des grains des plantes graminées, se propagent à travers les structures telles que les enveloppes florales, le rachis des épis et les tiges pendant la phase de sénescence de la plante. Les fragments végétaux issus des céréales à tige et du maïs qui se trouvent dans les champs constituent un vecteur de transmission du champignon. Les *Fusarium* se refroidissent facilement et persistent sur les résidus végétaux. Quand les conditions sont favorables pendant la période de croissance, les spores dérivant de ces résidus présents à la surface du sol arrivent aux épis (ou à la panicule de l'avoine) et peuvent contaminer les tissus floraux ainsi que les grains en cours de développement. Les spores peuvent être générées non seulement à partir des résidus du champ cultivé, mais également, sous certaines conditions, à partir des fragments des champs adjacents qui sont ensuite emportés par le vent. L'accumulation de résidus de culture en grande quantité favorise la production d'inoculum. Par conséquent, les résidus de culture du maïs peuvent représenter une source significative de spores vivantes pour les cultures avoisinantes, et ce, malgré la distance qui les sépare.

Le moment crucial pour l'infection, qui correspond à l'entrée du champignon dans les tissus de l'épi, commence lors de l'épiaison et couvre les jours suivants. Durant cette période, le niveau d'infection sera principalement influencé par la pluie et l'humidité. Le blé est

principalement infecté pendant la phase de floraison (pollinisation ou anthèse), qui ne dure que quelques jours. Dans des conditions d'humidité adéquates, les spores situées sur les étamines germent et génèrent du mycélium qui se développe jusqu'à la fleur en cours de formation. Par ailleurs, durant cette phase de développement, la fleur du blé est davantage développée et vulnérable à l'infestation par le champignon. C'est à ce moment précis du développement du blé que l'infection influence le plus la production de grains. (**'RAP Grandes cultures 2017'**)



Figure 12 : Zoom sur la Fusariose du blé

6.6.4 Les types de la fusariose qui affecte le blé en Algérie

Depuis quelques années, des agriculteurs cultivant des céréales ont remarqué dans leurs champs des zones clairsemées et insuffisamment germées ou des zones comportant des épis décolorés et vides (**Figure 13**). Ce phénomène est généralement lié à des problèmes phytosanitaires causés par des insectes tels que les vers blancs, les vers fil de fer et le volatile des chaumes, ou par certaines maladies fongiques comme les fusarioSES, dont les agents pathogènes se propagent via les plantes, les résidus de culture ou le sol. Ces pressions ont un impact défavorable sur la production de grains et de paille des blés. La diminution du rendement en grain est attribuée, d'une part, à la réduction de la densité des plants lors de la germination (**défaillance des semis**) et à la diminution du nombre de grains par épi et du poids de 1000 grains en cas d'agression post-épiage, d'autre part.



Figure 13 : des zones comportant des épis décolorés et vides.(DZ, 2021)

Des études effectuées en Algérie (**ENSA, INPV, Universités**) ont identifié la présence de **12** espèces de *Fusarium* spécifiques au blé, avec une prédominance importante de **3** espèces : *Fusarium culmorum*, *Fusarium graminearum* et *Microdochium nivale*. Parmi les facteurs favorables au développement des fusarioSES sur le blé en Algérie, on peut citer les suivants :

Dans les systèmes céréaliers algériens, divers éléments agro-climatiques et méthodes culturales participent de manière significative à la fréquence des fusarioSES. Les conditions sèches courantes dans les principales régions de culture de céréales favorisent un milieu propice à la prolifération des agents pathogènes fongiques. Ce phénomène est amplifié par des rotations culturales inadéquates, comme les séquences blé-blé, blé-jachère ou blé-pomme de terre, qui sont couramment observées dans le pays.(DZ, 2021)

Durant la phase cruciale de floraison, les conditions climatiques, notamment les précipitations et la chaleur intense, ainsi que les méthodes d'irrigation par aspersion (surtout dans les systèmes à pivot situés au Sud), agissent comme des facteurs aggravants. Un autre facteur participant à la persistance des inoculums est la pratique de cultiver le maïs en alternance avec le blé dans les régions du sud où sont installés des systèmes d'irrigation par pivot.

L'emploi de semences non protégées ou mal protégées, l'usage de fongicides de qualité inférieure pour le traitement des graines, ainsi que l'adoption de méthodes culturales simplifiées et de semis direct sans préparation du sol appropriée font aussi partie des pratiques agricoles inappropriées. Ces éléments conjugués génèrent un environnement particulièrement propice à l'apparition et à la prolifération des fusarioSES au cœur des cultures céréaliers en Algérie.

La culture de variétés résistantes, comme les blés durs locaux, ainsi que le traitement des semences avec un fongicide efficace, tel que le fludioxonil, constituent deux méthodes pour combattre les fusarioSES du blé. Cela permet non seulement de lutter contre les maladies, mais aussi d'assurer une germination rapide et uniforme de la culture. C'est la seule méthode de combat efficace, prompte et abordable contre ces maladies dangereuses.(DZ, 2021)

CHAPITRE 2 :

Techniques de

l'intelligence

artificielle

1. Introduction

L'intelligence artificielle (IA) se représente à la compétence des systèmes informatiques à imiter des capacités cognitives humaines comme l'apprentissage, le raisonnement et la prise de décisions.(stryker and kavlakoglu, 2024) Cette technologie s'appuie sur l'étude de grands volumes de données pour permettre aux machines de reproduire des comportements humains, comprenant des tâches simples de classification à des opérations industrielles avancées. L'intelligence artificielle, dans son nature même, cherche à instruire les ordinateurs de manière à ce qu'ils imitent des comportements semblables à ceux de l'homme.

Ce chapitre présente une étude bibliographique sur les techniques d'IA y compris l'histoire de l'intelligence artificielle, les différents types d'intelligence artificielle, les techniques d'IA en Bio-informatique comme les Apports de l'IA dans l'Analyse des Séquences, Applications de l'IA pour la Détection des Variants Génétiques, et l'IA et Annotation Génomique.

2. L'histoire de l'intelligence artificielle

Avant 1949, les ordinateurs avaient la capacité d'exécuter des instructions, mais ils ne pouvaient pas se souvenir de leurs actions, en raison de l'incapacité à stocker ces instructions. Dans son article de **1950** intitulé « **Computing Machinery and Intelligence** », **Alan Turing** a introduit l'idée de créer des machines capables d'intelligence et d'évaluer cette dernière. **Cinq ans plus tard**, le premier programme d'intelligence artificielle a été dévoilé dans le contexte du Dartmouth Summer Research Project on Artifical Intelligence (**DSPRAI**). Cet événement a agi comme un catalyseur pour la recherche en IA au cours des décennies à venir.

Dans les années **1980**, l'intelligence artificielle a été réanimée grâce à l'élargissement de la collection d'algorithmes et à l'augmentation de financements. Les méthodes de « **deep Learning** », qui permettent aux ordinateurs d'apprendre par l'expérience, ont été introduites par **John Hopfield** et **David Rumelhart**. Pour sa part, **Edward Feigenbaum** a instauré les « **systèmes experts** », qui reproduisent la prise de décision à l'exemple de l'humain. En contrainte d'un manque de fonds publics et de couverture médiatique, l'intelligence artificielle a connu un développement considérable et de nombreux principes historiques ont été dépassés au cours des vingt années suivantes. **En 1997**, le grand maître et champion du monde d'échecs **Gary Kasparov** a été défait par Deep Blue d'IBM, un programme informatique spécialement conçu pour jouer aux échecs. Cette même année, le programme de reconnaissance vocale conçu par Dragon Systems a été intégré à Windows. **Kismet**, un robot qui a la capacité de détecter et d'exprimer des émotions, a également été élaboré par **Cynthia Breazeal**.

En 2016, le projet **AlphaGo** de Google a succédé du maître du jeu de Go, Lee Se-dol. Puis, en 2017, Libratus, un supercalculateur spécialisé dans le poker, a surpassé les meilleurs joueurs humains.(**Glossaire HPE, 2025**)

3. Les techniques d'intelligence artificielle

Dans l'illustration présentée ci-dessous (**Figure 14**), on observe clairement les principes de base de l'intelligence artificielle :

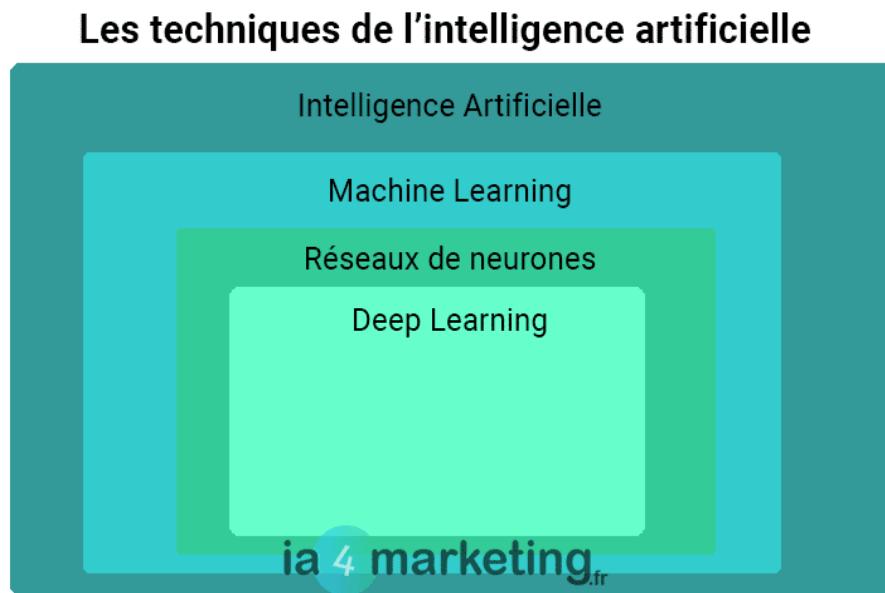


Figure 14 : schéma qui présente les principes de base d'IA

3.1. Apprentissage automatique (Machine Learning)

Le machine Learning est une technologie d'intelligence artificielle qui permet aux ordinateurs de se configurer à partir de données, sans qu'ils nécessitent besoin d'être spécifiquement programmés pour chaque mission. À Contraire de la programmation classique qui fixe des règles prédéfinies, l'apprentissage automatique offre aux systèmes la capacité de détecter des schémas dans les données, d'anticiper des résultats et d'évoluer de manière autonome grâce à l'expérience acquise. De manière plus précise, un algorithme de machine Learning 'se forme' en étudiant des exemples antérieurs : il en tire des principes généraux qu'il peut par la suite adapter à de nouvelles circonstances, modifiant ses réponses selon les résultats obtenus. L'aptitude à s'adapter de manière autonome confère au machine Learning une puissance considérable pour résoudre des problématiques complexes et en constante évolution.(**IBM, 2021**)

3.1.1. Étapes de la construction d'un modèle de machine Learning

La construction d'un modèle de Machine Learning s'effectue en quatre phases essentielles. Généralement, un Data Scientist est chargé de gérer et de superviser ce processus.

