

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Département d'Electronique

Mémoire de projet de fin d'études
pour l'obtention du diplôme d'ingénieur d'état en électronique

**Modélisation des paramètres physiques en utilisant des données issues
d'un réseau de capteurs sans fils en agriculture de précision**

MESBAH Djamel Abdenacer
MOUSSAOUI Dhaya Edine

Présenté et soutenu publiquement le (13/07/2021)

Composition du jury :

Mme. Rachida TOUHAMI	Professeur	ENP	Présidente
Mme. Nour El-Houda BENALIA	Docteur	ENP	Promotrice
Mlle. Soumaya Ferhat Taleb	Doctorante	ENP	Co-promotrice
M. Mourad HADDADI	Professeur	ENP	Examineur

ENP 2021

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Ecole Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Département d'Electronique

Mémoire de projet de fin d'études
pour l'obtention du diplôme d'ingénieur d'état en électronique

**Modélisation des paramètres physiques en utilisant des données issues
d'un réseau de capteurs sans fils en agriculture de précision**

MESBAH Djamel Abdenacer

MOUSSAOUI Dhaya Edine

Sous la direction de

Présenté et soutenu publiquement le (13/07/2021)

Composition du jury :

Mme. Rachida TOUHAMI	Professeur	ENP	Présidente
Mme. Nour El-Houda BENALIA	Docteur	ENP	Promotrice
Mlle. Soumaya Ferhat Taleb	Doctorante	ENP	Co-promotrice
M. Mourad HADDADI	Professeur	ENP	Examineur

ENP 2021

Dédicaces

Nous tenons à dédier ce travail à toutes notre famille, à nos amis, ainsi qu'à tous ceux qui nous ont octroyé leurs soutiens.

Remerciements

Nous tenons à exprimer toute notre gratitude à notre promotrice **Dr. Nour el houda BENALIA** et aussi à **Mlle. Soumaya Ferhat Taleb** pour tout le temps qu'elles ont consacré à la direction de ce travail. Nous leur sommes très reconnaissants pour leurs soutien constant, leurs patience et leurs motivation. Nous leur remercions pour tous les précieux conseils qu'elles nous ont apporté, et qui ont contribué à alimenter notre réflexion et nous ont aidé à surmonter bien des obstacles.

Nous remercions également l'ensemble des membres du jury de nous avoir fait l'honneur d'examiner ce travail, tout en leur adressant notre plus profond respect.

Nous souhaitons par la même occasion remercier tous nos enseignants du département des classes préparatoires ainsi que ceux du département d'électronique de **l'École Nationale Polytechnique** pour nous avoir inspiré et guidé vers la réussite au cours des cinq dernières années. Enfin, nous remercions tous les acteurs de l'école qui ont contribué à nous offrir tous les outils nécessaires à la réussite de notre formation d'ingénieur.

Pour finir, Nous remercions infiniment nos familles ainsi que tous nos proches de nous avoir soutenue sans relâche tout au long de cette belle aventure.

ملخص

تواجه الإنسانية تحديات كبيرة، أحد هذه التحديات هو توفير الغذاء لسكان العالم الحاليين والمستقبليين. من المتوقع أن تجلب الزراعة حالاً، ولتحقيق ذلك، يجب تطويرها باستخدام الأدوات التكنولوجية الأكثر تقدماً مع ضمان تنمية مستدامة، وهذا هو سياق مشروعنا إن التطور التكنولوجي الحالي، من خلال إنترنت الأشياء وامتدادها للبيانات الضخمة والذكاء الاصطناعي، يقدم حل للمشاكل الموجودة في الزراعة التقليدية يتكون مشروعنا من بناء نموذج مستمر من بيانات منفصلة من أجهزة استشعار مخصصة للزراعة الدقيقة. سيكون هذا النموذج قادراً على تقدير قيمة المعلمات الفيزيائية في أي نقطة من قطعة الأرض الزراعية دون وجود جهاز الاستشعار، باستخدام خوارزميات انحدار التعلم الآلي.

الكلمات الدالة: الذكاء الاصطناعي، الزراعة الدقيقة، خوارزميات الانحدار، التعلم الآلي.

Abstract

Providing food for the current and future world population is one of the major challenges that is facing humanity today. Agriculture is expected to bring a solution, and to do so, it must develop using the most advanced technological tools while ensuring sustainable development.

The current technological development, through IoT and their extension by big data and artificial intelligence, undeniably offers a solution to the constraints posed by classical agriculture. This is where our work comes in.

Our project consists in building a continuous model from discrete sensor data for precision agriculture. This model will be able to estimate the value of the physical parameters at any point of the agricultural plot without the presence of the sensor, using machine learning regression algorithms.

Keywords : artificial intelligence, precision agriculture, regression algorithms, machine learning.

Résumé

L'humanité fait face à des défis majeurs dont celui de pourvoir une alimentation pour la population mondiale actuelle et future. L'agriculture doit répondre à ces contraintes. Pour se faire, elle doit se développer en utilisant les moyens technologiques les plus avancées tout en assurant un développement durable.

Le développement technologique actuel, à travers les objets connectés (IoT) et leur prolongement constitué par le couple big data et intelligence artificielle, offre indéniablement une solution aux contraintes posées par l'agriculture classique. C'est dans ce contexte que s'inscrit notre travail.

Notre projet consiste à construire un modèle continu à partir des données discrètes des capteurs dédié à l'agriculture de précision. Ce modèle pourra estimer la valeur des paramètres physiques à n'importe quel point de la parcelle agricole sans la présence du capteur, en utilisant des algorithmes de régression de machine learning.

Mots clés : intelligence artificielle, agriculture de précision, algorithmes de régression, machine learning.

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

Introduction générale	11
1 Agriculture de précision	13
1.1 Introduction	14
1.2 Technologies utilisées en agriculture de précision	15
1.2.1 Robots agricoles	15
1.2.1.1 Robots de labourage	16
1.2.1.2 Robots d'analyse du sol	17
1.2.1.3 Robots pour le semis et la transplantation des plantes	18
1.2.1.4 Robots de surveillance des cultures	18
1.2.1.5 Robots pour la récolte	19
1.2.2 Télédétection	21
1.2.2.1 Satellites au service de l'agriculture	22
1.2.2.2 Indices de végétation	28
1.2.2.3 Applications les plus utilisées de la télédétection en agriculture	34
1.2.3 Réseaux de capteurs sans fil (RCSF)	40
1.2.3.1 Types de RCSFs	40
1.2.3.2 Différences entre les types de RCSFs	44
1.2.3.3 Conception d'un réseau de capteurs sans fil pour les applications agricoles	44
1.2.3.4 Applications communes des RCSFs en agriculture	48
1.2.3.5 Perspectives et problèmes des solutions existantes	50
1.3 Conclusion	51
2 Intelligence artificielle pour la prévision de données	53
2.1 Partie 1 : IA et ses utilisations dans l'agriculture de précision	54
2.1.1 Introduction	54
2.1.2 Utilité de l'IA en l'agriculture	55
2.1.3 Utilisation de l'IA pour résoudre les défis de l'agriculture	57

2.1.3.1	Prise de meilleures décisions	57
2.1.3.2	Réduire des coûts	57
2.1.3.3	Réponse aux pénuries de main-d'œuvre	57
2.1.4	Exploitations agricoles et l'adoption de l'IA	57
2.1.4.1	Longueur du processus d'adoption de la technologie	58
2.1.4.2	Manque d'expérience des technologies émergentes	58
2.1.4.3	Questions de confidentialité et de sécurité	58
2.1.5	Combinaison de l'IA avec d'autres technologies	59
2.1.5.1	Prise de décision éclairée sur des données volumineuses	59
2.1.5.2	Capteurs IoT pour la capture et l'analyse des données	59
2.1.6	Réalité et attentes de l'IA pour une agriculture durable	60
2.2	Partie 2 : Régression et classification — Apprentissage automatique supervisé	61
2.2.1	Introduction	61
2.2.2	Types d'algorithmes d'apprentissage supervisé	62
2.2.2.1	Classification	62
2.2.2.2	Régression	63
2.3	Partie 3 : Problématique et algorithmes choisis	68
2.3.1	Problématique et algorithmes utilisés	68
2.3.1.1	Régression linéaire	69
2.3.1.2	Arbres de décision (Decision Tree — DT)	70
2.3.1.3	Forêts aléatoires (Random Forest — RF)	71
2.3.1.4	XGBoost	73
2.4	Conclusion	76
3	Implémentation et résultats	77
3.1	Introduction	78
3.2	Description de la base de données	78
3.3	Critères d'évaluation	79
3.3.1	Erreur moyenne absolue — MAE	80
3.3.2	Erreur quadratique moyenne — RMSE	80
3.3.3	Erreur maximale — Max error	81
3.3.4	Variance expliquée — Explained variance score	81
3.4	Méthodologie de comparaison	81
3.5	Implémentation	82
3.5.1	Logiciels utilisés	82
3.5.2	Régression linéaire multiple	85
3.5.3	Forêts aléatoires	87
3.5.4	Extreme gradient boosting — XGBoost	89
3.5.5	Comparaison entre les trois méthodes	91
3.5.6	Ajustement des hyper paramètres	92
3.5.7	Influence du nombre de données	94
3.6	Conclusion	96
	Conclusion générale et perspectives	98

Liste des tableaux

1.1	Résolutions spatio-temporelles des capteurs satellitaires utilisés pour les applications d'agriculture de précision (AP). Les satellites qui offrent de hautes résolutions spatiales (<30 m) et temporelles sont plus adaptés à l'agriculture de précision	27
1.2	Quelques indices de végétation récemment utilisés pour des applications de télédétection dans l'agriculture de précision	33
1.3	Différences entre TWSNs et WUSNs.	44
1.4	Comparaison des plateformes de capteurs sans fil existantes	46
3.1	Résumé de quelques données statistiques du data set	79
3.2	Critère d'évaluation pour la régression linéaire multiple	85
3.3	Critère d'évaluation pour les Forêts aléatoires	87
3.4	Critère d'évaluation pour Extreme gradient boosting	90
3.5	Précision des différents modèles dans la phase de validation	91

Table des figures

1.1	Robot tracteur autonome	16
1.2	Robot Smartcore.	17
1.3	Robot TerraSentia.	19
1.4	Robocrop — Le robot guidé par des capteurs et des caméras 3D est actuellement en mesure de cueillir des framboises	21
1.5	Réflectance de l'eau, du sol et de la végétation dans différentes lon- gueurs d'onde.	22
1.7	RCSF typique déployé sur champ pour des applications agricoles . . .	41
1.8	Application agricole basée sur des réseaux de capteurs souterrains. . . .	43
1.9	Architecture d'un nœud de capteurs multi-puces embarqué typique . .	45
2.1	Rôle de l'IA dans le cycle de gestion de l'information agricole	56
2.2	Les Différents types d'apprentissages automatiques.	62
2.3	Diagramme de fonctionnement de la forêt aléatoire	73
2.4	Diagramme de fonctionnement de XGBoost.	76
3.1	Diagramme explicative du processus de traitement de données	83
3.2	Emplacement des capteurs	84
3.3	Matrice de corrélation	84
3.4	Histogramme de l'importance des caractéristiques pour la régression linéaire.	85
3.5	Distribution de la température sur la surface d'intérêt pour la régression linéaire.	86
3.6	Histogramme de l'importance des caractéristiques pour les forêts aléatoires.	88
3.7	Distribution de la température sur la surface d'intérêt pour les forêts aléatoires	88
3.8	Histogramme de l'importance des caractéristiques pour XGBoost	90
3.9	Distribution de la température sur la surface d'intérêt pour XGBoost .	91
3.10	Diagramme explicatif de la validation croisée à k-blocs	93
3.11	Courbe d'apprentissage pour la régression linéaire	95
3.12	Courbe d'apprentissage pour les forêts aléatoires	95
3.13	Courbe d'apprentissage pour XGBoost	96

Liste des abréviations

IOT : Internet of Things

IA : intelligence artificielle

AP : Agriculture de Précision

IR : infrarouge

NIR : near -infrared

LAI : Leaf Area Index

ET : évapotranspiration

S-SEBI : Simplified Surface Energy Balance Index

SSEBop : operational Simplified Surface Energy Balance

SEBAL : Surface Energy Balance Algorithm for Land

RVB : Rouge Vert Bleu

SDI : spectral disease indices — indices spectraux de maladie

RC : Réseau de capteurs

RCSF : RC sans fil

TWSN : RCSF terrestre

WUSN : RCSF sous terrain

MEMS : Microelectromechanical systems

RF : radio-fréquence

GUI : Graphical User Interface

ML : Machine learning

LASSO : Least Absolute Shrinkage and Selection Operator

DT : Decision Tree

RF : Random Forest

K - NN : K - nearest Neighbours

SVM : Support Vector Machine

SVR : Support Vector Regression

PCR : Principal Component Regression

OOB : Out - of - bag

XGBoost : extreme gradient boosting

Introduction générale

L'agriculture a joué un rôle clé dans le développement de la civilisation humaine.[1] L'agriculture moderne requiert une forte croissance de production pour accommoder la demande croissante en nourriture du fait d'une démographie élevée. la population mondiale est estimée à 9 milliards d'individus en 2050, nécessitant une augmentation de la production de l'ordre de 60% par rapport à la production actuelle [2].

De plus, les changements climatiques récents sont alarmants et affectent négativement la productivité des terres agricoles [3], ainsi que la pénurie d'eau [4], et les conséquences négatives sur l'environnement de l'application anarchique des fertilisants , nécessitent des solutions permettant une prise de décision intelligente. Parmi ces solutions on trouve la télédétection, la robotique et l'automatisation, les réseaux de capteurs sans fil, la technologies à taux variables, le GPS qui ont connues de plus en plus d'applications pratiques dans le domaine de l'agriculture.

L'utilisation des réseaux de capteurs (RCs) et leurs déclinaisons en objets connectés (IoT) dans l'assistance des pratiques agricoles est dans une direction très positive [5]. La réduction de la taille des capteurs et la disponibilité de capteurs à des prix abordables ont permis des applications très prometteuses, particulièrement dans le contexte du big data et de l'intelligence artificielle. Permettant ainsi une nouvelle vague de révolution en agriculture de précision, communément appelée agriculture 4.0.

Parmi les systèmes de surveillance existants, on trouve la technologie des Réseaux de Capteurs Sans Fils (RCSF). Ce réseau peut offrir un support important qui permettra la gestion précise des ressources, l'optimisation de l'irrigation et de l'utilisation des engrais, le suivi/prévention des maladies, la prédiction du moment adéquat de la

récolte, autrement dit : il peut améliorer la production agricole grâce aux informations fournies par les éléments qui composent ce réseau.

Vu que le nombre de capteurs utilisés dans un RCSF doit être le minimum possible pour des raisons financières. L'existence des points non échantillonnés par ces capteurs est une réalité.

Dans ce mémoire, notre but est de construire un modèle continu à partir des données discrètes des capteurs. Ce modèle pourra estimer la valeur des paramètres physiques en n'importe quel point de la parcelle agricole sans la présence du capteur.