- La première étape consiste à Sélectionner et préparer un ensemble de données d'apprentissage nécessaires à la résolution du problème. Ces données peuvent être étiquetées ou non étiquetées.
- La deuxième étape consiste à Choisir un algorithme à exécuter sur les données d'apprentissage :
 - Si les données sont étiquetées, l'algorithme peut être de type régression, arbre de décision ou basé sur des instances.
 - Si les données ne sont pas étiquetées, l'algorithme peut être un algorithme de regroupement, un algorithme d'association ou un réseau neural.
- La troisième étape est entraînée l'algorithme pour créer le modèle.
- La quatrième et dernière étape est Utiliser et améliorer le modèle.

Il existe trois méthodes de machine Learning. L'apprentissage « **supervisé** » fonctionne avec des données étiquetées et nécessite moins d'entraînement. L'apprentissage « **non supervisé** » est utilisé pour classer des données non étiquetées en identifiant des schémas et des relations. L'apprentissage « **semi-supervisé** » utilise un petit ensemble de données étiquetées pour guider la classification d'un plus grand ensemble de données non étiquetées.(**Robert, 2020**)

3.2. L'apprentissage profond

L'apprentissage profond, aussi connu sous le nom de deep Learning, est une variante de l'intelligence artificielle issue de l'apprentissage automatique (**machine Learning**). Contrairement à la programmation qui se limite à exécuter des règles préétablies, cette technologie permet à la machine d'acquérir une connaissance par elle-même.(**Céline Deluzarche, 2024**)

3.2.1. Historique d'apprentissage profond

L'idée de l'apprentissage profond est relativement nouvelle, ayant commencé à prendre forme dans les années **2000**. Malgré la nouveauté du concept d'apprentissage profond, les fondements sur lesquels il se construit, tels que les réseaux de neurones artificiels, ne sont pas récents. Les réseaux de neurones artificiels trouvent leurs racines **en 1943**, lorsque **Warren**

McCulloch et **Walter Pitts** ont introduit leur premier modèle mathématique et informatique du neurone biologique, appelé neurone formel. Le neurone formel s'inspire étroitement du fonctionnement des neurones biologiques. Pour rechercher ses origines, on doit remonter jusqu'au milieu du vingtième siècle. **En 1943**, le concept du "neurone formel" voit le jour, illustrant une représentation simplifiée de l'activité du cerveau humain. Aussi, **en 1957**, le « **perceptron** » a été créé et est reconnu comme le premier réseau de neurones artificiels.

Par ailleurs, c'est **en 1950** qu'apparaît le « test de Turing », un examen élaboré pour mesurer la capacité d'une machine à reproduire le comportement humain. Cet événement constitue un élément important dans l'histoire de l'intelligence artificielle, même si sa progression subit une décélération significative jusqu'aux années **1980**.

Au cours **des années 2000**, nous avons observé l'apparition officielle et l'évolution notable de progrès théoriques dans la compréhension du fonctionnement des réseaux de neurones profonds. Ces derniers ont commencé à prouver leur efficacité dans divers domaines tels que la reconnaissance vocale, la traduction automatique, l'identification d'objets sur les images, ainsi que plusieurs autres applications. La croissance d'Internet et l'accroissement de la numérisation des informations ont provoqué une explosion du nombre de jeux de données accessibles pour la machine Learning et le deep Learning. Les réseaux de neurones profonds ont été perfectionnés et élaborés, facilitant la formation de réseaux de neurones à plusieurs niveaux. De plus, l'apparition des réseaux de neurones convolutionnels, une structure spécifiquement développée pour le traitement de données structurées, a également été remarquée.**(McCulloch and Pitts, 1943)**

3.2.2. L'apprentissage profond dans la bio-informatique

À l'époque actuelle, convertir le Big Data en informations précieuses dans le domaine biomédical et la recherche biologique représente l'un des principaux enjeux en bio-informatique. Ainsi, l'utilisation de l'apprentissage profond en bio-informatique a permis une meilleure compréhension des données, ce qui lui a valu d'être mise en avant dans le milieu universitaire et industriel.

Dans le domaine de la bio-informatique, l'utilisation de l'apprentissage automatique pour extraire des informations à partir d'importantes quantités de données est très répandue. Des techniques comme les machines à vecteurs de support, les modèles de Markov cachés et les réseaux bayésiens sont fréquemment mises en œuvre dans les domaines de la génomique, de la protéomique et de la biologie des systèmes.**(Min, Lee and Yoon, 2016)**

Les algorithmes classiques de machine Learning se basent en grande partie sur des représentations de données nommées « **caractéristiques** », habituellement élaborées par des ingénieurs disposant d'une expertise sectorielle. Toutefois, il est toujours compliqué de déterminer les caractéristiques les plus appropriées. L'apprentissage profond, une sous-catégorie du Machine Learning, a vu le jour récemment grâce au Big Data, à la capacité de calcul parallel et distribué ainsi qu'aux algorithmes évolués. Il a transcendi les contraintes du passé, suscité un intérêt grandissant depuis le début **des années 2000** et facilité des avancées notables dans plusieurs secteurs.(**Haoyang Li et al., 2020**)

L'apprentissage profond a prouvé son utilité dans le domaine de la biologie. La figure ci-dessous illustre l'application de l'apprentissage profond en biologie :

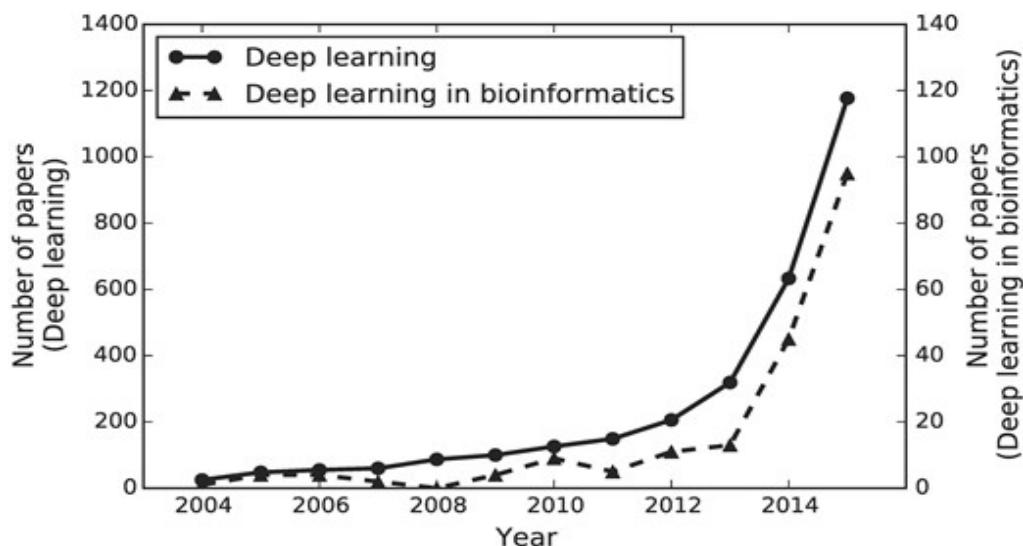


Figure15 : Nombre approximatif d'articles publiés sur l'apprentissage profond et l'apprentissage profond dans la bio-informatique par année.

3.2.3. Fonctionnement du l'apprentissage profond

Le deep Learning repose sur l'utilisation d'un réseau de neurones artificiels qui s'inspire du fonctionnement du cerveau humain. (**Voir la Figure 16**)

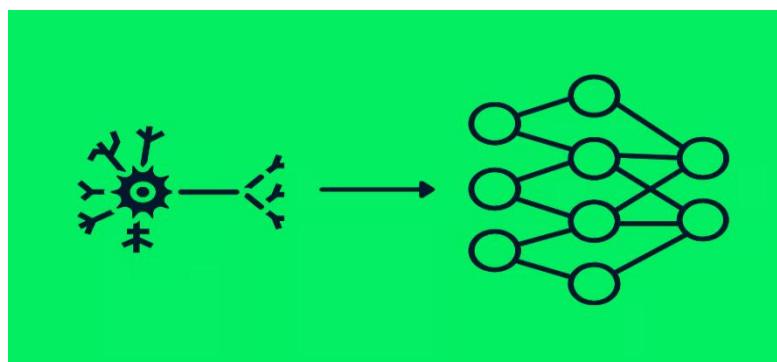


Figure 16 : La similitude entre les neurones et les réseaux neuronaux

Ce réseau est constitué de plusieurs dizaines, voire de centaines de « couches » neuronales, chaque couche traitante et analysant les données de la précédente. Par exemple, le système apprendra d'abord à identifier les lettres avant de se concentrer sur les mots dans un texte, ou à déterminer la présence d'un visage sur une image avant de tenter de reconnaître l'individu en question.(Céline Deluzarche, 2024)

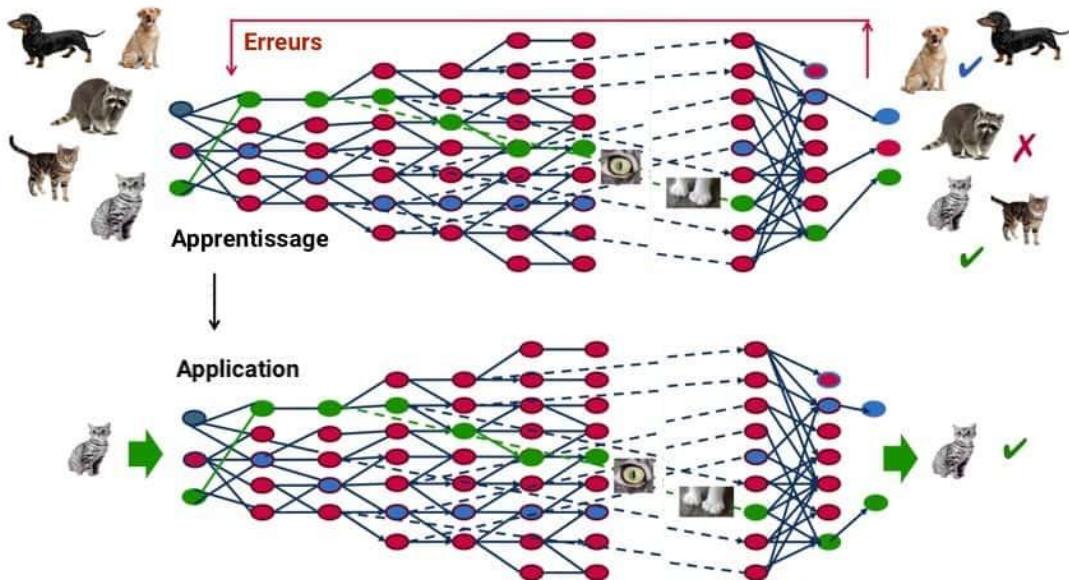


Figure 17 : un schéma représente le fonctionnement du deep Learning

En pratique, le deep Learning automatise la détection des caractéristiques via son architecture à plusieurs niveaux. Chaque niveau du réseau de neurones identifie graduellement des schémas de complexité croissante, en se basant sur le travail effectué par la couche antérieure. Cette méthode hiérarchique justifie pourquoi ces systèmes excellent dans la manipulation de données brutes, sans le besoin d'un prétraitement nécessaire pour l'apprentissage automatique traditionnel.(Marietti, 2025)

3.2.4. Techniques de l'apprentissage profond

3.2.4.1. Réseaux de neurones convolutionnels (CNN)

Un Réseau de Neurones Convolutionnel (CNN) est un modèle d'apprentissage automatique développé pour examiner des données en forme de grille, telles que les images, tiré de l'inspiration du cortex visuel des animaux.

Il s'appuie principalement sur trois sortes de couches : celles qui sont convolutionnelles, de regroupement et entièrement reliées. Les filtres appliqués par les couches convolutionnelles servent à extraire des caractéristiques, tandis que les couches de pooling diminuent la dimension des cartes de caractéristiques. Les couches

totallement connectées, quant à elles, associent ces caractéristiques aux résultats souhaités tels que les labels de classification. Cette structure hiérarchique permet aux CNN d'apprendre des motifs complexes et des associations spatiales, ce qui les rend performants pour la classification d'images, la détection d'objets et la segmentation d'images.(Yamashita *et al.*, 2018)

3.2.4.2. Réseaux de neurones récurrents (RNN)

Les RNN sont performants pour modéliser des séquences de données, comme les séries temporelles biologiques ou les séquences génomiques. On les utilise pour prédire des séquences, modéliser la structure secondaire des protéines, entre autres.

3.2.4.3. Réseaux de neurones génératifs (GAN)

Ces modèles sont utilisés pour produire de nouvelles données biologiques crédibles, telles que des images de cellules ou des séquences génomiques artificielles. Ils servent à élargir les ensembles de données ou à explorer le domaine des données biologiques.

3.2.4.4. Réseaux de neurones siamois

Ces réseaux, créés pour la mise en parallèle de paires de données, sont bénéfiques pour des missions comme la régulation de séquences génomiques ou la prévision de la ressemblance entre structures protéiques. Donc, sur la figure suivante, nous identifions certaines des méthodes que nous avons citées, qui doivent passer par des phases précises selon le fonctionnement de l'apprentissage profond :

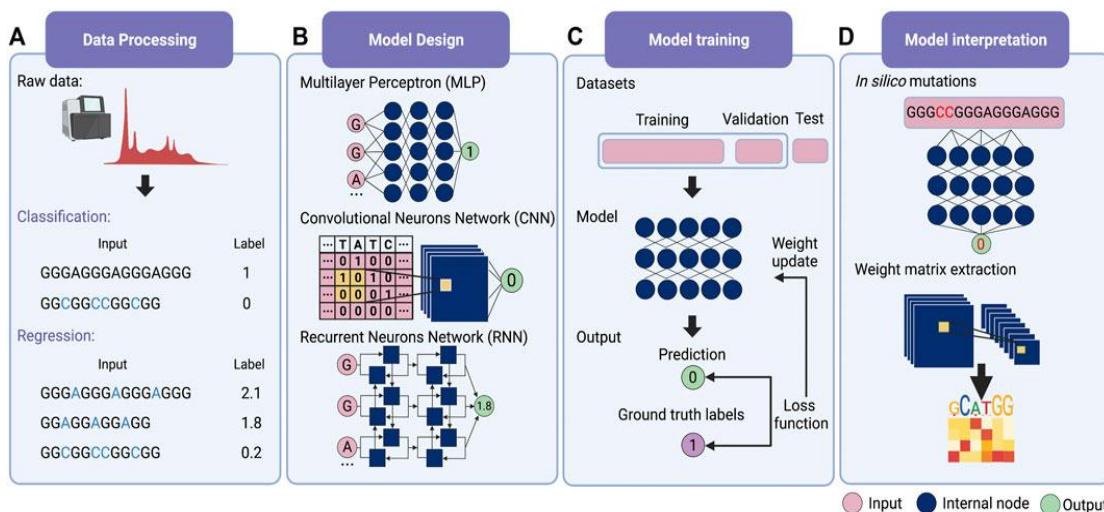


Figure 18 : Un ensemble schématique du flux de travail de l'apprentissage profond.(Yu, Qi and Ding, 2022)

Cet aperçu schématique du processus de travail en deep Learning montre que :

(A) Gestion des données. L'apprentissage supervisé exige une annotation claire des données, y compris les identifiants de classe pour les problèmes de classification et les valeurs pour les problèmes de régression. (B) Conception de modèles. Les trois principales structures d'apprentissage profond sont le Perceptron Multicouche (**MLP**), le Réseau de Neurones Convolutif (**CNN**) et le Réseau de Neurones Récurrent (**RNN**). Habituellement, les modèles de Deep Learning combinent diverses architectures basées sur des structures de données. (C) Entraînement du modèle. Le jeu de données d'entraînement est initialement segmenté en un ensemble d'apprentissage, un paquet de validation et un paquet de test. Par la suite, les informations d'entrée sont transférées au modèle afin de déterminer les valeurs anticipées. On utilise la fonction de perte pour mesurer l'écart entre les valeurs estimées et les valeurs réelles, facilitant ainsi la mise à jour des poids du modèle. (D) Interprétation des modèles. L'importance des caractéristiques peut être déterminée via des mutations *in silico*. Pour le modèle **CNN**, on peut aussi évaluer les caractéristiques en récupérant la matrice de poids du filtre. (Yu, Qi and Ding, 2022)

4. Contributions de l'IA en Bio-informatique

L'intégration de l'intelligence artificielle (**IA**) dans la bio-informatique transforme l'analyse des séquences d'ADN et d'ARN, rendant les processus plus rapides, plus précis, et ouvrant la voie à des découvertes révolutionnaires en génomique. L'intersection de l'intelligence artificielle et de la bio-informatique, couramment désignée comme « **IA bio-informatique** », facilite le traitement des volumes énormes de données produites par les techniques de séquençage à grande taille, tout en fournissant des perspectives novatrices sur la fonction et le développement des génomes.

4.1. Les apports de l'IA dans l'analyse des séquences

L'intelligence artificielle, notamment l'apprentissage automatique et l'apprentissage profond, est fondamentale dans le domaine de l'annotation génomique, la prédiction des structures de l'ARN et la détection des variations génétiques. Historiquement, l'analyse des séquences génétiques exigeait des algorithmes basés sur des règles précises, mais les modèles d'IA actuels offrent la possibilité de traiter des données complexes sans se baser sur des suppositions précédentes. Par exemple, les réseaux neuronaux profonds sont capables de repérer des motifs génétiques précis en lien avec des maladies complexes, rendant ainsi le diagnostic génétique plus simple.

Un exemple de cette application serait l'emploi d'algorithmes d'apprentissage approfondi pour anticiper les sites d'épissage des ARN. (**Zhang et al., 2019**) confirmé que les modèles de deep Learning surpassent les méthodes conventionnelles en matière de prédiction précise des sites d'épissage, ce qui est essentiel pour déchiffrer les mécanismes régulateurs de l'expression génique. Tout comme (**Eraslan et al., 2019**) qui ont fait appel à des réseaux neuronaux profonds pour prédire la structure tridimensionnelle de l'ARN à partir de séquences, ce qui a permis d'approfondir notre compréhension des interactions entre ARN et protéines. (**Poplin et al., 2018**)

4.2. Applications de l'IA pour la détection des variants génétiques

L'intelligence artificielle a aussi la capacité d'identifier et de catégoriser rapidement les variants génétiques, y compris les mutations rares, à partir des données de séquençage. Des applications telles que DeepVariant, élaboré par Google, exploitent les techniques d'apprentissage profond pour transformer les données brutes de séquençage en identifications de variants génétiques avec une précision exceptionnelle. Ces méthodes dépassent fréquemment les approches conventionnelles en matière de sensibilité et de spécificité, un atout majeur pour la détection de variants dans des situations cliniques.

L'incorporation de l'intelligence artificielle dans les pipelines de détection des variants influence directement la médecine sur mesure. En associant l'intelligence artificielle à des bases de données génomiques, les scientifiques sont en mesure de relier certains variants génétiques à des phénotypes cliniques. Cela ouvre la voie pour des thérapies personnalisées basées sur le profil génétique d'un patient. (**Poplin et al., 2018**)

4.3. IA et annotation génomique

L'IA a également permis d'améliorer l'annotation des génomes, qui implique la détection de gènes, d'exons, d'introns et d'autres composantes fonctionnelles dans une séquence ADN. Les instruments d'annotation fondés sur la machine Learning peuvent rendre automatique le processus de détection des éléments fonctionnels, diminuant par conséquent la nécessité de se fier à des bases de données existantes et à des annotations faites à la main. Pour illustrer, (**Friedberg et al., 2019**) a employé des algorithmes d'apprentissage supervisé afin d'optimiser l'annotation des génomes, ce qui a permis une identification plus précise des gènes et des éléments de régulation.

CHAPITRE 03 :

Contribution et étude expérimentale

1.Introduction

Face à l'augmentation de la population mondiale, le secteur agricole doit relever un défi majeur pour produire plus tout en optimisant ses ressources. La bio-informatique apporte des solutions cruciales pour comprendre et améliorer la croissance des plantes. Cependant, le suivi manuel des cultures montre ses limites en termes de temps et de fiabilité. C'est là que l'intelligence artificielle intervient en synergie avec la biologie en analysant automatiquement l'état des plantes, les besoins en eau et nutriments, ou les signes précoces de maladie, l'IA devient un outil précieux pour concilier productivité agricole et respect des processus biologiques naturels. Cette alliance entre biologie et technologie ouvre la voie à une agriculture plus intelligente et durable.