On commencera donc par définir l'agriculture de précision, suivie d'une vue d'ensemble des différentes technologies utilisées dans l'agriculture de précision ainsi que certaines recherches dans ce domaine, par la suite on passera aux détails sur le rôle de l'intelligence artificielle dans l'agriculture de précision ainsi que pour la prévision de données.

On procédera par la suite à détailler des méthodes de modélisation de paramètres physiques du sol de la surface d'intérêt ou de l'environnement local, qui permettrait d'extrapoler ces données discrètes sur toute la surface d'intérêt.

On finalise notre travail avec une conclusion et quelques perspectives.

Chapitre 1

Agriculture de précision

1.1 Introduction

L'agriculture est un processus par lequel les êtres humains gèrent leurs écosystèmes et contrôlent le cycle biologique des espèces domestiquées afin de produire des aliments et d'autres ressources utiles à leurs sociétés. Elle ne désigne pas que l'ensemble des compétences et des activités liées à la culture du sol, mais elle désigne aussi l'ensemble des travaux effectués sur le milieu naturel (pas seulement sur la terre) pour cultiver et récolter des êtres vivants (plantes, animaux, voire champignons ou microbes) utiles à l'homme[6].

Dans [7], l'agriculture des temps modernes a connu quatre révolutions qui ont permis d'augmenter de façon significative la production et la qualité des produits agricoles :

- L'introduction du premier tracteur par l'entreprise Waterloo Gasoline Engine en 1893 a marqué la première révolution, cette évolution technologique a permis une croissance significative sur les méthodes utilisées jusqu'alors basées sur l'utilisation des animaux.
- La deuxième révolution est venue de découvertes en chimie, permettant d'utiliser des produits nutritifs pour la culture, ainsi que de la protéger des dangers naturels avec l'introduction des pesticides.
- Les travaux de Norman Borlaug sur la modification génétique des plantes ont conduit à une importante amélioration du rendement en agriculture dans plusieurs régions du monde, évitant la mort de millions de personnes à cause de la famine.
- La quatrième révolution dans l'agriculture moderne concerne l'AP, qu'on abordera dans le reste de ce chapitre.

D'après [8], l'AP est le concept de gestion de l'agriculture moderne qui utilise des techniques numériques pour contrôler et optimiser les processus de production agricole.

D'après [7], elle est un concept de gestion de parcelles agricoles, fondé sur le constat de l'existence de variabilité intra-parcellaires.

1.2 Technologies utilisées en agriculture de précision

A travers les avancées ces dernières années, plusieurs technologies ont été introduites dans le domaine de l'AP. Tout au long de cette section, on s'intéressera aux technologies les plus prépondérantes à savoir : les robots, la télédétection ainsi que les réseaux de capteurs sans fil.

Les définitions de ces technologies seront accompagnés d'exemples de travaux déjà implémentés dans des processus agricoles.

1.2.1 Robots agricoles

Un robot est un dispositif mécatronique (alliant mécanique, électronique et informatique) conçu pour accomplir automatiquement des tâches imitant ou reproduisant, dans un domaine précis, des actions humaines. La conception de ces systèmes est l'objet d'une discipline scientifique, branche de l'automatisme nommé robotique.

Un robot agricole est un robot conçu pour accomplir certaines tâches dans le domaine de l'agriculture et de l'horticulture. Le principal domaine d'application des robots en agriculture est celui de la récolte. L'objectif principal de ces robots est donc de minimiser la main-d'œuvre et l'interaction directe avec la surface agricole.

La première commercialisation d'un robot agricole revient à Oz, une invention de la société toulousaine Naïo-Technologies. Ce robot peut désherber tous les rangs sans intervention humaine, à condition que l'espacement inter-rang ne soit pas trop large et qu'il ait en bout de rang de quoi faire demi-tour et s'engager dans la rangée suivante. Le robot peut aussi être guidé par l'utilisateur à l'aide d'une télécommande sans fil.

Ce robot haut de 60 cm et large de 40 cm a été conçu en 2013 et a depuis connu des successeurs plus avancés techniquement [9].

1.2.1.1 Robots de labourage

Le labour est une façon de travailler la terre. Avec une bêche, une charrue, un attelage ou un tracteur, le labour consiste à ouvrir la terre arable jusqu'à une certaine profondeur et à la retourner. La technique du labour ameublisse le sol et enfouit ce qu'il porte en surface. Dans les domaines de l'agriculture et de l'agronomie, cette technique est utilisée pour préparer le futur ensemencement du sol [10].

Exemple de ce types des robots Les auteurs de "Agricultural Robot for automatic ploughing and seeding" ont pu fabriquer un robot qui fait le labourage automatique, la distribution de graines mais aussi la cueillette de fruits et la pulvérisation de pesticides, le composant principal dont est constitué le robot est le mega microcontrôleur AVR At. La navigation se fait à l'aide d'un capteur ultrasonique qui envoie des données en continuité au microcontrôleur [11].



FIGURE 1.1: Robot tracteur autonome

D'après certains travaux, l'utilisation de robots de petites tailles aiderait à réduire la fréquence de labourage nécessaire à la surface agricole. il est aussi possible pour les robots de percer et placer les graines ou les semis sans avoir à labourer le sol, ou

bien optimiser le processus du labourage en faisant des micro-labourages autour de l'endroit du semis, qui a le plus d'influence sur la croissance de la plante.

1.2.1.2 Robots d'analyse du sol

Analyse du sol le labourage est ensuite suivi de l'analyse du sol, opération consistant à mesurer différentes propriétés physiques et chimiques du sol dans le but d'évaluer sa fertilité et ses conditions physiques. Dans certaines conditions, l'analyse du sol est faite à différentes étapes de la croissance de la plante pour évaluer la quantité et la composition des nutriments à appliquer [12].

Exemple de ce types des robots «Smartcore» est un robot autonome développé par une paire de diplômés du Purdue University College of Engineering, il est conçu pour collecter des échantillons de sol dans les champs et les amener au bord du champ pour l'expédition au laboratoire [13].



FIGURE 1.2: Robot Smartcore.

1.2.1.3 Robots pour le semis et la transplantation des plantes

Semis Le semis est une opération culturale qui consiste à mettre en terre les graines ou semences que ce soit dans un champ ou une surface de petite dimension. Le semis peut se faire à la main, ou de manière mécanisée à l'aide de semoirs. Le repiquage, appelé aussi **transplantation** ou replantation est une technique qui consiste à déplanter un végétal (appelé en cette circonstance transplant) et à le replanter dans un autre substrat de culture ou un autre endroit.

Exemple de ce types des robots Un robot autonome (agribot) a été développé par [14] pour effectuer la tâche complexe d'ensemencement pour quatre cultures : coton, maïs, soja et le blé. le setup expérimental est constitué d'un bloc principal composé d'une board ARM7 avec le LPC2148 comme microcontrôleur. Le réservoir de semences et le réservoir d'eau sont connecté pour le stockage des graines et de l'eau respectivement. L'entrée du type de culture est donnée manuellement en sélectionnant l'un des quatre commutateurs d'entrée. La distance en ligne et colonnes requises pour ces cultures est modélisée dans le système. Un moteur DC est connecté pour que le mécanisme de roue en rotation fasse tomber les graines. Une Pompe à eau DC submersible entièrement étanche est utilisée pour verser l'eau. Les capteurs IR sont connectés à proximité du réservoir de semences pour vérifier s'il est vide avant de commencer l'opération de semis. L'autre paire de capteurs IR est connectée à l'avant du véhicule pour détecter l'extrémité du champ. La navigation à l'aide de capteurs IR rend le robot moins volumineux que ses alternatives [15].

1.2.1.4 Robots de surveillance des cultures

Surveillance des cultures Les sélectionneurs de cultures entreprennent de grandes expériences comparant des milliers de variétés de plantes sur des centaines d'acres pour mesurer des traits clés tels que l'émergence et la hauteur des plantes. La majorité de cette opération est effectuée à la main, elle est considérée comme un processus extrêmement coûteux et chronophage qui peut parfois être imprécis et même inadéquat en raison de contraintes humaines [16].

Exemple de ce types des robots Une équipe de chercheurs de l'université d'Illinois ont développé un robot entièrement automatisé capable de surveiller les cultures sur le terrain pendant les périodes de croissance dans le but d'aider les sélectionneurs dans la vaste tâche de développement et de comparaison des cultivars de plantes. Le robot de 13 pouces de large appelé TerraSentia est transportable et autonome. Il est capable de se déplacer librement entre les rangées de cultures et de capturer l'information des plantes à l'aide d'une gamme de capteurs, d'algorithmes et de deep learning. Ce robot a appris à compter les plantes de maïs en utilisant seulement 300 images [17].



FIGURE 1.3: Robot TerraSentia.

1.2.1.5 Robots pour la récolte

La récolte désigne l'ensemble des travaux agricoles permettant à l'homme de collecter les plantes, fruits et légumes (objet d'une culture, le plus souvent) par cueillette, fauchage, arrachage, et de les transporter à l'abri pour stockage plus ou moins provisoire (dans un grenier, un hangar, une cave, un silo) avant consommation ou transformation (familiale, artisanale ou industrielle).

La récolte ne concerne pas que des produits destinés à l'alimentation. On parle, par exemple, d'une récolte de lin ou de coton. On parle aussi de « récolte » à propos d'autres produits naturels, par exemple le miel des ruches ou le sel des marais-salants.

Un robot récolteur est particulièrement complexe à mettre en œuvre car il doit détecter la position, la taille et la forme de la cible, doit être mobile pour pouvoir l'atteindre et doit avoir un mécanisme robotisé (à l'aide d'un bras mécanique par exemple) pour la cueillette.

Des solutions classiques, telles que les récolteuses et les moissonneuses batteuses ont déjà fait leur preuve, néanmoins elles ont besoin d'un pilote lors du processus.

Des chercheurs [18] se sont concentrés sur la récolte des fraises et ont développé un robot de récolte à deux bras à faible coût. Leur système était plus résistant aux variations d'éclairage en raison de la modélisation de la couleur en fonction de l'intensité lumineuse. Afin de faire face aux occlusions et à l'encombrement, le robot peut utiliser la pince pour écarter les obstacles environnants afin de cueillir des fraises situées en grappes. Le taux de réussite de la cueillette variait de 20% pour les scénarios les plus complexes avec une fraise mûre dans une grappe de fruits non mûrs à 100% pour les situations avec une fraise mûre isolée. En mode à deux bras, le système avait un temps de cycle de 4,6 s. Pour la récolte de tomates, un robot à deux bras utilisant un capteur de vision binoculaire pour la détection et la localisation des fruits a été développé. Dans une configuration expérimentale grandement simplifiée, le taux de réussite de ce robot était de 87,5% [19].

L'état de l'art dans les robots de récolte peut être lu dans [19].



FIGURE 1.4: Robocrop — Le robot guidé par des capteurs et des caméras 3D est actuellement en mesure de cueillir des framboises

1.2.2 Télédétection

Le recours à la télédétection dans le domaine agricole présente un précieux avantage : recueillir rapidement des informations fiables sur l'état des cultures tout en faisant ressortir la variabilité spatiale et temporelle de cet état. Ces informations sont utilisées comme indicateurs pour moduler un certain nombre d'opérations culturales et permettent d'établir un diagnostic des parcelles, par exemple : la comparaison inter-parcellaire, pour guider les décisions de gestion de production et d'optimisation de l'exploitation agricole et on peut noter aussi l'analyse intra-parcellaire pour optimiser l'utilisation des apports aux cultures (semences, engrais, eau, etc.).

L'utilisation de la télédétection dans l'AP permet la gestion et le suivi des pratiques agricoles à différentes échelles : nationale, régionale, locale et parcellaire. Cette dynamique permet de répondre à des problématiques diverses avec la précision adéquate : connaissance de la distribution spatiale des territoires agricoles, évolution de l'utilisation des terres agricoles, analyse de l'efficacité de mise en place de politiques agricoles, gestion de production et d'optimisation de l'exploitation, gestion spécialisée des interventions techniques et optimisation de la logistique.

La télédétection consiste à obtenir des informations sur l'objet de mesure sans avoir à être en contact directe avec le dit objet. Le porteur d'information, dans ce cas, est l'onde électromagnétique à différentes fréquences [7].

Selon les domaines dont lesquels on se trouve la télédétection, on ne s'intéressera pas aux mêmes gammes de longueurs d'ondes, comme montre la figure ci-dessous :

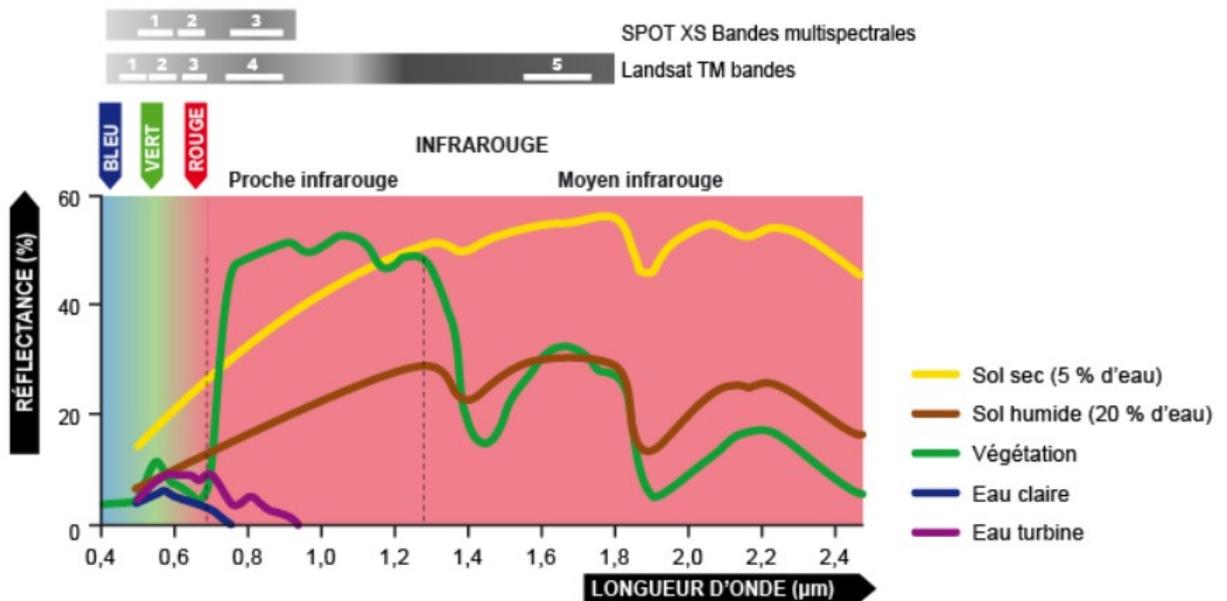


FIGURE 1.5: Réflectance de l'eau, du sol et de la végétation dans différentes longueurs d'onde

Dans le domaine agricole, on va plus particulièrement s'intéresser à la biomasse. Pour cela, on analyse la courbe de réflectance de la végétation, dans les gammes de longueurs d'ondes d'intérêt que sont le rouge (440-520 nm), le vert (510-600 nm), le bleu (630-690 nm) et le proche infra-rouge (760-860 nm). Celles-ci permettent, en les combinant, de former une image dite multi-spectrale, qui peut être utilisée de différentes manières pour aider à caractériser la quantité de biomasse qui a une activité photosynthétique importante.

1.2.2.1 Satellites au service de l'agriculture

Les chercheurs reconnaissent depuis longtemps la nécessité de cartographier les bases de données sur les sols et l'utilisation des terres pour la gestion durable des

ressources naturelles à l'échelle locale, régionale et nationale. La connaissance des propriétés physiques, biologiques et chimiques des sols est importante pour concevoir et mettre en œuvre des stratégies d'irrigation, de drainage, de gestion des nutriments et d'autres cultures, qui sont des composantes essentielles de l'AP. De même, la cartographie de l'utilisation des sols peut aider à évaluer les impacts de la gestion et des politiques existantes à l'échelle régionale ou nationale. L'approche traditionnelle de l'utilisation des techniques de télédétection en agriculture existait déjà avant 1958, date à laquelle le terme "télédétection" a été introduit pour la première fois. Par exemple, la photographie aérienne a été utilisée pour cartographier les sols, l'utilisation des terres et l'état des cultures aux États-Unis dans les années 1930 et 1940. Cependant, ces méthodes conventionnelles de cartographie des sols et de classification de l'utilisation des terres (par exemple, la photographie à basse altitude et les équipes au sol) impliquent généralement un important travail sur le terrain et des analyses en laboratoire, qui sont coûteuses et prennent beaucoup de temps [20].

L'avènement de la télédétection par satellite au cours des dernières années a facilité une cartographie plus efficace et plus rationnelle de l'utilisation et de la couverture des sols à l'échelle régionale, nationale et mondiale [21].

Le lancement de Vanguard 2 et TIROS 1 en 1959 et 1960, respectivement, a marqué le début de la télédétection par satellite pour les informations météorologiques. Cependant, l'ère de la télédétection par satellite pour l'agriculture a commencé avec le lancement de Landsat 1 (anciennement connu sous le nom de Earth Resources Technology Satellite-ERTS) le 23 juillet 1972 par la National Aeronautics and Space Administration (NASA).

Après Landsat 1, une série de satellites Landsat (Landsat 2-8) ont été lancés pour fournir des images de haute qualité aux chercheurs, aux gestionnaires des terres et aux décideurs politiques afin de faciliter la gestion des ressources naturelles dans le monde entier. Les images acquises par Landsat ont été utilisées pour la classification de l'utilisation des terres, la classification des cultures, la surveillance et l'estimation des besoins en eau d'irrigation dans de nombreuses régions du monde. Plus tard, en 1984, le cartographe thématique Landsat 5 a été lancé pour recueillir des images à

plus haute résolution (30 m) dans un plus grand nombre de bandes dans les régions du visible et du proche infrarouge. Actuellement, l'USGS-NASA prévoit de lancer Landsat 9 (résolution de 30 m, 100 m) à la mi-2021. En 1986 et 1988, la France et l'Inde ont également lancé les satellites SPOT 1 et IRS-1A, respectivement [22].

Satellite (années d'activité)	Capteur (résolution spatiale)	Résolution temporelle	Application dans l'agriculture de précision
Landsat 1 (1972–1978)	MS (80 m)	18 jours	Croissance des cultures
AVHRR (1979-présent)	MS (1.1 km)	1 jour	Gestion des nutriments
Landsat 5 TM (1984-2013)	MS et Thermal (60 m–Landsat 7,		
Landsat 7 (1999-présent)	100 m–Landsat 8,	16 jours	Biomasse; rendement des cultures;
Landsat 8 (2013-présent)	120 m–Landsat 5)		croissance des cultures.
SPOT 1 (1986–1990)	MS (20 m)	2-6 jours	Gestion de l'eau
SPOT-2 (1990–2009)			
IRS 1A (1988–1996)	MS (72 m)	22 jours	Gestion de l'eau, gestion des nutriments
LiDAR (1995)	VIS (10 cm)	N/A	Topographie, gestion des nutriments
RadarSAT (1995–2013)	C-band SAR (30 m)	1-6 jours	Croissance des cultures
IKONOS (1999–2015)	MS (3.2 m)	3 jours	Crop yield; soil properties; nutrient management; ET estimation
EO-1 Hyperion (2000–2017)	HS (30 m)	16 jours	Maladie
Terra/Aqua MODIS (Terra-1999-présent, Aqua-2002-présent)	MS (SpectroRadiomètre; 250-1000 m)	1-2 jours	Rendement des cultures; la croissance des cultures.
Terra-ASTER (2000–présent)	MS et Thermal (15 m–V, NIR, 30 m–SWIR, 90 m–TIR)	16 jours	Gestion de l'eau
QuickBird (2001–2014)	MS (2.44 m)	1-3.5 jours	Maladie

Satellite (années d'activité)	Capteur (résolution spatiale)	Résolution temporelle	Application dans l'agriculture de précision
AQUA AMSR-E (2002–2016)	MS (Radiomètre à micro-ondes ; 5,4 km-56 km)	1-2 jours	Gestion de l'eau
Spot-5 (2002–2015)	MS (V, NIR–10 m, SWIR–20 m)	2–3 jours	Rendement des cultures
ResourceSat-1 (2003–2013)	MS (5.6m–V, 23.5 m–SWIR)	5 jours	Gestion des nutriments
KOMPSAT-2 (2006–present)	MS (4 m)	5.5 jours	Rendement des cultures
Radarsat-2	C-band SAR (1–100 m)	3 jours	LAI et biomasse
RapidEye (2008–présent)	MS (6.5 m)	1–5.5 jours	la gestion de l'eau ; le rendement des cultures ; croissance des cultures et chlorophylle.
GeoEye-1 (2008–présent)	MS (1.65 m)	2.1-8.3 jours	Gestion des nutriments
WorldView-2 (2009–présent)	MS (1.4 m)	1.1 jours	Croissance des cultures
Pléiades-1A (2011-présent) Pléiades-1B (depuis 2012)	MS (2 m)	1 jour	Croissance des cultures
VIIRS Suomi-NPP (2011-présent)	MS (IR Radiometer, 375 m and 750 m)	16 jours	Gestion des cultures (NDVI)
VIIRS-JPSS-1 (2017-présent)			
KOMPSAT-3 (2012-présent)	MS (2.8 m)	1.4 jours	Croissance des cultures
Spot-6 (2012-présent), Spot-7 (depuis 2014)	MS (6 m)	1 jour	Maladie
SkySat-1 (2013-présent) SkySat-2 (depuis 2014)	MS (1 m)	sous-quotidiennement	Croissance des cultures

Satellite (années d'activité)	Capteur (résolution spatiale)	Résolution temporelle	Application dans l'agriculture de précision
Worldview-3 (2014-présent)	SS (1.24 m)	<1 jour	Croissance des cultures; gestion des mauvaises herbes.
Sentinel-1 (2014-présent)	C-band SAR (5-40 m)	1-3 jours	Croissance des cultures
Sentinel-2 (2015-présent)	MS (10 m-V and NIR, 20 m-Red edge et SWIR, 60 m-2 NIR)	2-5 jours	Rendement; Gestion de l'azote
KOMPSAT-3A (2015-présent)	MS (V NIR-2.2 m, SWIR-5.5 m)	1.4 jours	Maladie
SMAP (2015-présent)	L-band SAR (1-3 km) et radiometer (40 km)	2-3 jours	Rendement des culture; gestion de l'eau.
TripleSat (2015-présent)	MS (3.2 m)	1 jour	Croissance des cultures
ECOSTRESS-PHYTIR (2018-présent)	Thermique (38 x 69 m)	1-5 jours	ET

TABLE 1.1: Résolutions spatio-temporelles des capteurs satellitaires utilisés pour les applications d'agriculture de précision (AP). Les satellites qui offrent de hautes résolutions spatiales (<30 m) et temporelles sont plus adaptés à l'agriculture de précision

1.2.2.2 Indices de végétation

Les indices de végétation sont des expressions mathématiques qui combinent la réflectance mesurée dans de nombreuses bandes spectrales pour produire une valeur qui aide à évaluer la croissance et la vigueur des cultures (*Tableau 1.2*), ainsi que plusieurs autres propriétés de la végétation comme la biomasse et la teneur en chlorophylle. La cartographie de ces indices peut aider à comprendre la variabilité spatio-temporelle des conditions de culture, ce qui est crucial pour les applications d'AP [23].

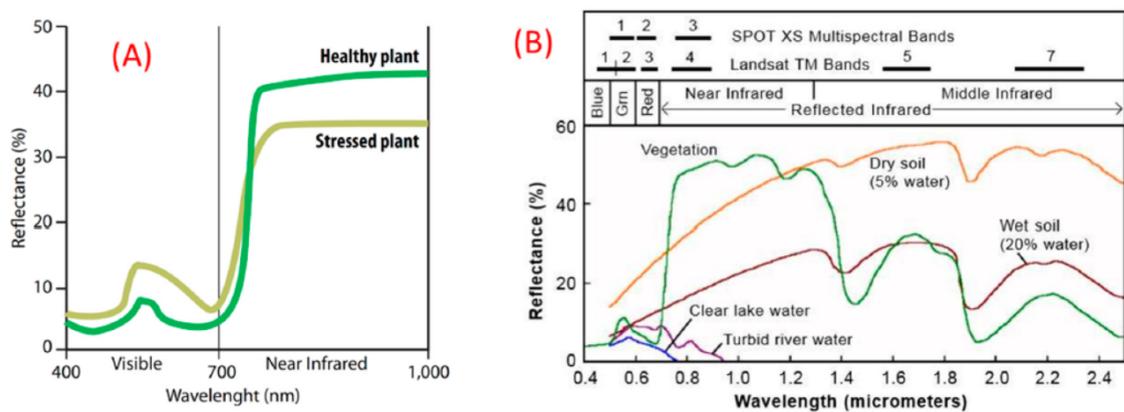


FIGURE 1.6: Spectre de réflectance typique (A) d'une plante saine et d'une plante stressée et (B) du sol, de l'eau et de la végétation

Les indices de végétation les plus utilisés, tels que l'indice de végétation de densité normalisée (NDVI), le NDVI vert (GNDVI) et l'indice de végétation ajusté au sol (SAVI), utilisent le fait que dans la gamme visible du spectre, la réflectance des plantes est faible dans les régions bleue et rouge, tandis qu'elle atteint son maximum dans la région verte (figure 1.7). Les pigments végétaux, principalement la chlorophylle et les caroténoïdes, s'adsorbent fortement dans la partie visible du spectre, sauf dans la région verte. Cependant, une adsorption aussi forte ne se produit pas dans la partie NIR du spectre, ce qui entraîne une réflectance élevée dans la région NIR des plantes vertes et saines (Figure 1.7). Le NDVI utilise les valeurs de réflectance mesurées dans les régions rouge et NIR pour fournir des informations précieuses sur la croissance des cultures (LAI, biomasse), la vigueur et la photosynthèse. La valeur du NDVI varie de -1 à 1, où les valeurs positives indiquent une verdure croissante, et les valeurs

négatives indiquent des surfaces non végétalisées telles que les zones urbaines, les sols/terres nus, l'eau et la glace [24].

Des facteurs externes aux conditions de végétation, tels que la géométrie solaire et la géométrie d'observation, le sol et les résidus de culture à la surface du sol, ainsi que l'atmosphère, sont également pris en compte. sur la surface du sol, et les effets atmosphériques peuvent causer des interférences dans les signaux spectraux. Le NDVI est sensible aux effets de confusion causés par les sols, l'atmosphère, les nuages et l'ombre du couvert végétal. qui peuvent entraîner des informations erronées sur l'état des cultures ou des plantes. De plus, le NDVI est également connu pour être insensible aux changements de LAI et de biomasse après avoir atteint un seuil (saturation), en particulier dans des conditions de végétation dense [25].

Indice	Définition/équation	Applications (Références)
Indice de végétation par différence normalisée (NDVI)	$\frac{R_{NIR} - R_{red}}{R_{NIR} + R_{red}}$	Biomasse ; sélection, phénotypage ; rendement ; maladies ; gestion de l'eau ; humidité du sol ; stress hydrique.
NDVI vert (GNDVI)	$\frac{R_{NIR} - R_{green}}{R_{NIR} + R_{green}}$	Stress hydrique ; rendement ; biomasse ; maladie
Différence normalisée du bord rouge (NDRE)	$\frac{R_{NIR} - R_{red\ edge}}{R_{NIR} + R_{red\ edge}}$	Rendement des cultures et biomasse ; gestion de l'azote ; maladies .
Indice de végétation par différence normalisée du bord rouge (RENDVI)	$\frac{R_{NIR} - R_{red\ edge}}{R_{NIR} + R_{red\ edge}}$	Rendement ; gestion de l'irrigation ; état et application de l'azot ; maladie
Indice de végétation ajusté au sol (SAVI)	$\frac{(R_{NIR} - R_{red})(1+L)}{R_{NIR} + R_{red} + L}$	Rendement ; biomasse ; maladie ; concentration et absorption de N l'absorption de l'azote ; stress hydrique.
Indice de végétation ajusté au sol modifié (MSAVI)	$\frac{2R_{NIR} + 1 - \sqrt{(2R_{NIR} + 1)^2 - 8(R_{NIR} - R_{red})}}{2}$	Biomasse ; rendement des cultures ; absorption de l'azote ; teneur en chlorophylle

Indice	Définition/équation	Applications (Références)
Indice de végétation par différence normalisée (RDVI)	$\frac{R_{NIR} - R_{red}}{\sqrt{R_{NIR} + R_{red}}}$	Rendement des cultures; absorption de l'azote; humidité du sol; biomasse.
Indice de végétation à large gamme dynamique (WDRVI)	$\frac{\alpha R_{NIR} - R_{red}}{\alpha R_{NIR} + R_{red}}$	Application d'azote, rendement; croissance de la culture LAI; maladies
Indice de végétation résistant à l'atmosphère (ARVI)	$\frac{R_{NIR} - R_{RedBlue}}{R_{NIR} + R_{RedBlue}}$	Maladie; cartographie des mauvaises herbes.
Indice de végétation résistant aux effets de l'atmosphère (IAVI)	$\frac{R_{NIR} - (R_{red} - \lambda(R_{blue} - R_{red}))}{R_{NIR} + (R_{red} - \lambda(R_{blue} - R_{red}))}$	Rendement des cultures
Rapport de l'indice de végétation (RVI)	$\frac{R_{NIR}}{R_{red}}$	Rendement des cultures; biomasse.
Indice de végétation de Dierence (DVI)	$R_{NIR} - R_{red}$	Maladie; rendement des cultures; LAI.
Indice végétatif transformé ajusté au sol (TSAVI)	$\frac{a(R_{NIR} - aR_{red} - b)}{R_{red} + aR_{NIR} - ab}$	Stress hydrique; rendement des cultures.
Indice de réflectance de la sénescence des plantes (PSRI)	$\frac{R_{680} - R_{550}}{R_{750}}$	Maladie; rendement; biomasse.
Indice du rapport normalisé pigments-chlorophylle (NPCl)	$\frac{R_{680} - R_{430}}{R_{680} + R_{430}}$	Stress hydrique