Dans ce chapitre Nous proposons ici notre apport fondé sur une approche d'intelligence artificielle en utilisant des réseaux de neurones convolutifs (**CNN**), en détaillant l'entraînement, l'évaluation et les résultats du modèle. À partir des séquences nucléiques **ITS** et des photos prise sur des instances réelles du Blé.

2.Matériel

2.1 Données utilisés

2.1.1 Données images

Notre étude s'appuie sur un jeu de données varié représentant différentes conditions du blé. Ces données incluent à la fois des images et des séquences nucléotidiques.

Nous avons utilisé des images de blé sains (**healthy – 1676 images**) et d'autre atteints des maladies du blé non sains (**unhealthy**) suivantes : rouille jaune du Blé (**yellow Rust - 1301 images**), la brûlure du Blé (**black Rust -798 images**), brûlure des feuilles Blé (**Leaf blight Spot - 1309 images**).

Tableau 2 : les données des images utilisées.

Nom du fichier	Taille du fichier	Nombre d'image
Healthy	0,98 GO	1676
Yellow Rust	0,58 GO	1301
Leaf Blight	0,58 GO	1309
Black Rust	0,25 GO	798

2.1.2 Données séquences

Nous avons également exploité des séquences nucléiques issues de la base **GENBANK** en format **FASTA**, incluant des séquences fongiques algérienne du blé sain (**Triticum**) et infecté par la fusariose de l'épi à l'aide des anciens littératures algériennes en mentionne (Dr.**BOUANAKA Hamza**, Dr.**BELLIL Ines**, Et Pr.**KHELIFI Douadi**), qui sont présentés dans le **tableau 3**

Tableau 3 : données des séquences utilisées

Nom du fichier	Taille du fichier
sequences_dataset.csv	386 KO

2.2. Configuration de la machine

Nous avons travaillé avec un ordinateur simple qui est détaillé dans le tableau suivant :

Tableau 4 : Les caractéristiques de l'ordinateur

Composants	Caractéristiques
Processeur	I7 11éme générations
RAM	8GO
ROM	256GO
Carte graphique	INTEL HD Graphics
Système d'exploitation	Windows 10

2.3 Langages de programmation et environnement de travail

Les langages de programmation utilisés sont détaillés dans ce qui suit :

- **Python :** Python est un langage de programmation open source, compatible avec tous les systèmes d'exploitation. Son approche orientée objet et ses nombreuses bibliothèques le rendent polyvalent pour divers usages : création de logiciels, analyse de données ou administration système.(**Futura, 2021**)
- **Visual Studio Code :** VS Code est un IDE open-source et cross-platform proposé par Microsoft. Initialement conçu pour le développement web, il prend désormais en charge divers langages (C++, Python, Java...) grâce à son système d'extensions modulaires. Sa compatibilité avec les trois principaux OS et sa communauté active en font un outil polyvalent et personnalisable.(**Bility, 2023**)

- **Google Colab** : Google Colab est une solution en ligne gratuite pour exécuter des notebooks Python/Jupyter. Sans nécessiter d'installation, elle donne accès à des ressources informatiques et aux bibliothèques ML courantes, le tout directement depuis un navigateur.(Jaillet, 2024)

2.4 Bibliothèques

Les bibliothèques utilisées sont détaillées dans ce qui suit :

- **Scikit-learn** : est une bibliothèque Python puissante et simple d'utilisation pour la machine learning. Elle offre une interface intuitive permettant d'implémenter facilement divers algorithmes d'apprentissage automatique (classification, régression, clustering...). Construite sur NumPy, SciPy et matplotlib, cette bibliothèque s'intègre parfaitement dans l'écosystème scientifique Python pour le traitement de données et la visualisation. Son approche unifiée et ses nombreuses fonctionnalités en font un outil incontournable pour les projets de data science.(Groupe GEMA – IA School, 2025a)
- **Pandas** : est une bibliothèque Python incontournable pour l'analyse et la manipulation de données. Conçue pour traiter efficacement des données tabulaires, des séries temporelles et divers formats de jeux de données, elle offre des fonctionnalités puissantes et une syntaxe intuitive qui en font un outil essentiel pour tout travail de data science en Python.(Andersen, 2023)
- **Matplotlib** : est la bibliothèque de visualisation Python la plus complète, permettant de créer des graphiques professionnels de qualité publication. Cette librairie polyvalente prend en charge tous les types de visualisations - des simples courbes **2D** aux diagrammes complexes **3D** - avec une précision et une personnalisation poussée. Son intégration native avec NumPy et Pandas en fait l'outil idéal pour les scientifiques et data analystes exigeants.(Wild Code School, 2023)
- **NumPy** : constitue le socle du calcul scientifique en Python, offrant des tableaux multidimensionnels performants (**ndarrays**) et une vaste collection de fonctions mathématiques optimisées. Cette bibliothèque fondamentale permet d'effectuer des opérations vectorielles et matricielles à haute vitesse, formant la base indispensable de l'écosystème data science Python. Son architecture optimisée en fait l'outil privilégié pour le traitement numérique intensif.(Wild Code School, 2023)
- **Seaborn** : est une puissante bibliothèque Python spécialisée dans la visualisation statistique, permettant de créer des graphiques à la fois élégants et riches en informations. Construite au-dessus de **Matplotlib** et parfaitement intégrée avec Pandas,

elle simplifie la génération de visualisations complexes tout en offrant des designs sophistiqués. Son interface intuitive en fait l'outil idéal pour explorer et présenter des données statistiques de manière claire et visuellement impactante.(**Wild Code School, 2023**)

- **Tqdm** : est une bibliothèque Python légère et efficace qui ajoute des barres de progression intuitives à vos boucles et itérations. Son implémentation simple mais puissante permet de visualiser instantanément l'avancement de votre code, offrant un feedback clair sur l'exécution des tâches tout en conservant une grande flexibilité d'utilisation.(**Franschläder, 2025**)
- **TensorFlow** : est une bibliothèque open-source dédiée à l'apprentissage automatique qui utilise une approche innovante basée sur des graphes de calcul.(**Groupe GEMA – IA School, 2025b**)

3. Méthodes

Dans le cadre de ce mémoire, deux approches d'intelligence artificielle ont été utilisées pour prédire les maladies du blé. La première méthode permet d'identifier le type de maladie à partir d'une image, tandis que la seconde détermine la pathologie à partir des séquences protéiques.

3.1 Modèle de classification par séquence

Ce modèle utilise un réseau de neurones convolutifs (**CNN**) pour classer des séquences du blé en "saines" ou "malades" (**fusarium**), à partir d'un ensemble des étapes suivant :

3.1.1 Préparation des données

Nous avons utilisé un dataset qui contient un ensemble de séquences **ITS** Téléchargées depuis **GenBank** sous format **FASTA**, puis on a converti les séquences en format numérique (**One-hot encoding**) c'est à dire que Chaque base (**A, C, G, T**) est convertie en un vecteur binaire à l'aide des bibliothèques Python pour créer le fichier **CSV** et avoir une base de données bien organiser.

- **Importation des Bibliothèques**

Voici comment on a importé les bibliothèques nécessaires pour le nettoyage des données dans la figure suivante :

```

import pandas as pd
import numpy as np

data = pd.read_csv('sequences_dataset.csv')

def one_hot_encode(seq):
    mapping = {'A': [1,0,0,0], 'C': [0,1,0,0], 'G': [0,0,1,0], 'T': [0,0,0,1]}
    return np.array([mapping.get(nuc, [0,0,0,0]) for nuc in seq])

print("Aperçu des données :")
print(data.head())
import numpy as np

```

Figure 19 : les bibliothèques nécessaires

- **Structure du fichier CSV**

Le fichier est structuré en ordre suivant : séquence, maladie (**HEALTHY** ou **UNHEALTHY**)

	sequence	maladie
0	TAAAAAGGTCGACAATGGAATAGAAGGTACTATTCCCACGCTAACT...	unhealthy
1	AATTCGTTACGTTATCTCTGTCAGTCAGGGGGCCTTATGGCA...	unhealthy
2	TAGTAGTGAACACTTCCTAAACCGCTTGTCCTCAGGCTTATAACC...	unhealthy
3	TGATGGTTCGCAGTGCAGGGGACGGGTGAGTTCTACAATAATTGCC...	unhealthy
4	CGGACGCCGGTCCCCTTATCGTAGATTCCCTAACGTTAGCCAATG...	healthy
5	CAAAGCTGTAAGCGCGTACGTTGTTCCAAGTCCTCACCTTCTT...	healthy
6	CATGAAAATTGACCAAAAGGTTGTAAGTCAGAAATTGTCGAAGCC...	healthy
7	GAGGCAAATTGATATCGCAATATCCCCAACTTGACTGAGTGTAA...	unhealthy
8	ATCGAGCCAACACAGAAAAGTGCTTGTAAATCAAATTGGGAACCCAA...	unhealthy
9	GCAGGAATCACATAGCTCATCCGAGGAGCTTGAGGTCGCGTCCC...	healthy

Figure 20 : structure du fichier csv

3.1.2 Construction du modèle

Pour la construction du modèle nous avons crées un modèle séquentiel de deux couches de convolution 1D (voir la **figure 21**)

```

model = Sequential([
    Conv1D(64, kernel_size=5, activation='relu', input_shape=(max_length, 1)),
    MaxPooling1D(pool_size=2),
    Dropout(0.3),
    Conv1D(128, kernel_size=3, activation='relu'),
    MaxPooling1D(pool_size=2),
    Dropout(0.3),
    Flatten(),
    Dense(256, activation='relu'),
    Dropout(0.4),
    Dense(len(label_encoder.classes_), activation='softmax')
])

```

Figure 21 : modèle séquentiel

- ✓ Première couche de convolution 1D :

- Elle utilise **64** filtres pour détecter des motifs dans la séquence.
- Chaque filtre analyse un groupe de **5** éléments consécutifs
- Un maxpooling qui réduit la dimension tout en conservant les informations importantes.
- ✓ Deuxième couche de convolution 1D :
- Elle Utilise **128** filtres et Chaque filtre analyse **3** éléments consécutifs.
- maxpooling pour réduire la taille, suivi d'un dropout pour la régularisation.
- La couche flatten pour transformer les données multi-dimensionnelles de la sortie des couches précédentes en un vecteur.