Indice	Définition/équation	Applications (Références)
Indice du rapport d'absorption de la chlorophylle (CARI)	$\frac{R_{700}}{R_{670}} \times \frac{aR_{670} + bR_{670}}{\sqrt{a^2 + 1}}$	Teneur en chlorophylle
Indice de végétation chlorophyllienne (CVI)	$\frac{R_{NIR}}{R_{Green}} \times \frac{R_{Red}}{R_{Green}}$	Rendement des cultures; croissance des cultures croissance - teneur en chlorophylle; rendement
Indice de chlorophylle (CI)	$\frac{R_{NIR}}{R_{RedEdge}} - 1$	Chlorophylle et teneur en Azote
Indice de végétation ajusté au sol optimisé (OSAVI)	$\frac{1.16(R_{NIR} - R_{Red})}{R_{NIR} + R_{Red} + 0.16}$	Maladie; rendement des cultures; biomasse, absorption de l'azote; humidité du sol; stress hydrique .
Indice de réflectance photochimique (PRI)	$\frac{R_{531} - R_{570}}{R_{531} + R_{570}}$	maladie; stress hydrique foliaire (PRInorm), température de la canopée et rendement (PRI550); stress hydrique (PRI, PRI550-515, PRInorm .
Indice du bilan hydrique	$\frac{R_{1500} - R_{531}}{R_{1500} + R_{531}}$	Programmation de l'irrigation
Différence normalisée de teneur en eau (NDWI)	NDWI	Teneur en eau de la végétation
Indice de stress hydrique par infrarouge à ondes courtes (SIWSI)	$\frac{R_{858.5} - R_{1640}}{R_{858.5} + R_{1640}}$	Teneur en eau des feuilles (stress hydrique);

Indice	Définition/équation	Applications (Références)
Degrés au-dessus de la canopée non stressée (DANS)	$\min(0, T_c - T_{cNonStressed})$	Stress hydrique; ET.
Degrés au-dessus du seuil de canopée (DACT)	$\max(0, T_c - T_{critical})$	ET; stress hydrique.
Indice de végétation triangulaire (TVI)	$0.5[120(R_{NIR} - R_{Green}) - 200(R_{Red} - R_{Green})]$	Maladie
Point d'inflexion du bord rouge (REIP)	REIP	Rendement et biomasse
Indice de végétation amélioré (EVI)	$\frac{2.5(R_{NIR} - R_{Red})}{(R_{NIR} + 6R_{Red} - 7.5R_{Blue} + 1)}$	Maladie; biomasse.
Indice d'eau normalisé (IEN)	$\frac{R_{970} - R_{900}}{R_{970} + R_{900}}$	Humidité du sol et rendement des cultures

TABLE 1.2: Quelques indices de végétation récemment utilisés pour des applications de télédétection dans l'agriculture de précision

La mesure de la réflectance ou de l'émissivité dans les bandes de l'infrarouge proche et moyen est particulièrement utile pour développer des indices qui aident à comprendre plusieurs caractéristiques intrinsèques des plantes telles que la teneur en eau, les pigments, les teneurs en sucre, en glucides et en protéines. Le rayonnement réfléchi ou émis dans les bandes infrarouges thermiques est directement lié à la température de la plante. Comme la température de la plante est liée à son taux de transpiration, les indices obtenus à partir des données de réflectance thermique/infrarouge peuvent être utilisés pour comprendre l'état hydrique de la plante et d'autres stress biotiques et abiotiques tels que les maladies [26].

1.2.2.3 Applications les plus utilisées de la télédétection en agriculture

Gestion de l'eau d'irrigation Le moment et le taux d'application de l'irrigation jouent un rôle important dans l'atténuation du stress hydrique des cultures et dans l'obtention d'une croissance et d'un rendement optimaux des cultures. Diverses pratiques de gestion de l'irrigation sont utilisées par les agriculteurs en fonction de nombreux facteurs, notamment la disponibilité de l'eau, l'infrastructure de gestion de l'eau existante à la ferme (p. ex. le statut économique, la taille de l'exploitation, les connaissances de l'agriculteur, etc) [27].

De nombreux agriculteurs appliquent une irrigation uniforme à intervalles réguliers en fonction de leurs connaissances ou de leur expérience de l'agriculture, des sols et du climat du lieu. Les grands exploitants commerciaux déploient des systèmes de surveillance de l'humidité du sol pour irriguer (en mode automatique ou manuel) en fonction des données mesurées sur l'humidité du sol et des besoins en eau des plantes.

Presque toutes ces pratiques agricoles conventionnelles ne tiennent pas compte de la variabilité au sein d'un champ et utilisent un taux d'irrigation uniforme pour l'ensemble du champ. Les données de télédétection peuvent aider à discerner la variabilité à l'intérieur du champ et à appliquer un taux d'irrigation variable avec les systèmes d'irrigation couramment utilisés tels que le pivot central. L'application d'un taux variable peut aider à atténuer le stress hydrique résultant de conditions extrêmes

d'humidité et de sécheresse afin d'obtenir des rendements uniformément élevés dans toutes les parties du champ tout en réduisant les pertes d'eau et de nutriments. Les images de télédétection, recueillies plusieurs fois au cours d'une saison de croissance, sont utilisées pour déterminer divers indicateurs de la demande en eau des cultures, tels que l'ET, l'humidité du sol et le stress hydrique des cultures. Ces indicateurs sont utilisés pour estimer les besoins en eau des cultures et programmer l'irrigation avec précision.

Évapotranspiration (ET) L'évapotranspiration (ET) est la quantité d'eau transférée vers l'atmosphère, par l'évaporation au niveau du sol et au niveau de l'interception des précipitations, et par la transpiration des plantes. Les méthodes conventionnelles de mesure de l'ET (par exemple, lysimètre à pesée et covariance de Foucault) sont généralement coûteuses et ne fournissent pas d'estimations de l'ET variables dans l'espace, résultant des différences dans l'utilisation des terres, les sols, la topographie et d'autres processus hydrologiques. Les données de télédétection sont largement utilisées pour estimer l'ET, ce qui est nécessaire pour déterminer les besoins en eau des cultures et programmer l'irrigation [28].

Les approches d'estimation de l'ET, basées sur les données de télédétection, peuvent être regroupées en trois catégories :

- le bilan énergétique de surface,
- le coefficient de récolte,
- et la méthode de Penman-Monteith

De nombreuses études ont passé en revue les techniques d'estimation de l'ET basées sur la télédétection, y compris l'article [29] qui a discuté des théories de développement de plusieurs approches/méthodes d'estimation de l'ET ainsi que de leurs avantages et limites. L'approche du bilan énergétique de surface a été largement utilisée pour l'estimation de l'ET au cours des cinq dernières années. Certaines études ont utilisé des méthodes hybrides pour combiner les approches du coefficient de récolte et du bilan énergétique pour l'estimation de l'ET. Dans une approche de bilan énergétique de surface, le flux de rayonnement net (R_n), le flux de chaleur du sol (G) et le flux

de chaleur sensible (H) sont généralement calculés à partir de données de télédétection.

Ces flux sont généralement calculés à partir de données de télédétection dans les bandes du visible, du proche infrarouge et de l'infrarouge thermique, tandis que le flux de chaleur latente (ET) est calculé comme un rappel du terme dans l'équation du bilan énergétique.

$$\lambda ET = R_n - G - H \quad (1.1)$$

Gestion des maladies Les maladies peuvent entraîner une perte importante de la production agricole et des bénéfices de l'exploitation agricole. La détection précoce des maladies des plantes et de leur étendue spatiale peut aider à contenir la propagation de la maladie et à réduire les pertes de production. Le repérage sur le terrain, une méthode conventionnelle de détection des maladies, prend du temps, demande beaucoup de travail et est sujet à l'erreur humaine. En outre, il peut être difficile de détecter la maladie aux premiers stades, lorsque les symptômes ne sont pas entièrement visibles. En outre, certaines maladies ne présentent aucun symptôme visible, ou l'effet peut ne pas être perceptible jusqu'à ce qu'il soit trop tard pour agir. Il est également difficile de cartographier l'étendue spatiale et la gravité de la propagation de la maladie avec la méthode traditionnelle de repérage sur le terrain [30].

La télédétection pourrait être utilisée pour surveiller la maladie de manière efficace, en particulier dans les premiers stades de développement de la maladie, lorsqu'il peut être difficile de discerner les signes de la maladie avec le repérage sur le terrain. De multiples techniques utilisant l'imagerie RVB, multi-spectrale, hyperspectrale, thermique et de fluorescence ont été utilisées pour identifier les maladies dans une variété de cultures. Une bonne corrélation entre la maladie de la raie de la vigne et le NDVI généré à partir d'images de drones a été découverte en Italie [31].

Par rapport aux indices de végétation généralement utilisés (par exemple, NDVI), le développement d'indices spectraux de maladies (SDI) spécifiques aux maladies

peut augmenter la précision de la détection et de la différenciation des maladies dans des conditions réelles de terrain. Une plus grande précision dans la détection des maladies pendant les stades critiques de la croissance des cultures, comme la floraison, peut aider à mettre en œuvre des plans de gestion efficaces en temps voulu pour atténuer ou éviter les pertes importantes de rendement et de qualité des récoltes [32].

L'utilisation de SDI, à la place des VI classiques, peut également réduire la complexité des méthodes de détection des maladies et augmenter l'efficacité du système en réduisant la demande de calcul. Bien que des études aient été menées pour la classification des maladies des plantes, des efforts supplémentaires sont nécessaires pour développer des méthodologies plus précises, automatisées et reproductibles pour la détection des maladies dans diverses conditions climatiques et de terrain [33].

Suivi des cultures et rendement Le suivi de la croissance et du rendement des cultures est nécessaire pour comprendre la réponse des cultures à l'environnement et aux pratiques agronomiques et pour développer des plans de gestion efficaces pour les travaux sur le terrain et/ou les remèdes. Le LAI et la biomasse sont deux indicateurs essentiels de la santé et du développement des cultures. Le LAI est également utilisé comme une entrée dans de nombreux modèles de prévision de la croissance et du rendement des cultures. Les méthodes In-Situ d'estimation du LAI (physiques et optiques) [34] demandent beaucoup de temps et de travail, tout comme les méthodes destructives sur le terrain utilisées pour l'estimation de la biomasse. De plus, ces méthodes ne fournissent pas de carte de variabilité spatiale de la croissance des cultures et de la biomasse.

Les données de télédétection sur la croissance des cultures (ex. : LAI) et la biomasse peuvent aider à obtenir des informations précieuses sur les propriétés spécifiques du site (ex. : sols, topographie), la gestion (ex. : eau, nutriments et autres intrants), et les divers facteurs de stress biotiques et abiotiques (ex. : température, humidité, etc.). De même, les données de télédétection peuvent également être utilisées pour cartographier les différences dans le travail du sol, c'est à dire le comportement du

sol, et la gestion des résidus et leurs effets sur la croissance des cultures [35].

Plusieurs études ont utilisé des images hyperspectrales avec diverses techniques d'apprentissage automatique et de classification pour cartographier le travail du sol et les résidus de culture dans les champs agricoles. Ces informations sur l'état des cultures et les pratiques de travail du sol peuvent aider à élaborer des plans de gestion spécifiques au site, y compris l'application variable d'eau, d'éléments nutritifs et de pesticides pour augmenter la production et l'efficacité de la gestion. En général, ces études utilisent un ensemble de données de référence (par exemple, l'indice LAI mesuré et les indices de végétation correspondants) pour développer une approche basée sur la régression ou l'apprentissage automatique afin d'estimer l'indice LAI et/ou la biomasse pour un champ cible.

Des chercheurs ont utilisé plusieurs indices spectraux en conjonction avec la hauteur mesurée des plantes pour estimer la biomasse dans des parcelles à irrigations multiples et à traitement nutritif pour le blé d'hiver cultivé en Chine [?].

Pour l'AP, la connaissance de la variabilité spatiale du rendement des cultures est importante pour comprendre la réponse des cultures aux pratiques de gestion et aux facteurs de stress environnementaux. Les paramètres biophysiques des cultures dérivés de la télédétection, ou les indices de végétation, ont une forte corrélation avec le rendement des cultures et la biomasse observés, ce qui indique leur potentiel d'utilisation dans l'estimation du rendement[36].

L'estimation du rendement des cultures à partir de données de télédétection a généralement été réalisée de deux manières.

Premièrement, les paramètres biophysiques (par exemple, l'indice de végétation) dérivés des données de télédétection sont utilisés dans un modèle de culture pour estimer le rendement des cultures et la biomasse. Deuxièmement, des relations statistiques (par exemple, régression) ou empiriques sont développées entre les paramètres/indices de culture dérivés de la télédétection (par exemple, NDVI, LAI) et le rendement des

cultures et la biomasse observés dans un champ de culture représentatif. Le modèle de régression ou la relation empirique développés peuvent ensuite être utilisés pour cartographier le rendement d'un champ de culture cible. La modélisation des cultures est une approche gourmande en données, qui nécessite une grande quantité d'informations comme les paramètres d'entrée du modèle, les données météorologiques et les données de rendement et de biomasse observées.

Gestion des mauvaises herbes Les approches conventionnelles de gestion des mauvaises herbes impliquant une application uniforme d'herbicide sont une pratique inefficace qui augmente le risque de pertes de pesticides sur site. L'application d'herbicides à un taux variable en fonction des besoins peut contribuer à améliorer l'efficacité du traitement et à réduire les coûts des intrants et la pollution de l'environnement. La télédétection a été largement utilisée pour cartographier les parcelles d'adventices dans les champs de culture pour une gestion des adventices spécifique au site.

Les mauvaises herbes peuvent être identifiées ou différenciées des plantes cultivées en fonction de leur signature spectrale particulière liée à leurs attributs phénologiques ou morphologiques qui sont différents de ceux de la culture. Au cours des dernières années, les approches d'apprentissage automatique sont apparues comme une méthode très précise et efficace de classification d'images pour la cartographie des mauvaises herbes. Deux types généraux d'approches de classification d'images sont couramment utilisés pour la cartographie des mauvaises herbes : la classification supervisée et la classification non supervisée. Bien que chaque méthode ait ses propres forces et faiblesses, la classification supervisée prend beaucoup de temps et nécessite plus de travail manuel [37].

Les drones ont été la plateforme de télédétection la plus populaire pour la cartographie et la gestion des mauvaises herbes, principalement en raison de leur capacité à produire des images d'une résolution de l'ordre du centimètre (5 cm) nécessaires à la détection et à la cartographie des mauvaises herbes. En utilisant l'approche du réseau entièrement convolutif, une étude a permis de cartographier les

mauvaises herbes dans une rizière en Chine avec une précision allant jusqu'à 90% [38].

Un pulvérisateur de mauvaises herbes ciblé en utilisant une approche de réseau neuronal d'apprentissage profond pour la détection des mauvaises herbes à partir de capteurs au sol a pu être mis en place, il a donné une précision d'application de 71% dans des champs expérimentaux en Floride, aux États-Unis [39].

1.2.3 Réseaux de capteurs sans fil (RCSF)

Avec les avancées technologiques qu'a connu l'électronique, notamment dans le domaine des capteurs et du traitement de signal, mais aussi grâce à la disponibilité de différents types de capteurs à prix bas, l'utilisation de ces derniers dans l'agriculture, plus particulièrement les RCSFs notamment pour la facilité de leur déploiement dans un environnement délicat comme les surfaces agricoles.

D'après [40] les RCSFs sont un réseau de capteurs alimentés par batterie et interconnectés par un support sans fil pour servir un objectif d'application spécifique.

En effet, de nos jours, les RCSFs sont largement utilisés dans la surveillance de l'agriculture pour améliorer la qualité et la productivité de l'agriculture. Généralement, les capteurs collectent différents types de données (c.-à-d. Humidité, niveau de dioxyde de carbone, température, pH) dans des scénarios en temps réel pour ensuite transmettre ces données à un bloc centralisé pour traiter ces données et prendre des actions en fonction des informations tirées.

1.2.3.1 Types de RCSFs

Dans cette section, nous discutons de deux variantes largement utilisées des RCSFs, les RCSFs terrestres (TWSN) et les RCSFs sous terrains (WUSN), spécifiquement utilisés dans les applications agricoles.

RCSFs terrestres ou terrestrial wireless sensor network (TWSN) Dans les TWSNs, les nœuds sont déployés au-dessus de la surface du sol. Les progrès de la

technologie MEMS ont permis la création de capteurs intelligents, de petite taille et à faible coût. Ces capteurs puissants permettent à un nœud de capteurs de collecter avec précision les données environnantes. Sur la base des informations détectées, ces nœuds se mettent en réseau pour répondre aux exigences de l'application.

Par exemple, dans un environnement d'AP, des RCSFs sont déployés dans tout le champ pour automatiser le système d'irrigation. Tous ces capteurs déterminent le taux d'humidité du sol, et ensuite, décident en coordination du moment et de la durée de la programmation de l'irrigation sur ce champ. Ensuite, à l'aide du même réseau, la décision est transmise au nœud de capteur relié à une pompe à eau. Dans [41], les auteurs ont proposé un tel système d'irrigation automatisé en utilisant un RCSF et un module GPRS.

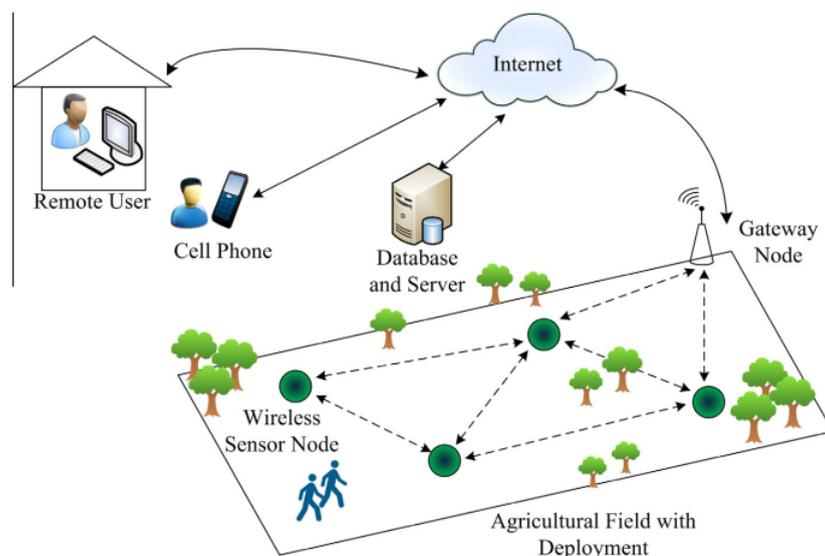


FIGURE 1.7: RCSF typique déployé sur champ pour des applications agricoles

La *figure1.8* représente un RCSF typique déployé sur champ pour des applications agricoles. Le champ est constitué de nœuds de capteurs alimentés par des capteurs embarqués spécifiques à l'application.

Les nœuds du réseau de capteurs sur le terrain communiquent entre eux en utilisant des liaisons radiofréquences (RF) des bandes radio industrielles, scientifiques et médicales (ISM) (telles que 902-928 MHz et 2,4-2,5 GHz). En général, un nœud

passerelle est également déployé avec les nœuds de capteur pour permettre une connexion entre le réseau de capteurs et le monde extérieur [40].

Ainsi, le nœud passerelle est alimenté à la fois par la RF et par le système mondial de communications mobiles (GSM) ou GPRS.

Un utilisateur distant peut surveiller l'état du champ et contrôler les capteurs et les dispositifs d'actionnement sur le terrain [41].

Par exemple, un utilisateur peut activer ou désactiver une pompe ou une vanne lorsque le niveau d'eau appliqué au champ atteint une valeur seuil prédéfinie. Les utilisateurs munis d'un téléphone portable peuvent également surveiller et contrôler à distance les capteurs sur le terrain. L'utilisateur mobile est connecté par GPRS ou même par SMS. Il est également possible de concevoir une mise à jour périodique des informations fournies par les capteurs et un contrôle du système à la demande pour les deux types d'utilisateurs (distants ou présents dans l'exploitation).

RCSFs souterrains ou wireless underground sensor network (WUSN) Une autre variante des RCSFs est son homologue souterrain. Les RCSFs souterrains (WUSN) [42]. Dans cette version, les capteurs sans fil sont plantés dans le sol. Dans ce contexte, les hautes fréquences subissent une forte atténuation, et les fréquences comparativement plus basses sont capables de pénétrer à travers le sol [43]. Ainsi, le rayon de communication est limité et le réseau nécessite un plus grand nombre de nœuds pour couvrir une grande zone. L'application de capteurs filaires augmente la couverture du réseau en nécessitant un nombre relativement plus faible de capteurs.

Cependant, dans cette conception, les capteurs et les fils peuvent être vulnérables aux activités agricoles.

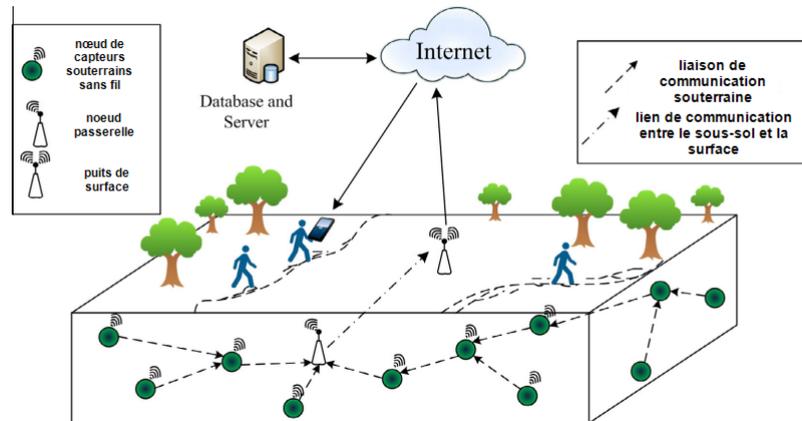


FIGURE 1.8: Application agricole basée sur des réseaux de capteurs souterrains

Une application agricole typique basée sur des réseaux de capteurs souterrains est illustrée à la *figure1.9*. Contrairement aux applications basées sur les TWSNs présentées à la *figure1.8*, les nœuds de capteurs sont enterrés dans le sol.

Un nœud passerelle est également déployé pour transmettre les informations collectées par les nœuds de capteurs souterrains au puits de surface placé au-dessus du sol. Par la suite, les informations peuvent être transmises par Internet pour être stockées dans des bases de données distantes, et peuvent être utilisées pour avertir l'utilisateur d'un téléphone portable. Cependant, en raison de la distance de communication comparativement plus courte, un plus grand nombre de nœuds doivent être déployés pour être utilisés dans les WUSN.

1.2.3.2 Différences entre les types de RCSFs

Caractéristique	TWSNs	WUSNs
Déploiement	placés au-dessus du sol	enterrés sous terre
Profondeur	n'importe où en surface	Terre végétale (0-30 cm) et sous-sol (>30 cm)
plage de communication	~100 m	~0.1–10 m
fréquence de communication	plus grande (868-915 MHz, 2.4 GHz)	plus basse (433 MHz, 8–300 kHz)
Consommation d'énergie	Plus basse	plus grande
Coût	Moins cher	plus cher

TABLE 1.3: Différences entre TWSNs et WUSNs.

1.2.3.3 Conception d'un réseau de capteurs sans fil pour les applications agricoles

Architecture de réseau pour les applications agricoles Dans cette section, nous discutons de l'architecture de réseau considérée dans diverses applications agricoles.

- Architecture stationnaire : Dans l'architecture stationnaire, les nœuds de capteurs sont déployés à une position fixe et ne changent pas de position pendant la durée de l'application. Typiquement, les applications telles que le système de gestion de l'irrigation, la surveillance de la qualité de l'eau souterraine, et le contrôle de l'utilisation d'engrais nécessitent des architectures stationnaires. Dans ces applications avec les TWSNs, les nœuds capteurs enregistreurs de données (collecteurs de données) sont généralement placés sur le terrain. Cependant, dans les WUSN, les nœuds de capteurs collecteurs de données sont placés sous terre.
- Architecture mobile : Les architectures mobiles comprennent des dispositifs qui changent de position avec le temps. Un exemple d'applications basées sur ce type

- d'architecture serait un réseau autonome de tracteurs et d'agriculteurs munis d'un téléphone portable servant à des opérations agricoles omniprésentes.
- Architecture hybride : Dans l'architecture hybride, les nœuds stationnaires et mobiles sont présents. Par exemple, ce type d'architecture est applicable aux applications agricoles composées de capteurs de champ fixes, d'équipements agricoles mobiles, d'utilisateurs transportant des téléphones portables et de bétail en mouvement.

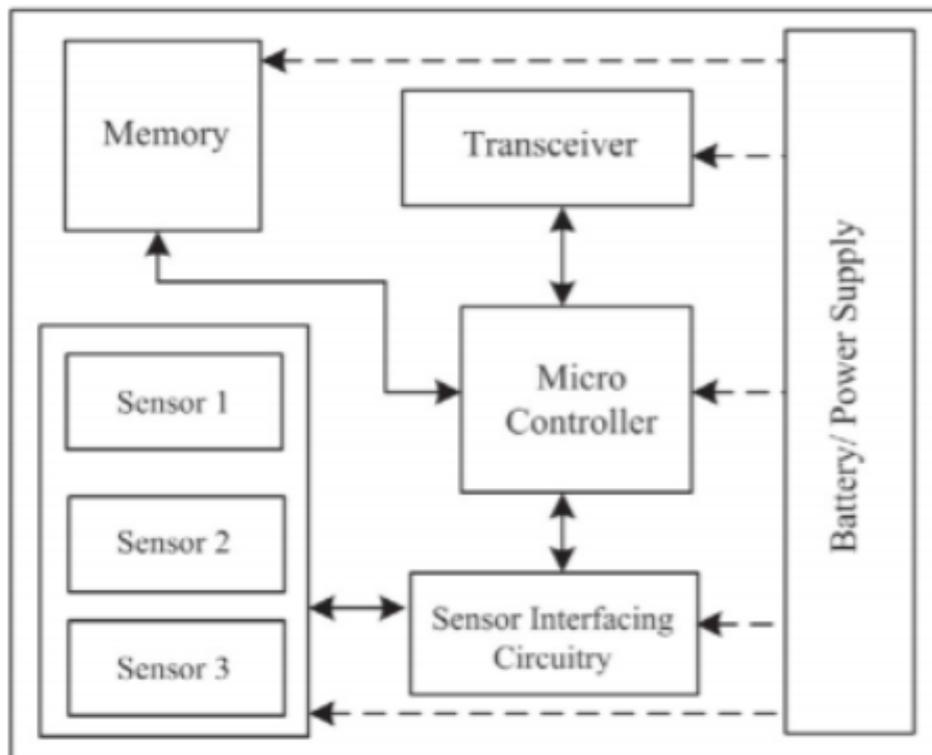


FIGURE 1.9: Architecture d'un nœud de capteurs multi-puces embarqué typique

Fonctionnalité	MICA2	MICAZ	TelosB	IRIS	LOTUS	Imote2	SunSPOT
Processeur	ATmega128L	ATmega128L	TIMSP430	ATmega128L	Cortex M3 LPC 17xx	Marvell/ XScalePXA271	ARM 920T
Programmation							JAVA
Vitesse d'horloge (MHz)	7.373	7.373	6.717	7.373	10-100	13-416	180
Largeur du bus (bits)	8	8	16	8	32	32	32
Mémoire système (kB)	4	4	10	4	64	256	512
Mémoire flash (kB)	Program :128 Serial :512	Program :128 Serial :512	Program :48 Serial :1024	Program :128 Serial :512	Program :512 Serial :64 ×1024	Program : programmable Serial :31 ×1024	Program :4096 -
Bande de fréquence de fonctionnement (MHz)	868/915	2400	2400	2400	2400	2400	2400
Puce émettrice-réceptrice	CC1000	CC2420	CC2420	Atmel RF230	Atmel RF231	CC2420	802.15.4
Nombre de canaux	4/50	Programmable	Programmable	Programmable	-	In steps of 5 MHz	-
Débit de données (kpbs)	38.4 (Baud)	250	250	250	250	250	250
Connectivité E/S	UART, I2C, SPI, DIO	3xUART, SPI, I2C, I2S, GPIO, ADC	UART 3x, I2C, GPIO, SPI2x, DIO, JTAG	DIO, I2C, GPIO			

TABLE 1.4: Comparaison des plateformes de capteurs sans fil existantes

Architecture des nœuds de capteurs

Nœuds de capteurs multi-puces embarqués Les composants d'un nœud de capteurs multi-puces typique sont illustrés à la Figure 1.9. En général, un nœud de capteurs se compose d'un réseau de capteurs spécifiques à une application et d'un émetteur-récepteur pour la communication. Un processeur ou un microcontrôleur est utilisé comme le "cerveau" du nœud. En option, une carte de capteurs comprend des unités de mémoire pour stocker des données. Selon la demande de l'application, l'architecture des nœuds de capteurs varie pour répondre aux exigences. Par exemple, la puissance de traitement et la taille de la mémoire embarquée sont augmentées pour répondre aux exigences des applications plus intenses. À cet égard, une autre technologie importante est le système en boîtier (SiP), qui est défini comme toute combinaison de plusieurs puces, y compris les composants passifs (tels que les résistances et condensateurs), montées ensemble tout en gardant la possibilité de fixer des composants externes ultérieurement. Le SiP réduit le coût du produit en optimisant sa taille et ses performances. Ainsi, la technologie SiP a un potentiel d'applications dans les scénarios agricoles. Les systèmes agricoles basés sur SiP peuvent être utilisés dans différentes applications, simplement en attachant différents capteurs au boîtier principal. Dans ce qui suit, nous examinons les facteurs associés à la sélection des composants d'un nœud de capteurs en fonction des exigences des applications agricoles.