3.1.3 Entraînement et évaluation du modèle

On a créé une règle pour arrêter automatiquement l'entraînement du modèle si la performance sur les données de validation qui s'appelle early stopping

```
early_stop = EarlyStopping(monitor='val_loss', patience=5, restore_best_weights=True)

history = model.fit(
    X_train, y_train,
    epochs=10,
    batch_size=32,
    validation_data=(X_test, y_test),
    callbacks=[early_stop],
    verbose=2
)
```

Figure 22 : la fonction principale d'entraînement du modèle.

- ✓ Paramètre d'entraînement :
- epochs=10 : le modèle sera entraîné pendant maximum **10** passes sur les données, sauf si early stopping intervient plus tôt.
- batch_size=32 : l'entraînement se fait par mini-lots de **32** exemples (pour optimiser la mémoire et la vitesse).

Ces données et paramètres ont été adopter par le script qui a entraîné le modèle du CNN sur **50 lots** avec une durée moyenne de **8** millisecondes par époche ce qui a conduit à une perte finale de **0.376** et une précision finale de **99.5%**.

```
Epoch 8/10
50/50 - 1s - loss: 0.2542 - accuracy: 0.9031 - val_loss: 0.1297 - val_accuracy: 0.9900 - 1s/epoch - 28ms/step
Epoch 9/10
50/50 - 1s - loss: 0.1877 - accuracy: 0.9287 - val_loss: 0.0674 - val_accuracy: 0.9900 - 1s/epoch - 28ms/step
Epoch 10/10
50/50 - 1s - loss: 0.1149 - accuracy: 0.9613 - val_loss: 0.0376 - val_accuracy: 0.9950 - 1s/epoch - 28ms/step
13/13 [=====] - 0s 8ms/step - loss: 0.0376 - accuracy: 0.9950

✓ Test Accuracy: 99.50%
13/13 [=====] - 0s 7ms/step
```

Figure 23 : Script de l'entraînement

3.2. Modèle de classification par image

Ce modèle utilise un réseau de neurones convolutifs (**CNN**) pour classer des images du blé téléchargés depuis la plateforme Kaggle (**Wheat Disease Detection**) <https://www.kaggle.com/code/kushagra3204/wheat-disease-detection/input> à partir d'un ensemble des étapes suivant :

3.2.1 Préparation des données

La méthode met en œuvre un **Dataset** contient des images saines "**HEALTHY**" et non saines "**UNHEALTHY**" collectées pour réseaux neurones convolutions (**CNN**) et les organiser dans deux fichiers pour l'entraînement et validation du modèle en fournissant des fonctionnalités requises qui sont importées par les bibliothèques dans **la figure 24**

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report, confusion_matrix
import tensorflow as tf
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
import os
```

Figure 24 : Bibliothèques et modules importés

- On ajoute un code de détection des plantes avec **MobileNetV2** pour vérifier si une image contient une plante ou non, il est capable d'extraire les trois prédictions les plus probables pour l'image fournie. Ensuite, il vérifie si au moins l'une de ces prédictions correspond à des mots-clés associés au monde végétal tels que "**Plant**", "**Tree**", "**Leaf**", ou d'autres termes similaires qui sont illustrés dans la figure. Cette approche permet d'identifier rapidement et automatiquement les images contenant des plantes, sans avoir besoin d'entraîner un nouveau modèle.

```

model = MobileNetV2(weights='imagenet')
PLANT_KEYWORDS = [
    "flower", "tree", "leaf", "plant", "garden", "forest", "shrub", "vine",
    "daisy", "rose", "sunflower", "tulip", "orchid", "lily", "fern", "moss",
    "cactus", "bamboo", "palm", "pine", "oak", "maple", "willow", "birch",
    "corn", "wheat", "rice", "soybean", "tomato", "potato", "carrot", "lettuce",
    "grass", "herb", "crop", "vegetable", "fruit", "seedling", "sapling",
    "bonsai", "ivy", "algae", "seaweed", "succulent", "flora", "botanical"
]

def is_plant_image(image_data, confidence_threshold=0.05):
    try:
        img = image_data.resize((224, 224))
        img_array = image.img_to_array(img)
        img_array_expanded_dims = np.expand_dims(img_array, axis=0)
        processed_img = preprocess_input(img_array_expanded_dims)

        predictions = model.predict(processed_img)
        decoded_predictions = decode_predictions(predictions, top=3)[0]

        is_plant = False
        top_results = []
        for _, label, score in decoded_predictions:
            top_results.append((label, float(score)))
    
```

Figure 25 : Code de détection des plantes

✓ Organisation des fichiers

Le fichier a été organisé sous classe bien organiser pour l'entraînement du modèle

```

fusariose_detection/
└── dataset/
    ├── train/
    │   ├── healthy/
    │   │   ├── image1.jpg
    │   │   ├── image2.jpg
    │   │   └── ...
    │   └── unhealthy/
    │       ├── image1.jpg
    │       ├── image2.jpg
    │       └── ...
    └── validation/
        ├── healthy/
        │   ├── image1.jpg
        │   └── ...
        └── unhealthy/
            ├── image1.jpg
            └── ...

```

Figure 26 : Structure du fichier

✓ Chargement des données

On charge les chemins du l'entraînement et validation vers le dossier des images contenant les sous dossiers par classe “**HEALTHY**” et “**UNHEALTHY**”

```

train_generator = train_datagen.flow_from_directory(
    'C:/Users/Manar Baali/Desktop/fusariose_detection/dataset/train',
    target_size=(256, 256),
    batch_size=32,
    class_mode='categorical',
    subset='training'
)

valid_generator = train_datagen.flow_from_directory(
    'C:/Users/Manar Baali/Desktop/fusariose_detection/dataset/validation',
    target_size=(256, 256),
    batch_size=32,
    class_mode='categorical',
    subset='validation'
)

```

Figure 27 : Le chemin du dossier

✓ **Les paramètres**

- Target size= **(256, 256)** : redimensionne toutes les images à **256x256** pixels.
- Batch size=**32** : envoie **32** images par lot à chaque étape de l'entraînement.
- Class mode=**'catégorial'** : indique que la sortie du modèle est multi classe génère des vecteurs one-hot (**[1, 0], [0, 1]**).

3.2.2 Construction du Modèle

Un modèle du l'api séquentiel de **keras** de trois couches de convolution **2D** pour la construction du modèle, traite des images de taille **256x256** avec **3** canaux de couleur. Avec un maxpooling pour réduire la taille du dimensionnalité.

```

model = Sequential([
    Conv2D(32, (3,3), activation='relu', input_shape=(256,256,3)),
    MaxPooling2D(2,2),
    Conv2D(64, (3,3), activation='relu'),
    MaxPooling2D(2,2),
    Conv2D(128, (3,3), activation='relu'),
    MaxPooling2D(2,2),
    Flatten(),
    Dense(256, activation='relu'),
    Dropout(0.5),
    Dense(num_classes, activation='softmax')
])

```

Figure 28 : Modèle de construction

✓ **Les parameters**

- **Conv2D + MaxPool** pour extraire des motifs visuels
- **Flatten** pour Aplatir les données pour les couches denses

- **Dense** + **Dropout** pour Apprendre les combinaisons discriminantes
 - **Softmax** pour Prédire la classe (**HEALTHY /UNHEALTHY**)
- ✓ **Le script du séquentiel**

(**Modèle. summary ()**) sous forme de tableau

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_1 (Conv2D)	(None, 125, 125, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d_2 (Conv2D)	(None, 60, 60, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 128)	0
flatten (Flatten)	(None, 115200)	0
dense (Dense)	(None, 256)	29491456
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 2)	514

Figure 29 : Séquentiel du modèle

3.2.3. Entraînement et évaluation du modèle

Le modèle va faire **20** époche sur le Dataset d'entraînement Généré à l'aide de (**Train_Generator**), de façon ces données de validation utilisées à la fin de chaque époche pour évaluer les performances sur des données jamais vues.

```
# Entraînement
history = model.fit(
    train_generator,
    epochs=20,
    validation_data=valid_generator,
    callbacks=callbacks
)
```

Figure 30 : Code d'entraînement

Une itération complète sur l'ensemble des données d'entraînement. Ici, le modèle a été entraîné pendant **20** époches et Mesure d'erreur du modèle sur les données d'entraînement

avec une perte de **0.17** et précision avec une durée moyenne de **3** secondes par époche. (Voir la **Figure 31**)

```
Epoch 15/20
76/76 [=====] - 252s 3s/step - loss: 0.1810 - accuracy: 0.9281 - val_loss: 0.3524 - val_accuracy: 0.8725 - lr: 1.0000e-05
Epoch 16/20
76/76 [=====] - 250s 3s/step - loss: 0.1748 - accuracy: 0.9314 - val_loss: 0.3739 - val_accuracy: 0.8593 - lr: 1.0000e-05
Epoch 17/20
76/76 [=====] - 250s 3s/step - loss: 0.1790 - accuracy: 0.9257 - val_loss: 0.3537 - val_accuracy: 0.8626 - lr: 1.0000e-05
Epoch 18/20
76/76 [=====] - 251s 3s/step - loss: 0.1797 - accuracy: 0.9285 - val_loss: 0.3490 - val_accuracy: 0.8709 - lr: 1.0000e-05
Epoch 19/20
76/76 [=====] - 250s 3s/step - loss: 0.1729 - accuracy: 0.9310 - val_loss: 0.3604 - val_accuracy: 0.8709 - lr: 1.0000e-05
Epoch 20/20
76/76 [=====] - 252s 3s/step - loss: 0.1794 - accuracy: 0.9240 - val_loss: 0.3466 - val_accuracy: 0.8791 - lr: 1.0000e-05
```

Figure 31 : Le script de l'entraînement

3.3. L'interface

Dans le but de rendre l'utilisation de nos modèles plus accessible, nous avons développé une interface conviviale intégrant l'ensemble des fonctionnalités mises en place. Cette interface permet aux utilisateurs d'effectuer des prédictions sur les maladies du blé, soit à partir d'images de plantes prises sur le terrain, soit à partir de séquences nucléiques.

Grâce à cette application web, il est possible d'estimer avec précision le risque d'apparition d'une maladie, apportant ainsi une aide précieuse au diagnostic. Les figures ci-après présentent des captures d'écran illustrant l'interface développée.

3.3.1 Construction de l'interface

Ce code représente une interface web simple et intuitive destinée à la prédiction du stress chez les plantes à partir d'une image. Grâce au technologies **Front-end**.

✓ HTML

La structure de la page permet à l'utilisateur de téléverser une image et de lancer une prédiction en un clic.