- Le processeur : La puissance de calcul du nœud de capteurs dépend uniquement du choix de l'unité de traitement. Un microcontrôleur offre quelques avantages tels qu'un faible coût, la flexibilité de communiquer avec d'autres nœuds, la facilité de programmation et l'absence d'erreurs. une faible consommation d'énergie par rapport aux processeurs traditionnels. La plupart du temps, ces microcontrôleurs fonctionnent sur 3,5-5 V. Cependant, la consommation d'énergie est l'un des facteurs les plus importants dans les nœuds de capteurs. Compte tenu de ce fait, les microcontrôleurs sont préférés aux processeurs à usage général.
- Émetteur-récepteur : La transmission et la réception sont les deux principales raisons de la consommation d'énergie dans les nœuds de capteurs. Dans les applications agricoles, le planificateur du réseau choisit le déploiement pour assurer une consommation d'énergie optimale des nœuds de capteurs.

- La mémoire : Les nœuds de capteurs disposent de deux types de mémoire embarquée : la mémoire associée au processeur et la mémoire externe. Selon les besoins de l'application, les nœuds de capteurs doivent stocker les données historiques pour une prise de décision intelligente. À cet égard, les mémoires flash sont utilisées pour un stockage supplémentaire.
- Puissance : il s'agit également d'un facteur important pour la sélection des nœuds de capteurs car la puissance de la batterie des nœuds de capteurs est limitée. Dans de nombreuses applications agricoles, les nœuds possèdent des sources d'énergie alternatives telles que l'énergie solaire. Cependant, l'énergie solaire n'est disponible que pendant la journée. Les nœuds dépendent de l'alimentation par batterie. De plus, le changement fréquent de la batterie augmente le coût de la maintenance. Par conséquent, nous avons besoin d'algorithmes efficaces sur le plan énergétique afin de réduire la consommation d'énergie des nœuds de capteurs.
- Le coût : Un facteur de sélection très important des nœuds de capteurs est le coût total du matériel. Une conception d'application à faible coût est toujours préférable, quel que soit le niveau d'application. Par conséquent, il s'agit de la question la plus importante en termes d'applications ciblant les marchés des pays à revenu faible ou intermédiaire.

1.2.3.4 Applications communes des RCSFs en agriculture

Système de gestion de l'irrigation L'agriculture moderne nécessite un système de gestion de l'irrigation amélioré pour optimiser l'utilisation de l'eau en agriculture[44]. La réduction alarmante du niveau des eaux souterraines est une autre motivation pour l'exigence d'un système avancé.

Dans ce contexte, les techniques de micro-irrigation sont rentables et efficaces en termes d'utilisation de l'eau[45]. Il est aussi possible d'améliorer l'efficacité de la micro-irrigation en fonction des informations sur l'environnement et le sol.

A cet égard, les RCSFs sont appliqués comme technologie de coordination, dans [46] les auteurs ont décrit le développement et le déploiement d'un système d'irrigation

automatisé comprenant un RCSF, d'une passerelle et d'un serveur distant.

Le projet avait pour objectif de mettre en œuvre un système RCSF capable de réduire la consommation d'eau. Le RCSF se compose de capteurs d'humidité et de température du sol enterrés dans le sol pour prendre des mesures à différentes profondeurs. Le nœud de la passerelle a des installations embarquées qui supportent à la fois ZigBee (IEEE Standard for Information technology) et les communications GPRS. Il est également doté de prise de décision intelligente, comme l'activation automatique de l'irrigation en fonction des valeurs d'humidité et de température du sol dépassant un certain seuil prédéfini.

Le serveur distant est utilisé pour stocker toutes les informations et les afficher dans une interface utilisateur graphique (GUI). L'avantage de cette application est sa fonction d'analyse des données en temps réel.

Fertilisation La croissance des plantes et la qualité des récoltes dépendent étroitement de l'utilisation d'engrais. Cependant, trouver l'apport optimal d'engrais aux endroits appropriés dans les champs est une tâche délicate.

L'utilisation d'engrais pour l'agriculture peut être contrôlée en surveillant la variation des nutriments du sol tels que l'azote (N), le phosphore (P), le potassium (K) et le pH. À travers ce contrôle, l'équilibre nutritionnel du sol peut être atteint, et donc la qualité de la production agricole maintenue. Les auteurs de [47] ont étudié l'efficacité des nœuds mobiles pour améliorer la productivité agricole dans un système intelligent avec des pulvérisations de précision, ils ont trouvé que la distribution des nœuds a légèrement affecté l'efficacité du réseau, le réseau homogène étant subtilement plus efficace que le réseau hétérogène.

Prédictions de parasites/maladies dans l'agriculture Ces dernières années, les maladies des cultures et les insectes nuisibles ainsi que leur degré d'infection ont connu une forte augmentation, ce qui a causé d'énormes pertes économiques. C'est ce qui a conduit à de nombreuses recherches quant à des produits plus efficaces pour la lutte contre ces phénomènes. L'utilisation des RCSFs dans ce contexte est dans un but de suivi de ces phénomènes dans les parcelles agricoles

ciblées et d'utiliser en conséquence les produits avec précision pour réduire la maladie.

Une tentative a été faite dans [48] pour comprendre les relations cachées entre la maladie Bud Necrosis Virus (BNV), les ravageurs et les paramètres météorologiques de la culture de l'arachide.

Un RCSF a été mis en place dans le banc d'essai pour obtenir en temps réel des paramètres météorologiques (température, humidité et humidité des feuilles) au niveau microclimatique et quelques paramètres météorologiques connexes de la station météorologique voisine. La dynamique culture-météo- peste-maladie et les relations cachées ont été obtenues et quantifiées à l'aide de techniques de data mining. L'approche statistique a permis de développer un modèle de régression multivariée qui a été utilisé pour développer un modèle de prédiction empirique (non cumulatif) afin de prévoir l'accroissement de la population, l'initiation, la sévérité des ravageur et les maladies.

1.2.3.5 Perspectives et problèmes des solutions existantes

Nous discutons des perspectives de l'application des RCSFs dans ce qui suit. Cependant, il est encore possible d'améliorer les solutions. Nous énumérons quelques points pour améliorer les innovations existantes.

- Des solutions rentables pour les PFR sont nécessaires : Les efforts actuels de recherche et de développement visent à réduire les coûts de matériel et de logiciel, tout en maximisant le rendement du système. Le coût global du système augmente en raison de l'utilisation de dispositifs importés de l'étranger pour construire les systèmes. Cependant, pour les applications dans les PFR, nous devons réduire davantage le coût du système. Ainsi, le défi consiste à réduire davantage le coût.
- Tolérance aux pannes : La tolérance aux pannes est une caractéristique nécessaire des RCSFs pour réaliser une agriculture de précision. Différents types de pannes pouvant survenir dans un système d'agriculture de précision basé sur un RCSF sont les suivantes défaillance d'un nœud due à l'épuisement de la batterie ou à toute autre raison, défaillance du matériel du capteur génération

- d'une valeur erronée, calibrage défectueux du capteur, échec de la communication. Dans la littérature existante, les auteurs de [?] ont présenté un système de gestion de l'irrigation qui est tolérant aux pannes de nœuds et de communication. En cas de panne, le système suit le programme d'irrigation par défaut. Pour réduire les risques de défaillance des nœuds en raison de l'épuisement de l'énergie des nœuds alimentés par l'énergie solaire ont été utilisés [?]. De plus, les schémas de gestion de la topologie et d'agrégation des données doivent être tolérants aux pannes dans un déploiement à grande échelle.
- Gestion de l'énergie et récolte de l'énergie : La gestion de l'énergie est une question importante dans tout système basé sur les RCSFs. Les composants et les algorithmes du système doivent être conçus en gardant cette question à l'esprit. Par ailleurs, les solutions potentielles de récolte d'énergie solaire, l'énergie éolienne la biomasse et les vibrations doivent également être prises en compte lors de la conception de systèmes d'agriculture de précision basés sur les RCSFs. Parmi les travaux existants, les RCSFs à énergie solaire sont déjà utilisés. Par conséquent, les travaux futurs ont la possibilité de travailler sur ces WSNs basés sur des ressources non conventionnelles.
 - Une simplification des solutions existantes est nécessaire : L'utilisateur final de la plupart des applications agricoles ciblées sont les agriculteurs. Ainsi, les plates-formes et les solutions conçues sont envisagées pour être simplifiées en termes de facilité d'utilisation. À cet égard, les questions d'interaction homme-ordinateur tels que l'accessibilité et la facilité d'utilisation doivent être prises en compte.

1.3 Conclusion

Nous venons, à travers ce premier chapitre, d'introduire les éléments nécessaires pour comprendre les enjeux et les contraintes associés à l'agriculture classique. Le développement actuel permet de lever ces contraintes à travers aussi bien la robotisation que par des systèmes traitant les données de bout en bout (du capteur au traitement en passant par les communications et l'agrégation de données).

Dans le chapitre suivant, on abordera les outils d'apprentissage automatique nous permettant de traiter ces données agrégées à travers les RCSFs et les exploiter.

Chapitre 2

Intelligence artificielle pour la prévision de données

2.1 Partie 1 : IA et ses utilisations dans l'agriculture de précision

2.1.1 Introduction

L'industrie de l'agriculture est l'une des industries comportant le plus de risques.

En même temps, la population mondiale accroit et l'urbanisation se poursuit. Le revenu disponible augmente et les habitudes de consommation changent. Les agriculteurs sont soumis à une forte pression pour répondre à la demande croissante, et ils ont besoin d'un moyen d'accroître leur productivité. Dans trente ans, il y aura plus de gens à nourrir. Et comme la quantité de sols fertiles est limitée, il faudra aussi aller au-delà de l'agriculture traditionnelle [49].

Il est donc impératif de chercher des moyens d'aider les agriculteurs à minimiser leurs risques, ou du moins à les rendre plus gérables. La mise en œuvre de l'IA dans l'agriculture à l'échelle mondiale est l'une des opportunités les plus prometteuses.

L'IA peut potentiellement changer la façon dont nous voyons l'agriculture, en permettant aux agriculteurs d'obtenir plus de résultats avec moins d'efforts tout en apportant de nombreux autres avantages. Toutefois, l'IA n'est pas une technologie qui fonctionne de manière autonome. En tant que prochaine étape sur le chemin de l'agriculture traditionnelle vers l'innovation, l'IA peut compléter les technologies déjà mises en œuvre, elle peut apporter des avantages tangibles aux petites choses du quotidien et simplifier la vie des agriculteurs à bien des égards. Nous allons donc voir dans cette partie comment pouvoir utiliser l'IA pour une agriculture durable, ses opportunités dans l'agriculture et ses potentielles contributions pour relever les défis existants.

La définition et la signification précises de l'intelligence artificielle fait aujourd'hui l'objet de nombreuses discussions et suscite beaucoup de confusion. Un seul dictionnaire [The New International Webster's Comprehensive Dictionary of the English Language, Encyclopedic Edition], par exemple, donne quatre définitions de l'intelligence

artificielle :

- Un domaine d'étude dans le domaine de l'informatique. L'IA s'intéresse au développement d'ordinateurs capables de s'engager dans des processus de pensée semblables à ceux des humains, tels que l'apprentissage, le raisonnement et l'autocorrection.
- Le concept selon lequel les machines peuvent être améliorées pour assumer certaines capacités normalement considérées comme relevant de l'intelligence humaine, telles que l'apprentissage, l'adaptation, l'auto-correction, etc.
- L'extension de l'intelligence humaine par l'utilisation d'ordinateurs, comme la puissance physique était autrefois étendue par l'utilisation d'outils mécaniques.
- Dans un sens restreint, l'étude des techniques permettant d'utiliser les ordinateurs de manière plus efficace en améliorant les techniques de programmation.

Les définitions ont également changé au fil du temps, en raison des évolutions rapides. Les définitions les plus récentes parlent d'imitation du comportement humain intelligent, ce qui correspond déjà à une définition beaucoup plus forte [50].

2.1.2 Utilité de l'IA en l'agriculture

L'agriculture implique un certain nombre de processus et d'étapes, dont la plus grande partie est manuelle. En complétant les technologies adoptées, l'IA peut faciliter les tâches les plus complexes et les plus courantes. Elle peut rassembler et traiter de grandes données sur une plateforme numérique, proposer le meilleur plan d'action, et même lancer cette action lorsqu'elle est combinée à d'autres technologies [51].

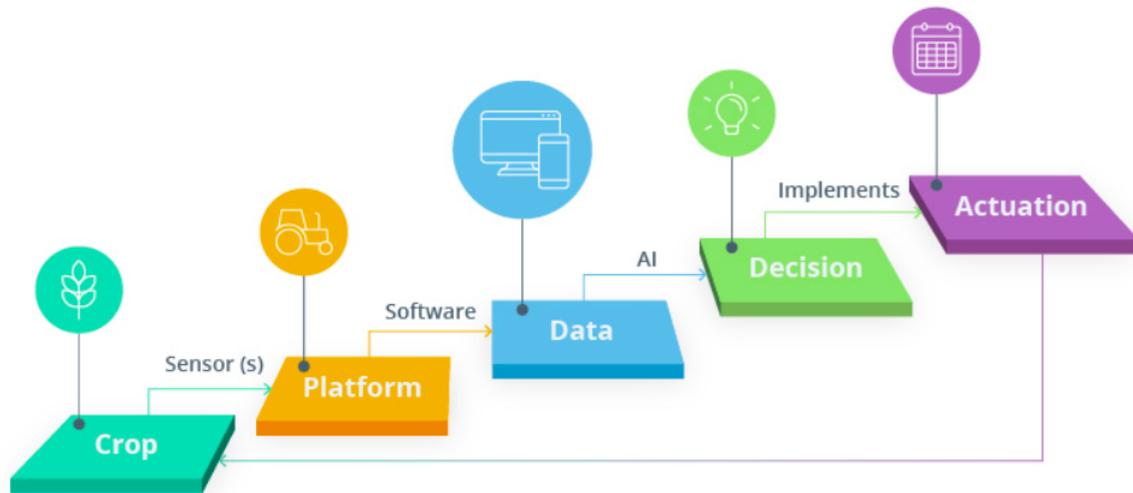


FIGURE 2.1: Rôle de l'IA dans le cycle de gestion de l'information agricole.

L'utilisation de l'IA en agriculture aide les agriculteurs à collecter et comprendre les données, à partir de capteurs et logiciels (Plateformes), telle que la température, les précipitations, la vitesse du vent et le rayonnement solaire. L'analyse des données historiques offre une meilleure comparaison des résultats souhaités. La mise en œuvre de l'IA dans l'agriculture ne supprimera pas les emplois des agriculteurs humains, mais améliorera leurs processus.

La combinaison de l'IA et de l'agriculture peut être bénéfique pour plusieurs processus, par exemple dans l'analyse de la demande du marché, l'IA peut simplifier la sélection des cultures et aider les agriculteurs à identifier les produits qui seront les plus rentables ou les plus demandés. Elle permet aussi une meilleure gestion des risques, en effet, les agriculteurs peuvent utiliser les prévisions et l'analyse prédictive pour réduire les erreurs dans les processus commerciaux et minimiser les risques de mauvaises récoltes. Dans la sélection des semences, en recueillant des données sur la croissance des plantes, elle peut aider à produire des cultures moins sujettes aux maladies et mieux adaptées aux conditions climatiques. Son implémentation dans des systèmes robustes nous octroie la possibilité de faire des analyses chimiques du sol, fournir des estimations précises des nutriments manquants, surveiller l'état des plantes et identifier les schémas d'irrigations optimaux nous permettant ainsi de simplifier la protection et le nourrissage des cultures.

2.1.3 Utilisation de l'IA pour résoudre les défis de l'agriculture

2.1.3.1 Prise de meilleures décisions

L'analyse prédictive peut aider à la gestion dans le domaine agricole. Les agriculteurs peuvent collecter et traiter beaucoup plus de données et le faire plus rapidement avec l'IA qu'ils ne le feraient autrement. L'analyse de la demande du marché, la prévision des prix et la détermination du moment optimal pour les semis et les récoltes sont les principaux défis que les agriculteurs peuvent résoudre grâce à elle.

2.1.3.2 Réduire des coûts

L'IA peut fournir aux agriculteurs des informations en temps réel sur leurs champs, leur permettant d'identifier les zones qui ont besoin d'être irriguées, fertilisées ou traitées avec des pesticides avec une grande précision minimisant de ce fait les pertes superflues du au traitement approximatif engendré par l'agriculture classique.

2.1.3.3 Réponse aux pénuries de main-d'œuvre

Le travail agricole est difficile, et les pénuries de main-d'œuvre dans ce secteur ne sont pas nouvelles. Les agriculteurs peuvent résoudre ce problème avec l'aide de l'automatisation à travers l'IA, les tracteurs sans conducteur, les systèmes d'irrigation et de fertilisation intelligents, la pulvérisation intelligente et les robots basés sur l'IA pour la récolte sont quelques exemples de la manière dont les agriculteurs peuvent effectuer le travail sans avoir à embaucher davantage de personnes. Par rapport à n'importe quel travailleur agricole humain, les outils pilotés par l'IA sont plus rapides, plus résistants et plus précis.

2.1.4 Exploitations agricoles et l'adoption de l'IA

Compte tenu des avantages de l'IA pour l'agriculture durable, l'adoption de cette technologie peut sembler être une étape logique pour chaque agriculteur. Cependant, il existe encore quelques contraintes importantes.

2.1.4.1 Longueur du processus d'adoption de la technologie

Les agriculteurs doivent comprendre que l'IA n'est qu'une partie avancée de technologies plus simples pour le traitement, la collecte et le suivi des données sur le terrain. Pour fonctionner, l'IA a besoin d'une infrastructure technologique appropriée. C'est pourquoi même les exploitations qui disposent déjà d'une certaine technologie peuvent avoir du mal à aller de l'avant. C'est également un défi pour les sociétés de logiciels. Elles doivent approcher les agriculteurs progressivement, en leur proposant d'abord des technologies plus simples, comme une plateforme de commerce agricole. Une fois que les agriculteurs se seront habitués à une solution moins compliquée, il sera raisonnable de passer à la vitesse supérieure et de leur proposer autre chose, notamment des fonctions d'IA.

2.1.4.2 Manque d'expérience des technologies émergentes

Le secteur agricole des pays en développement est différent de celui de l'Europe occidentale et des États-Unis. Certaines régions pourraient bénéficier d'une agriculture fondée sur l'IA, mais il peut être difficile de vendre une telle technologie dans des régions où la technologie agricole n'est pas courante. Les agriculteurs auront très probablement besoin d'aide pour l'adopter.

Par conséquent, les entreprises technologiques qui espèrent faire des affaires dans les régions où l'économie agricole est émergente devront peut-être adopter une approche proactive. En plus de fournir leurs produits, elles devront assurer la formation et le soutien continu des agriculteurs et des agro-industriels qui sont prêts à adopter des solutions innovantes.

2.1.4.3 Questions de confidentialité et de sécurité

Comme il n'existe pas de politiques et de réglementations claires autour de l'utilisation de l'IA, non seulement dans l'agriculture mais en général, l'AP et l'agriculture intelligente soulèvent diverses questions juridiques qui restent souvent sans réponse. Les menaces pour la vie privée et la sécurité, comme les cyberattaques et les fuites de données, peuvent causer de graves problèmes aux agriculteurs. Malheureusement, de nombreuses exploitations agricoles sont vulnérables à ces menaces.

2.1.5 Combinaison de l'IA avec d'autres technologies

Comme nous l'avons mentionné, l'IA ne peut pas exister sans d'autres technologies déjà en place, comme le big data, les capteurs et les logiciels. De même, les autres technologies ont besoin de l'IA pour fonctionner correctement. Par exemple, dans le cas du big data, les données en elles-mêmes ne sont pas particulièrement utiles. Ce qui compte en réalité, c'est la manière dont elles sont traitées et leur pertinence. Le moment, le lieu et les critères de sélection déterminent tous si les recommandations de l'IA basées sur un ensemble de données vont être utiles. C'est pourquoi il est également important de disposer de bons ingénieurs et analystes de données pour que la technologie de l'IA fonctionne. Parlons plus en détail des utilisations de l'IA dans l'agriculture.

2.1.5.1 Prise de décision éclairée sur des données volumineuses.

Le véritable objectif de la production et de la collecte de données est de les mettre à profit. Dans le domaine de l'agriculture, l'analyse des données peut entraîner des augmentations massives de la productivité et d'importantes économies de coûts. En combinant l'IA avec le big data, les agriculteurs peuvent obtenir des recommandations valables basées sur des informations en temps réel bien triées sur les besoins des cultures. En retour, cela élimine les conjectures et permet des pratiques agricoles plus précises telles que l'irrigation, la fertilisation, la protection des cultures et la récolte.

2.1.5.2 Capteurs IoT pour la capture et l'analyse des données

Les agriculteurs peuvent utiliser des capteurs IoT et d'autres technologies de soutien (par exemple, des drones, des SIG (Système d'information géographique) et d'autres outils) pour surveiller, mesurer et stocker les données des champs sur une variété de paramètres en temps réel. En combinant des outils agricoles d'IA avec des appareils et des logiciels IoT, les agriculteurs peuvent obtenir des informations plus précises plus rapidement. De meilleures données signifient de meilleures décisions et moins de temps et d'argent consacrés aux essais et aux erreurs.

2.1.6 Réalité et attentes de l'IA pour une agriculture durable

Les avantages de l'IA dans l'agriculture sont indéniables. Les outils agricoles intelligents peuvent effectuer de petites tâches répétitives et chronophages afin que les ouvriers agricoles puissent utiliser leur temps pour des opérations plus stratégiques qui requièrent de l'intelligence humaine. Cependant, il est important de réaliser que, contrairement à un tracteur, on ne peut pas simplement acheter l'IA et la mettre en marche.

L'IA n'est pas quelque chose de concret et matériel. Il s'agit d'un ensemble de technologies qui sont automatisées par la programmation.

L'IA est essentiellement une simulation de la pensée; il s'agit d'apprendre et de résoudre des problèmes à partir de données. L'IA n'est que l'étape suivante dans le développement de l'agriculture intelligente, et elle a besoin d'autres technologies pour fonctionner réellement. En d'autres termes, pour profiter de tous les avantages de l'IA, les agriculteurs ont d'abord besoin d'une infrastructure technologique. Il faudra un certain temps, voire des années, pour développer cette infrastructure. Mais en faisant cela, les agriculteurs seront en mesure de construire un écosystème technologique robuste qui résistera à l'épreuve du temps.

Pour l'instant, les fournisseurs de technologies doivent réfléchir à quelques points : comment améliorer leurs outils, comment aider les agriculteurs à relever leurs défis, et comment faire comprendre facilement et de manière compréhensible que l'apprentissage automatique aide à résoudre des luttes réelles, comme la réduction du travail manuel. L'avenir de l'IA dans l'agriculture ne peut qu'être fructueux.

2.2 Partie 2 : Régression et classification — Apprentissage automatique supervisé

2.2.1 Introduction

Les scientifiques des données utilisent de nombreux types d'algorithmes d'apprentissage automatique pour découvrir des modèles dans les données volumineuses qui mènent à des informations exploitables. À un niveau élevé, ces différents algorithmes peuvent être classés en deux groupes en fonction de la façon dont ils apprennent sur les données pour faire des prédictions : l'apprentissage supervisé et non supervisé. Etant donné que l'apprentissage non supervisé n'est pas le sujet de notre recherche, on va s'intéresser seulement sur le cas supervisé.

L'apprentissage supervisé consiste à disposer de variables d'entrée (x) et d'une variable de sortie (Y) et à utiliser un algorithme pour apprendre la fonction de correspondance entre l'entrée et la sortie $Y = f(X)$. L'objectif est d'approcher la fonction de correspondance si bien que, lorsque vous disposez de nouvelles données d'entrée (x), vous pouvez prédire les variables de sortie (Y) pour ces données. L'une des principales caractéristiques des algorithmes d'apprentissage supervisé est qu'ils modélisent les dépendances et les relations entre la sortie cible et les caractéristiques d'entrée pour prédire la valeur des nouvelles données.

Les techniques d'algorithmes d'apprentissage automatique supervisé comprennent la régression linéaire et logistique, la classification multi-classes, les arbres de décision et les machines à vecteurs de support. L'apprentissage supervisé exige que les données utilisées pour former l'algorithme soient déjà étiquetées avec des réponses correctes. Les problèmes d'apprentissage supervisé peuvent être regroupés en problèmes de régression et de classification. Les deux problèmes ont pour objectif la construction d'un modèle succinct qui peut prédire la valeur de l'attribut dépendant à partir des variables d'attribut. La différence entre ces deux tâches réside dans le fait que l'attribut dépendant est numérique pour la régression et catégorique pour la classification.

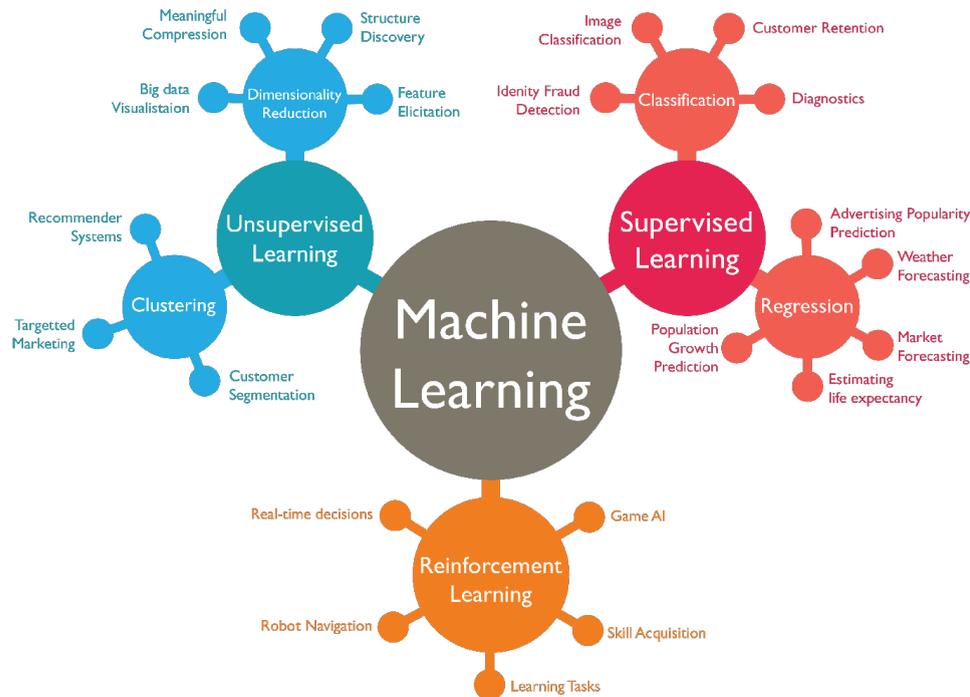


FIGURE 2.2: Les différents types d'apprentissages automatiques.

2.2.2 Types d'algorithmes d'apprentissage supervisé

2.2.2.1 Classification

En apprentissage automatique, la classification est un concept d'apprentissage supervisé qui permet de classer un ensemble de données en classes. Il peut s'agir d'un problème de classification binaire ou d'un problème multi-classes.

Les algorithmes de classification utilisés dans l'apprentissage automatique utilisent des données d'entraînement en entrée dans le but de prédire la probabilité que les données qui suivent entrent dans l'une des catégories prédéterminées. En résumé, la classification est une sorte de reconnaissance des formes. Ici, les algorithmes de classification appliqués aux données d'apprentissage trouvent le même modèle dans les ensembles de données futurs.

Il y a un tas d'algorithmes d'apprentissage automatique pour la classification, nous pouvons citer les plus utilisés : Logistic Regression, Naive Bayes Classifier, Stochastic Gradient Descent, K-Nearest Neighbor ($K - NN$), Decision Tree (DT), Random

Forest (*RF*), Neural Networks (*NN*), Bagging, Support Vector Machine (*SVM*), etc [?].

Dans [52], les auteurs ont fait une étude comparative des algorithmes de classification et de leur performance dans la prédiction du rendement en agriculture de précision. Les algorithmes utilisés sont : SVM, RF, NN, REPTree, Bagging, Naive Bayes Classifier.

les résultats de cette étude ont montré que l’algorithme de classification Bagging, est le mieux adapté à la prédiction du rendement.

2.2.2.2 Régression

La régression sert à trouver la relation d’une variable par rapport à une ou plusieurs autres. Dans l’apprentissage automatique, le but de la régression est d’estimer une valeur (numérique) de sortie à partir des valeurs d’un ensemble de caractéristiques en entrée. Par exemple, estimer le prix d’une maison en se basant sur sa surface, nombre des étages, son emplacement, etc. Donc, le problème revient à estimer une fonction de calcul en se basant sur des données d’entraînement.

$$\hat{y}(x) = f(x_1, x_2, \dots, x_m) \quad (2.1)$$

Il existe de nombreux types d’analyse de régression, dont quelques algorithmes seront examinés ci-dessous [?].

Régression linéaire Il s’agit de l’un des algorithmes de régression les plus utilisés en apprentissage automatique. Une variable significative de l’ensemble des données est choisie pour prédire les variables de sortie (valeurs futures). L’algorithme de régression linéaire est utilisé si les étiquettes sont continues.

Les algorithmes de régression linéaire supposent qu’il existe une relation linéaire entre l’entrée et la sortie. Si les variables dépendantes et indépendantes ne sont pas tracées sur la même ligne dans la régression linéaire, il y aura une perte de rendement.

Dans [53], les auteurs ont utilisé la régression linéaire pour construire un modèle de prévision de la croissance des cultures en phase végétative pour soutenir l'application de l'agriculture de précision en usine et ont pu arriver à un score r^2 de 0.9875

Régression Ridge La régression ridge est un autre algorithme de régression linéaire très utilisé en apprentissage automatique. Si une seule variable indépendante est utilisée pour prédire la sortie, on parlera d'un algorithme MLR (multiple linear regression) . La régression Ridge minimise la perte rencontrée dans la régression linéaire.