```
<div class="container">
  <h1>🌿 Détection du stress des plantes</h1>

  <input type="file" id="imageInput" accept="image/*"><br>
  <img id="imagePreview" alt="Aperçu de l'image"><br>

  <button onclick="predict()">Prédire</button>
  <div id="result"></div>
</div>
```

Figure 32 : Structure HTML

✓ CSS

Le style **CSS** rend l'interface moderne et agréable visuellement, avec un fond vert clair, une boîte blanche centrée et des boutons arrondis au design professionnel.

```
<style>
  body {
    font-family: Arial, sans-serif;
    background: #f0fdf4;
    display: flex;
    justify-content: center;
    align-items: center;
    height: 100vh;
    margin: 0;
  }
}
```

Figure 33 : Partie d'une structure CSS

✓ JAVA SCRIPT

Le JavaScript permet d'ajouter de l'interactivité :

```
function predict() {
  if (imagePreview.src === "" || imagePreview.style.display === "none") {
    alert("⚠ Veuillez d'abord sélectionner une image.");
    return;
  }

  const options = ["HEALTHY", "UNHEALTHY"];
}
```

Figure 34 : Partie d'un code JAVASCRIPT

Il affiche automatiquement un aperçu de l'image sélectionnée et simule une prédiction ("HEALTHY" ou "UNHEALTHY"). Lorsque l'utilisateur clique sur le bouton. Cette simulation permet de tester le comportement de l'interface avant de la connecter à un véritable modèle d'intelligence artificielle. L'ensemble forme une base solide pour construire une application de détection visuelle des maladies des plantes.

3.3.2. Structure de l'interface

Nous avons conçu deux interfaces utilisateur complémentaires : l'une dédiée à la prédiction à partir d'images, l'autre à partir de séquences protéiques.

- **Interface de prédiction par image** : elle permet à l'utilisateur de diagnostiquer l'état de santé d'une feuille de blé qui est présentée dans la figure en important une image via un bouton « **choisir un fichier** ». Dès que l'image est chargée, on clique sur le Button « **prédir** » puis le modèle de prédiction se déclenche automatiquement pour analyser la feuille et déterminer si elle est saine ou malade.
- **Si on importe une image d'un autre objet par exemple** : une image d'une voiture va marquer une erreur sur la barre des résultats.



Figure 35 : L'interface en cas d'image

- **Interface de prédiction par séquence** : cette interface met à disposition un champ de saisie où l'utilisateur peut entrer une séquence nucléique. Après validation via le bouton « **prédir** », le système traite la séquence et indique si elle est liée à une infection par la fusariose.



Figure 36 : L'interface en cas de la séquence

4. Résultat

4.1 Modèle de classification par séquence

4.1.1 Matrice de confusion

Le modèle présente des résultats remarquables :

Il a correctement identifié **213** séquences "**HEALTHY**" (**vrais positifs**) et **185** séquences "**UNHEALTHY**" (**vrais négatifs**). Seules deux séquences "**HEALTHY**" ont été mal classées comme "**UNHEALTHY**", ce qui correspond à deux faux positifs. Fait impressionnant, aucune séquence "**UNHEALTHY**" n'a été classée à tort comme "**HEALTHY**", ce qui signifie qu'il n'y a eu aucun faux négatif. Ce profil qui est figuré indique une capacité de détection très fiable, particulièrement pour les séquences pathologiques.

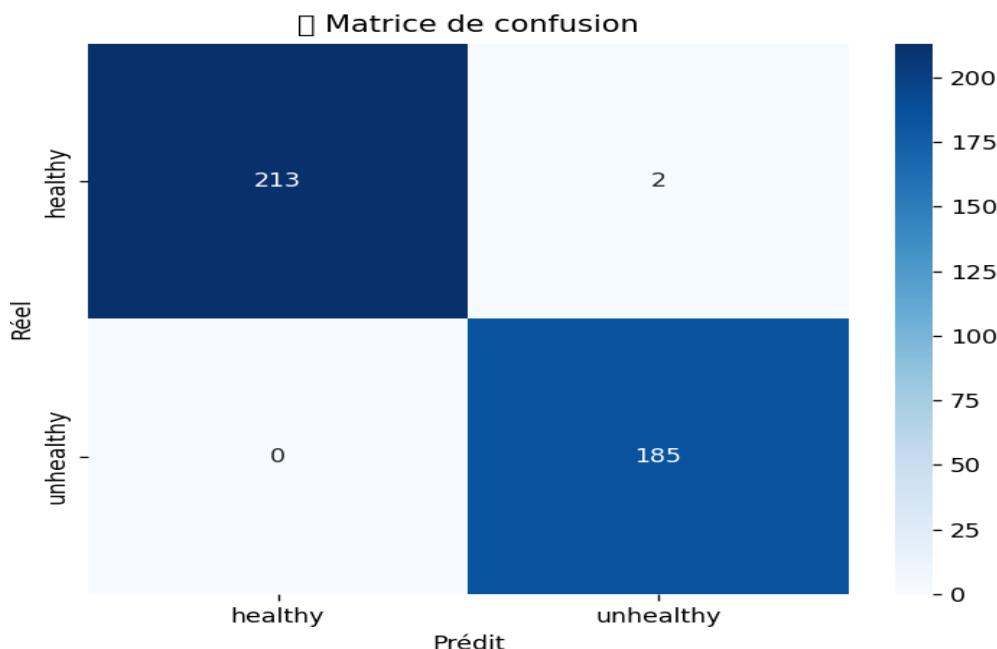


Figure 37 : Matrice de confusion

✓ Résumé des résultats :

- **213 vrais positifs (VP)** : les séquences "**HEALTHY**" bien classées.
- **185 vrais négatifs (VN)** : les séquences "**UNHEALTHY**" bien classées.
- **2 faux positifs (FP)** : des "**HEALTHY**" mal classées comme "**UNHEALTHY**".
- **0 faux négatifs (FN)** : aucune séquence "**UNHEALTHY**" n'a été classée à tort comme "**HEALTHY**".

4.1.2 Valeurs obtenues du modèle de séquence

Le tableau (5) suivant montre les résultats des valeurs obtenues du modèles.

Tableau 5 : Valeurs obtenues du modèle séquence

Paramètre	Valeur
Précision	0.99
Val précision	0.93
Perte	0.11
Val perte	0.03

4.1.3 Plateforme du modèle de séquence

Dans la **figure 38** ci-dessous lorsque l'utilisateur clique sur le bouton '**Prédire**', le modèle associé à notre interface évaluera la séquence fournie en s'appuyant sur ses données d'entraînement.

**Figure 38 :** Résultat de la prédiction cas de la séquence

4.2 Modèle de classification par image

4.2.1 Diagramme de précision (Training Accuracy et Validation Accuracy)

- **Courbe d'entraînement (Training Accuracy)** : Cette courbe représente l'évolution de la justesse des prédictions du modèle sur les données d'apprentissage au cours des époques. Son augmentation indique que le modèle affine progressivement sa capacité à reconnaître les motifs présents dans les données d'entraînement.
- **Courbe de validation (Validation Accuracy)** : Cette métrique évalue la performance du modèle sur un ensemble de données distinct, non utilisé pendant l'apprentissage. Une progression régulière de cette courbe est souhaitable, car elle témoigne d'une bonne

généralisation du modèle à de nouvelles données, au-delà du simple apprentissage par cœur.

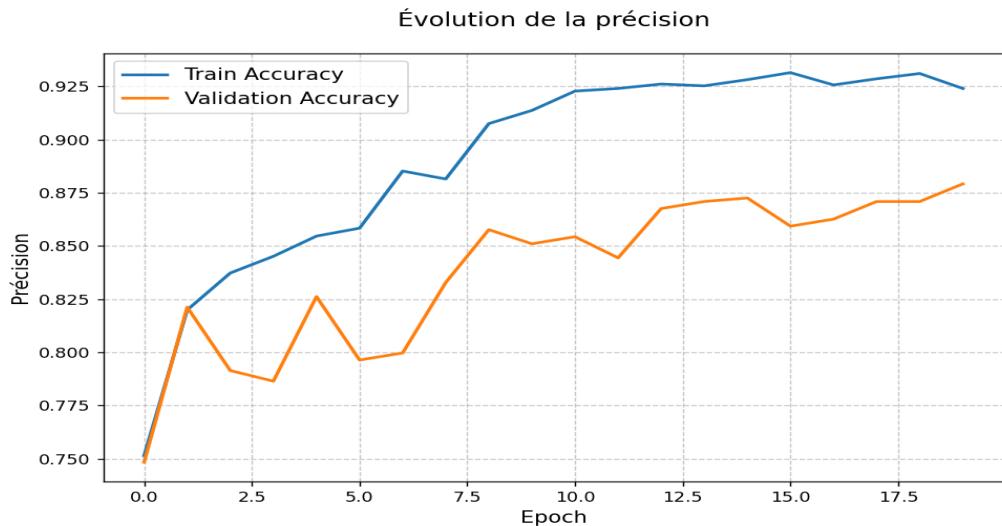


Figure 39 : Diagramme de précision (Training Accuracy et Validation Accuracy)

4.2.2 Diagramme de perte (Training Loss et Validation Loss)

- **Courbe de perte d'entraînement (Training Loss)** Montre comment l'erreur du modèle diminue au fil des époques sur les données d'apprentissage. Quand cette courbe baisse régulièrement, c'est que le modèle apprend correctement en ajustant ses paramètres.
- **Courbe de perte de validation (Validation Loss)** Indique comment l'erreur évolue sur des données que le modèle ne voit pas pendant l'entraînement. Dans le meilleur des cas, cette courbe devrait aussi diminuer, montrant que le modèle généralise bien à de nouvelles données.

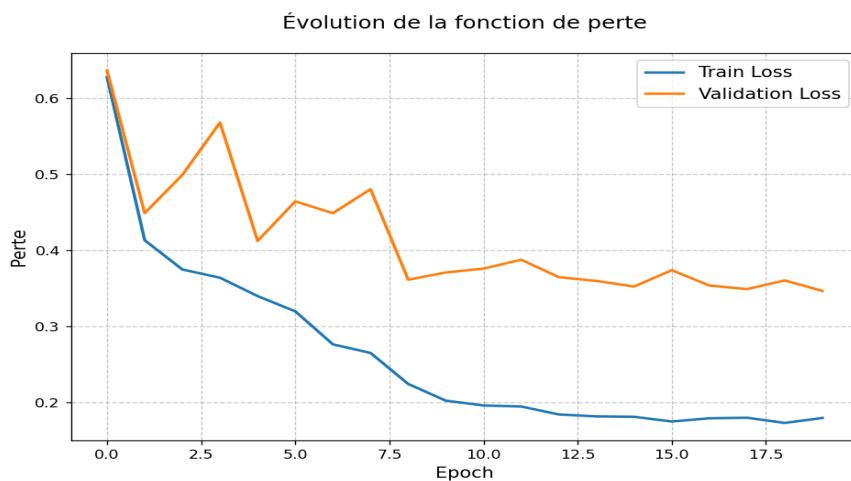


Figure 40 : Courbe de perte (Training Loss et Validation Loss)

4.2.3. Valeurs obtenues du modèle d'image

Le tableau (6) suivant montre les résultats des valeurs obtenues du modèles.

Tableau 6 : les valeurs obtenues du modèle d'image

Paramètre	Valeur
Précision	0.92
Valeur précision	0.88
Perte	0.17
Valeur perte	0.34

4.2.4 Plateforme du modèle d'image

L'utilisateur peut y téléverser une photo via un bouton “**choisir un fichier**”, et celle-ci est immédiatement affichée pour confirmation visuelle. Sur la figure présentée après avoir cliqué sur le bouton "Prédire", le système reposant le modèle entraîné et analyse l'image et retourne un diagnostic visuel clair : "HEALTHY", affiché en vert pour renforcer l'aspect positif du résultat.



Figure 41 : Résultat de la prédiction d'image

Si l'image ne correspond pas aux critères attendus par exemple (**absence de plante détectable**) Lorsque l'utilisateur clique sur le bouton "Prédire", le système renvoie un message d'erreur clair qui est visible dans la figure "**Cette image ne semble pas représenter une plante**", invitant l'utilisateur à vérifier la qualité ou le contenu du fichier soumis. Cette étape de validation permet d'assurer des résultats cohérents et d'éviter les analyses erronées.



Figure 42 : Résultat d’interface qui distingue entre une plante et un objet

5. Discussions

Les analyses présentées dans les sections précédentes évaluent l’efficacité de notre modèle d’image dans la détection de la fusariose affectant le blé. Après un entraînement approfondi sur **20** cycles, le système a affiché des performances exceptionnelles en termes de justesse et de fiabilité. En phase finale, il a obtenu un taux de réussite de **92,4 %** sur les données d’apprentissage et de **88%** en validation, accompagné d’un score de perte minimal de **0,17**. Ces indicateurs confirment la robustesse de l’algorithme et sa capacité à réduire significativement les marges d’erreur. Grâce à ses couches convolutionnelles et de maxpooling, elle identifie avec finesse les motifs discriminants sur les feuilles, assurant des diagnostics précis. Ces avancées ouvrent des perspectives concrètes pour une aide à la décision agricole, en particulier dans la gestion précoce des infections.

Parallèlement, l’intégration d’un réseau de neurones convolutifs (**CNN**) a renforcé la discrimination entre séquences sains et contaminés. Son aptitude à traiter des schémas complexes se reflète dans un taux de précision atteignant **98,04 %** et **0.11** taux de perte validé par des métriques rigoureuses (**sensibilité, spécificité**). Cette performance souligne la pertinence de notre méthode pour des applications opérationnelles.

Pour en faciliter l’usage, une plateforme interactive a été conçue. Elle permet d’obtenir un diagnostic sur l’état sanitaire d’une séquence nucléiques, avec une transparence totale sur la fiabilité des résultats. Les tests confirment que le système détecte avec une grande acuité les infections liées aux fusarioSES, offrant un outil précieux pour limiter les impacts économiques. Bien que centrée sur le blé, cette méthodologie est transposable à d’autres cultures et pathogènes, élargissant ainsi son champ d’application.

Les modèles d'apprentissage automatique (**Machine Learning**) et d'apprentissage profond (**Deep Learning**) diffèrent principalement par leur complexité et leur approche d'apprentissage. L'apprentissage automatique des algorithmes classiques (comme **Random Forests** ou **Xgboost**) qui nécessitent une extraction manuelle des caractéristiques et fonctionnent bien sur des jeux de données de taille modérée. En revanche, l'apprentissage profond s'appuie sur des réseaux de neurones profonds (comme les **CNN** ou les **Transformers**) capables d'apprendre automatiquement des caractéristiques à partir de données brutes, ce qui le rend bien plus performant pour des tâches complexes (**traitement d'images, langage naturel**) mais exige dénormes quantités de données et une puissance de calcul élevée (**GPU/TPU**). Ainsi, le **ML** est souvent plus adapté pour des problèmes simples avec peu de données, tandis que le **DL** excelle dans les cas complexes nécessitant une analyse approfondie.

Notre approche peut être évaluée sur plusieurs aspects, notamment le choix du modèle et la précision (**Accuracy**). Toutefois, puisque notre travail se concentre spécifiquement sur la prédiction abiotique du blé à partir des séquences **ITS** ou par des images.

Voici le **tableau 07** qui montre une comparaison de notre modèle **CNN** et un modèle d'apprentissage automatique **Random Forest**.

Tableau 7 : Comparaison notre travaille avec une autre méthode

Modèle	Les données utilisées	Précision
CNN	2000 séquences ITS	98%
Random Forest	2000 séquences ITS	87%

Conclusion et Perspective

Conclusion et Perspectives

En conclusion, nous avons combiné des approches de bio-informatique et d'imagerie pour développer un modèle intelligent capable de diagnostiquer les maladies du blé à l'aide de l'apprentissage profond. Notre objectif principal était d'améliorer la précision et la rapidité du diagnostic pour les maladies spécifiques du blé à partir d'images, tout en intégrant des maladies détectées via l'analyse des séquences nucléiques.

Nous avons collecté des données visuelles détaillées sur la santé des plants de blé. Ces données ont permis l'entraînement d'un modèle de réseaux de neurones convolutifs (**CNN**) capable de détecter avec une grande fiabilité, les signes distinctifs des maladies à partir d'images, grâce à L'intégration de l'apprentissage profond qui a appris à distinguer les plantes saines des plantes malades, avec une précision validée par des références établies. Par ailleurs, nous avons enrichi notre approche en y intégrant un large ensemble données séquences nucléiques, afin de prédire le stress abiotiques chez le blé Conçu par la fusariose.

En perspective, nous envisageons d'améliorer notre modèle pour permettre une détection encore plus précoce aux maladies, afin de limiter les pertes agricoles et renforcer la sécurité alimentaire. Les méthodes développées ici ne se limitent pas au blé, elles peuvent être transposées à d'autres plantes et domaines tels que l'agriculture de précision, la pharmacie ou l'industrie et la biologie.

Enfin, l'intégration de nouveaux modèles d'intelligence artificielle pourrait permettre la détection d'un plus grand nombre de maladies, et l'extension de cette approche à d'autres cultures comme le blé, le riz, le seigle ou l'avoine. Nous visions à améliorer l'interface développée dans le cadre de ce projet en proposant une interface intuitive, fluide et plus responsive dans les serveurs du web, compatible aussi bien avec les appareils mobiles que les ordinateurs. Ensuite, l'ajout d'éléments visuels tels que l'affichage de l'image ou de la séquence soumise, des icônes de validation, des barres de confiance, ou des explications visuelles permettrait une meilleure interprétation des résultats. En résumé, ce projet démontre le potentiel remarquable de l'association entre bio-informatique et apprentissage profond pour la gestion intelligente et durable des cultures.

Références

Bibliographiques

Références Bibliographiques

- AgrifoodMarket (no date) Variétés de blé : types et utilisations principales, agrifoodmarket. Available at: <https://www.agrifoodmarket.fr/quelles-sont-les-principales-varietes-de-ble-et-leurs-utilisations/>.
- AgronoBlog (2025) ‘La rouille : Une maladie qui réduit vos rendements agricoles’, AgronoBlog-Blog de agricultura, 14 January. Available at: <https://agronoblog.com/problemes-phytosanitaires/la-rouille-une-maladie-qui-reduit-vos-rendements-agricoles/>.
- Andersen, N.B. (2023) What Is Pandas? (Definition, Features, Tutorial), Built In. Available at: <https://builtin.com/data-science/pandas>.
- Baloch, U.K. (1999) ‘WHEAT: Post-harvest Operations’.
- Battenfield, S.D. *et al.* (2016) ‘Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program’, *The Plant Genome*, 9(2), p. plantgenome2016.01.0005. Available at : <https://doi.org/10.3835/plantgenome2016.01.0005>.
- Bayer (2021) Maladies fongiques-Oidium du blé, Crop Science. Available at: <https://www.cropscience.bayer.dz/fr-dz/cultures/problematique/maladies-fongiques-oidium-du-ble.html>.
- Bility (2023) ‘Définition Visual Studio Code - Bility - Agence de développement web sur-mesure’. Available at: <https://bility.fr/definition-visual-studio-code/>.
- Bonneuil, C. *et al.* (2006) ‘Innover autrement ? La recherche face à l'avènement d'un nouveau régime de production et de régulation des savoirs en génétique végétale’, *Dossiers de l'environnement de l'INRA*, (n° 30), pp. 29–51.
- Borlaug, N.E. (1968) ‘Wheat Breeding and its Impact on World Food Supply’.
- Bouanaka H, Bellil I, Khelifi D (2021b). Multiple methods for varietal resistance assessment of durum wheat cultivars against Fusarium culmorum the causal agent of Fusarium Head Blight Crown Rot in Algeria. *PhysiolMol Plant Path* 115 :101683. Available at: <https://doi.org/10.1016/j.pmpp.2021.101683>
- Bouanaka H, Bellil I, Harrat W, Boussaha S, Benbelkacem A *et al.* (2021) On the biocontrol by Trichoderma afroharzianum against Fusarium culmorum responsible of fusarium head blight and crown rot of wheat in Algeria, *Egypt J Biol Pest Control* 31, 68. Available at: <https://doi.org/10.1186/s41938-021-00416-3>

Références bibliographiques

- Boulal, H. et al. (2007) Guide pratique de la conduite des céréales d'automne (bles et orge) dans le Maghreb (Algérie, Maroc, Tunisie). Institut Technique des Grandes Cultures (ITGC). Available at : <https://hdl.handle.net/20.500.11766/69463>.
- Céline Deluzarche (2024) Définition | Deep Learning - Apprentissage profond. Available at: <https://www.futura-sciences.com/tech/definitions/intelligence-artificielle-deep-learning-17262>.
- Das, S. et al. (2019) ‘DiscoverSL: an R package for multi-omic data driven prediction of synthetic lethality in cancers’, *Bioinformatics*. Edited by R. Schwartz, 35(4), pp. 701–702. Available at: <https://doi.org/10.1093/bioinformatics/bty673>.
- Doussinault, G. et al. (1992) ‘Les céréales à paille : présentation générale’, in. INRA. Available at: <https://hal.inrae.fr/hal-02846465>.
- DZ, edisoft (2021) Agrichem - Votre partenaire agricole. Available at: <https://agrichem.dz>.
- Ephytia (2017) Vigi-Semences - Wheat mosaic virus / Wheat streak mosaic virus. Available at: <https://ephytia.inra.fr/fr/C/25086/Vigi-Semences-Wheat-mosaic-virus-Wheat-streak-mosaic-virus>.
- Ephytia (2018) Vigi-Semences - Xanthomonas translucens pv. translucens (Glume noire des céréales). Available at: <https://ephytia.inra.fr/fr/C/25082/Vigi-Semences-Xanthomonas-translucens-pv-translucens-Glume-noire-des-cereales>.
- Eraslan, G. et al. (2019) ‘Single-cell RNA-seq denoising using a deep count autoencoder’, *Nature Communications*, 10(1), p. 390. Available at: <https://doi.org/10.1038/s41467-018-07931-2>.
- Feillet P., 2000. Le grain de blé composition et utilisation. Ed. INRA. Paris. 308p
- Figueroa, M., Hammond-Kosack, K. E., & Solomon, P. S. (2018). A review of wheat diseases – a field perspective. *Molecular Plant Pathology*, 19(6), 1523-1536 Available at : <https://doi.org/10.1111/mpp.12618>
- Franschläder, T. (2025) *Python : Un guide avec des exemples pratiques*. Available at: <https://www.datacamp.com/tutorial/tqdm-python>.
- Futura, la rédaction (2021) *Définition / Python / Futura Tech, Futura*. Available at: <https://www.futura-sciences.com/tech/definitions/informatique-python-19349>.

Références bibliographiques

- Glossaire HPE (2025) *What is Artificial Intelligence (AI)? / Glossary, Glossaire HPE.* Available at: <https://www.hpe.com/ca/fr/what-is/artificial-intelligence.html>.
- Groupe GEMA – IA School (2025a) ‘Qu’est-ce que Scikit-Learn en machine learning?’, *IA School*. Available at: <https://www.intelligence-artificielle-school.com/ecoole/technologies/quest-ce-que-scikit-learn-en-machine-learning/>.
- Groupe GEMA – IA School (2025b) ‘TensorFlow: tout savoir sur cet outil Open-Source IA School’, *IA School*. Available at: <https://www.intelligence-artificielle-school.com/ecoole/technologies/tensorflow/>.
- Guellier, N. (2017) *Le Blé (Triticum), base de l’alimentation depuis l’Antiquité : plantation, culture, Binette & Jardin*. Available at: <https://jardinage.lemonde.fr/dossier-1415-ble-triticum.html>.
- Haoyang Li *et al.* (2020) *Modern deep learning in bioinformatics / Journal of Molecular Cell Biology* Oxford Academic. Available at: <https://academic.oup.com/jmcb/article/12/11/823/5861537>.
- Henry et De Buyser ... 2001. L'origine des blés. In : Belin. Pour la science (Ed). De la graine à la plante. Ed. Belin. Paris. pp. 69-72.
- IBM (2021) *Qu’est-ce que le machine learning (ML) ? / IBM*. Available at: <https://www.ibm.com/fr-fr/think/topics/machine-learning>.
- INRAE (2018) *De nouvelles perspectives pour l’amélioration variétale du blé et sa culture : le séquençage du génome du blé est aujourd’hui réalisé / INRAE*. Available at: <https://www.inrae.fr/actualites/nouvelles-perspectives-lamelioration-varietale-du-ble-sa-culture-sequencage-du-genome-du-ble-est-aujourd’hui-realise>.
- IWGSC *et al.* (2018) ‘Shifting the limits in wheat research and breeding using a fully annotated reference genome’, *Science*, 361(6403), p. eaar7191. Available at : <https://doi.org/10.1126/science.aar7191>.
- Jaillet, A. (2024) *Google Colab : la force du cloud pour l’apprentissage automatique, DataScientest News*. Available at: <https://datascientest.com/google-colab-tout-savoir>.
- Kaggle: *Wheat Disease Detection*. Available at: <https://www.kaggle.com/code/kushagra3204/wheat-disease-detection/input>
- Lery (1982) *L’agriculture au Maghreb ou pour une agronomie méditerranéenne*. Available at: <https://agsis.fao.org/search/en/providers/122621/records/647396d9ce9437aa76005c83>.

Références bibliographiques

- Marietti, M. (2025) ‘Deep Learning: Tout Comprendre de l’Apprentissage Profond’, *Proactive Academy*, 4 March. Available at: <https://www.proactiveacademy.fr/blog/ia/deep-learning-apprentissage>.
- McCulloch, W.S. and Pitts, W. (1943) ‘A logical calculus of the ideas immanent in nervous activity’, *The bulletin of mathematical biophysics*, 5(4), pp. 115–133. Available at: <https://doi.org/10.1007/BF02478259>.
- Min, S., Lee, B. and Yoon, S. (2016) *Deep learning in bioinformatics / Briefings in Bioinformatics* / Oxford Academic. Available at: <https://academic.oup.com/bib/article/18/5/851/2562808>.
- Poplin, R. et al. (2018) ‘A universal SNP and small-indel variant caller using deep neural networks’, *Nature Biotechnology*, 36(10), pp. 983–987. Available at: <https://doi.org/10.1038/nbt.4235>.
- ‘RAP Grandes cultures 2017’ (no date).
- Ray, D.K. et al. (2019) ‘Climate change has likely already affected global food production’, *PLOS ONE*, 14(5), p. e0217148. Available at : <https://doi.org/10.1371/journal.pone.0217148>.
- Robert, J. (2020) ‘Machine Learning : Définition, fonctionnement, utilisations’, *DataScientest*, 18 November. Available at : <https://datascientest.com/machine-learning-tout-savoir>.
- Shewry, P. R., & Hey, S. J. (2015). The contribution of wheat to human diet and health. *Food and Energy Security*, 4(3), 178-202. Available at: <https://doi.org/10.1002/fes3.64>
- Stryker, cole and kavlakoglu, eda (2024) *Qu'est-ce que l'intelligence artificielle (IA)? / IBM*. Available at: <https://www.ibm.com/fr-fr/think/topics/artificial-intelligence>.
- Tyczewska, A. (2023) *MATE transporter GFD1 cooperates with sugar transporters, mediates carbohydrate partitioning and controls grain-filling duration, grain size and number in rice - Sun - 2023 - Plant Biotechnology Journal - Wiley Online Library*. Available at: <https://onlinelibrary.wiley.com/doi/10.1111/pbi.13976>.
- Wild Code School (2023) *Definition Matplotlib - Qu'est-ce que Matplotlib / Wild Code School, Wild Code School*. Available at: <https://www.wildcodeschool.com/lexique/definition-matplotlib>.

Références bibliographiques

- Wise, T. *et al.* (2019) ‘A computational account of threat-related attentional bias’, *PLOS Computational Biology*. Edited by M. Browning, 15(10), p. e1007341. Available at: <https://doi.org/10.1371/journal.pcbi.1007341>.
- Yamashita, R. *et al.* (2018) ‘Convolutional neural networks: an overview and application in radiology’, *Insights into Imaging*, 9(4), pp. 611–629. Available at: <https://doi.org/10.1007/s13244-018-0639-9>.
- Yu, H., Qi, Y. and Ding, Y. (2022) ‘Deep Learning in RNA Structure Studies’, *Frontiers in Molecular Biosciences*, 9. Available at: <https://doi.org/10.3389/fmolb.2022.869601>.
- Zhou, X. *et al.* (2023) ‘CRISPR-mediated acceleration of wheat improvement: advances and perspectives’, *Journal of Genetics and Genomics*, 50(11), pp. 815–834. Available at : <https://doi.org/10.1016/j.jgg.2023.09.007>.

Présenté par : BAALI Khalil Sid Ali TALBI Abd El Hamid	Date de soutenance : 25/06/2025
Thème : Méthode d'apprentissage profond pour la prédition de stress abiotique causé par la fusariose. Cas d'étude : Le blé	
Nature du diplôme : Master	
Domaine : Science de la nature et de la vie	
Spécialité : Bio-informatique	
Résumé : Cette étude propose une approche innovante combinant l'intelligence artificielle et la bio-informatique pour prédire le stress abiotique causé par la fusariose chez le blé. Nous avons développé deux méthodes distinctes. Une première approche basée sur l'analyse d'images des plantes, et une seconde exploitant des séquences nucléiques, toutes deux utilisant des réseaux de neurones convolutifs (CNN). Les résultats démontrent l'efficacité de ces techniques, avec une précision de 92.4% pour le modèle d'analyse d'images et 99% pour celui utilisant les données nucléiques. Ces performances élevées soulignent le rôle crucial que peuvent jouer l'IA avancée et les outils bio-informatiques dans l'agriculture moderne, notamment pour le diagnostic précoce des pathologies végétales. Nos travaux ouvrent des perspectives prometteuses pour l'intégration de ces technologies dans la gestion optimisée des cultures, et posent les bases pour des recherches futures visant à développer des systèmes d'aide à la décision encore plus performants pour les agriculteurs et toutes les catégories de la société intéressés par ce domaine de la prédition.	
Mots clés : Intelligence Artificiel, prédition, méthode Convolutional Neural Network (CNN), images, bio-informatique, la fusariose, blé, séquences nucléiques, agriculture moderne.	
Jury d'évaluation : ➤ Président : Dr. BENHAMDI Asma (MCA - Université Frères Mentouri, Constantine 1). ➤ Encadreur : Dr. GHERBOUDJ Amira (MCA - Université Frères Mentouri, Constantine 1). ➤ Examinateur : Dr. DAAS Mohamed Skander (MCA - Université Frères Mentouri, Constantine 1).	