La complexité du modèle MLR peut également être réduite par la régression ridge. Il convient de noter que tous les coefficients ne sont pas réduits dans la régression ridge, mais qu'ils le sont dans une plus large mesure que dans les autres modèles.

Dans [54], l'étude menée par les auteurs a permis de développer des modèles de prévision pour la superficie et la production de blé dans le district de Chakwal de la région de Rawalpindi en tenant compte des hypothèses de l'estimation des MCO. Ces modèles linéaires peuvent être utilisés pour prévoir la superficie et la production de blé pour les cinq prochaines années. L'engrais Urée, l'engrais DAP et le fumier jouent un rôle important dans l'amélioration de la production de blé. Le nombre de charrues dans les champs de blé est un facteur important pour augmenter la production de blé. De bonnes pluies au cours des mois d'octobre et de novembre contribuent de manière significative à l'augmentation de la production de blé et la température maximale moyenne au cours du mois de mars est un facteur important de réduction de la production de blé.

Least Absolute Shrinkage and Selection Operator Regression — Régression LASSO La régression LASSO est une autre régression multi-linéaire, largement utilisée (une variable d'entrée). La somme des valeurs des coefficients est pénalisée dans la régression LASSO pour éviter les erreurs de prédiction. Les coefficients de détermination dans la régression LASSO sont réduits à zéro en utilisant la technique du "shrinkage". Les coefficients de régression sont réduits par la régression LASSO pour qu'ils s'adaptent parfaitement à divers ensembles de données.

L'algorithme de régression LASSO est utilisé lorsqu'il existe une forte multi-colinéarité dans l'ensemble de données donné. La multi-colinéarité dans l'ensemble de données signifie que les variables indépendantes sont fortement liées les unes aux autres, et qu'un petit changement dans les données peut entraîner un changement important dans les coefficients de régression. La régression par l'algorithme Lasso peut être utilisée dans les applications de prévision en ML.

Dans [55], les auteurs ont mis en place un système de détection des éléments nutritifs du sol pour l'agriculture de précision à l'aide de la spectroscopie par claquage induit par laser (LIBS) et de méthodes de régression multivariées (PLSR, Lasso et GPR).

Régression par arbre de décision La régression non linéaire en apprentissage automatique peut être effectuée à l'aide de la régression par arbre de décision. La fonction principale de l'algorithme de régression par arbre de décision est de diviser l'ensemble de données en plus petits ensembles. Les sous-ensembles de l'ensemble de données sont créés pour tracer la valeur de tout point de données qui se rapporte à l'énoncé du problème. La division de l'ensemble de données par cet algorithme donne lieu à un arbre de décision comportant des nœuds de décision et des nœuds de feuille.

Il faut savoir que même un léger changement dans les données peut entraîner un changement majeur dans la structure de l'arbre de décision qui en résulte. Il ne faut pas non plus trop réduire les variables de l'arbre décisionnel, car il ne restera plus assez de nœuds finaux pour effectuer la prédiction. Pour avoir plusieurs nœuds finaux (valeurs de sortie de la régression), il ne faut pas élaguer excessivement les variables (régresseurs) de l'arbre décisionnel.

Dans [56], les auteurs ont montré que les modèles d'arbres de décision de l'approche d'exploration de données CART sont des outils assez efficaces pour la détection du stress hydrique, azoté et des mauvaises herbes dans les cultures de maïs.

Forêt aléatoire La forêt aléatoire est également un algorithme largement utilisé pour la régression non linéaire en apprentissage automatique. Contrairement à la régression par arbre de décision (arbre unique), une forêt aléatoire utilise plusieurs arbres de décision pour prédire la sortie. Des points de données aléatoires sont

sélectionnés dans l'ensemble de données donné (disons que k points de données sont sélectionnés), et un arbre de décision est construit avec eux via cet algorithme. Plusieurs arbres de décision sont ensuite modélisés pour prédire la valeur de tout nouveau point de données.

Comme il existe plusieurs arbres de décision, plusieurs valeurs de sortie seront prédites via un algorithme de forêt aléatoire. Vous devez trouver la moyenne de toutes les valeurs prédites pour un nouveau point de données afin de calculer la sortie finale. Le seul inconvénient de l'utilisation d'un algorithme de forêt aléatoire est qu'il nécessite plus de données en termes de formation. Cela se produit en raison du grand nombre d'arbres de décision mappés sous cet algorithme, car il nécessite une plus grande puissance de calcul.

Par exemple, une approche d'exploration de données à l'aide des forêts aléatoires a permis d'expliquer la productivité de la canne à sucre dans les Tropiques humides, en Australie. Facilitant ainsi les décisions de l'exploitation, comme la quantité d'engrais azotés à appliquer et aider les meuniers à planifier soigneusement l'entretien et de travail pour être prêts pour le début de la saison de broyage [57].

Méthode des k plus proches voisins (k-Nearest Neighbours — KNN) Le modèle KNN est couramment utilisé pour la régression non linéaire en apprentissage automatique. KNN est une approche facile à mettre en œuvre pour la régression non linéaire en apprentissage automatique. KNN suppose que le nouveau point de données est similaire aux points de données existants. Le nouveau point de données est comparé aux catégories existantes et est placé dans une catégorie relative. La valeur moyenne des k voisins les plus proches est prise comme entrée dans cet algorithme. Les voisins dans les modèles KNN reçoivent un poids particulier qui définit leur contribution à la valeur moyenne.

Une pratique courante pour attribuer des poids aux voisins dans un modèle KNN est $1/d$, où d est la distance du voisin par rapport à l'objet dont la valeur doit être prédite. Pour déterminer la valeur d'un nouveau point de données via le modèle KNN, il faut savoir que les voisins les plus proches contribueront davantage que les voisins éloignés.

Dans [58], les auteurs ont utilisé $K - NN$ Afin de prédire le rendement des cultures et ils sont parvenus à des résultats convaincants.

Régression par vecteur de support (Support Vector Regression — SVR)

Les SVM peuvent être classés dans les types de régression linéaire et non linéaire de l'apprentissage automatique. L'algorithme SVM est utilisé lorsqu'il faut identifier la sortie dans un espace multidimensionnel. Les points de données sont représentés comme un vecteur dans un espace multidimensionnel.

Ce modèle crée un hyperplan à marge maximale qui sépare les classes et attribue une valeur à chacune d'elles. Le modèle SVM n'est pas aussi performant que prévu lorsque l'ensemble de données contient plus de bruit.

Dans [59], les auteurs ont mis en œuvre un système d'irrigation intelligent basé sur Régression par vecteur de support (SVR - support vector regression).

Régression sur composantes principales (Principal component regression — PCR)

La régression en composantes principales est une extension de l'analyse en composantes principales et de la régression linéaire multiple. La PCR modélise une variable cible lorsqu'il existe un grand nombre de variables prédictives et que ces prédictives sont fortement corrélés, ou bien même colinéaires. Cette méthode construit de nouvelles variables prédictives, appelées composantes, en tant que combinaisons linéaires des variables prédictives originales. La PCR crée des composantes pour expliquer la variabilité observée dans les variables prédictives, sans tenir compte du tout de la variable cible.

Dans la première étape, les composantes principales sont calculées. Les scores des composantes principales les plus importantes sont utilisés comme base pour la régression linéaire multiple avec la variable cible. Le point le plus important de la PCR est la sélection correcte des vecteurs propres à inclure. Un graphique des valeurs propres indique généralement le "meilleur" nombre de vecteurs propres. L'avantage de la PCR par rapport à la régression linéaire multiple est que le bruit reste dans les résidus puisque les vecteurs propres à faible valeur propre ne représentent que des parties des données à faible variance. De plus, les coefficients de régression sont plus

stables. Cela est dû au fait que les vecteurs propres sont orthogonaux les uns par rapport aux autres.

Dans [60], les auteurs ont fait une comparaison entre les méthodes de régression en composantes principales (PCR), de régression partielle des moindres carrés (PLSR) et de réseau neuronal à propagation arrière (BPNN) pour la précision de prédiction du carbone organique, des formes extractibles de K, Na, Mg et P à l'aide de la spectroscopie de réflectance diffuse dans le visible et le proche infrarouge a permis de tirer la conclusion suivante : Les modèles BPNN-variables latentes ont surpassé les modèles PLSR et BPNN-PCs pour toutes les propriétés du sol.

Régression polynomiale La régression polynomiale est le même concept que la régression linéaire sauf qu'elle utilise une ligne courbe au lieu d'une ligne droite (qui est utilisée par la régression linéaire). La régression polynomiale utilise plus de paramètres pour tracer une ligne de régression non linéaire. Elle est avantageuse pour les données qui ne peuvent pas être résumées par une ligne droite. Le nombre de paramètres (également appelés degrés) doit être déterminé. Un modèle à degré plus élevé est plus complexe, mais peut s'adapter excessivement aux données.

Dans [61], les auteurs ont mis en œuvre des modèles de régression polynomiale afin d'obtenir des informations à partir des données générées par le réseau pour la reconnaissance des formes et les tâches de prédiction. Cela permettra à l'agriculteur de réduire le coût global du système central et de donner un aperçu des données directement sur les nœuds cibles.

2.3 Partie 3 : Problématique et algorithmes choisis

2.3.1 Problématique et algorithmes utilisés

Le déploiement d'un RCSF dans une parcelle agricole offre un support important pour une gestion précise des ressources, afin d'optimiser le processus d'irrigation, le suivi/prévention des maladies ainsi que le moment adéquat de la récolte améliorant par conséquent la production agricole vis à vis les informations fournies par les éléments qui composent ce réseau.

Ces capteurs pourront nous octroyer des informations sur des paramètres physiques importants tels que la température, l'humidité, le pH ...etc Néanmoins, le nombre de capteurs utilisés dans ce réseau, doit être minime pour des raisons financières. Il existe donc des points non échantillonnés par ces capteurs.

Notre problématique ici consiste ainsi à construire un modèle continu à partir des données discrètes des capteurs. C'est à dire, à partir des paramètres physiques tirés des quelques capteurs déployés, il nous faut les extrapoler afin d'avoir une estimation de leur distribution sur toute la surface d'intérêt.

Le modèle ainsi obtenu nous permettra d'estimer la valeur des paramètres physiques à n'importe quel point de la parcelle agricole.

Ce modèle doit donc être robuste, mais assez rapide d'exécution puisque l'agriculture de précision nécessite un entretien constant et méticuleux. Ce modèle doit aussi nécessiter peu de données pour donner des résultats satisfaisants compte tenu de la contrainte du nombre de capteurs.

Pour ces raisons, nous avons opté pour les algorithmes ci-dessous.

2.3.1.1 Régression linéaire

L'analyse de régression est l'une des méthodes statistiques les plus importantes pour la recherche appliquée depuis de nombreuses décennies, capable de décrire la relation entre deux ou plusieurs variables. Le modèle de régression linéaire univarié étudie la relation linéaire entre une variable dépendante et une seule variable indépendante. Dans le modèle de régression linéaire multivarié, la variable dépendante Y est modélisée comme une fonction linéaire des variables indépendantes X_i

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + \epsilon \quad (2.2)$$

où y est la variable dépendante à valeur réelle, X_i sont des variables indépendantes à valeur réelle, a est une constante (appelée intercept) et b_i sont les coefficients de régression de la variable dépendante X_i . Le ϵ est le résidu de moyenne nulle qui ne peut être exprimé par les variables du modèle.

L'objectif de cette méthode est d'estimer les paramètres du modèle sur la base des n ensembles de valeurs observées et en appliquant une certaine fonction de critère. Une telle fonction de critère,

$$\phi = \sum_{i=1}^n (y_i - Y_i)^2 \quad (2.3)$$

Où y_i sont les valeurs observées et Y_i sont les valeurs à estimer à partir des n observations, peut être résolu en utilisant la méthode bien connue d'estimation par les moindres carrés.

2.3.1.2 Arbres de décision (Decision Tree — DT)

Les arbres de décision sont les algorithmes d'apprentissage automatique les plus utilisés en géosciences [62], [63].

L'utilisation croissante des arbres de décision est liée à leur simplicité et à leur interprétabilité, à leur faible coût de calcul et à la possibilité d'être représentés graphiquement. Un arbre de décision représente un ensemble de restrictions ou de conditions qui sont organisées de manière hiérarchique, et qui sont appliquées successivement d'une racine à un nœud terminal ou à une feuille de l'arbre [64].

Le principal avantage de l'utilisation d'une structure arborescente hiérarchique pour effectuer des décisions de classification ou régression est que la structure arborescente est transparente. Afin d'induire les arbres de décision à partir d'un ensemble de données, une mesure d'évaluation de chacun des éléments probants est utilisée pour maximiser l'hétérogénéité entre les nœuds. On peut distinguer deux méthodologies différentes au sein des DT : les arbres de classification et les arbres de régression (RT). Afin d'induire les arbres de décision, un partitionnement récursif et des régressions multiples sont effectués à partir de l'ensemble de données. À partir du nœud racine, le processus de partitionnement des données dans chaque nœud interne d'une règle de l'arbre est répété jusqu'à ce qu'une condition d'arrêt préalablement spécifiée soit atteinte. À chacun des nœuds terminaux, ou feuilles, est attaché un modèle de régression simple qui ne s'applique qu'à ce nœud. Une fois le processus d'induction de l'arbre terminé, l'élagage peut être appliqué dans le but d'améliorer la capacité de généralisation de l'arbre en réduisant sa complexité structurelle. Le nombre de cas dans les nœuds peut être considéré comme un critère d'élagage.

l'induction du DT implique d'abord la sélection de vecteurs de mesure de fractionnement optimaux. Le processus commence par la division de la caractéristique dépendante, ou du nœud parent (racine), en morceaux binaires, où les nœuds enfants sont plus "purs" que le nœud parent. Grâce à ce processus, les DT recherchent tous les fractionnements candidats pour trouver le fractionnement optimal, s^* , qui maximise la pureté de l'arbre résultant (définie par la plus grande diminution de l'impureté).

$$\Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2.4)$$

Dans cette équation, s est la scission candidate au nœud t , puis le nœud t est divisé par s en un nœud enfant gauche t_L avec une proportion de p_L , et un nœud enfant droit t_R avec une proportion de p_R . $i(t)$ est une mesure de l'impureté avant la scission, $i(t_L)$ et $i(t_R)$ sont des mesures de l'impureté après la scission, et $\Delta i(s, t)$ mesure la diminution de l'impureté à partir de la scission s .

Il existe de nombreuses approximations pour mesurer l'impureté. Certaines des plus fréquentes sont le ratio de gain [65], l'indice de Gini [66] et le chi carré [67]. La mesure la plus courante est l'indice de Gini. L'indice de Gini utilisé dans cette recherche mesure $I_G(t)$ comme :

$$I_G(t_{X(x_i)}) = 1 - \sum_{j=1}^m f(t_{X(x_i)}, j)^2 \quad (2.5)$$

Où $f(t_{X(x_i)}, j)$ est la proportion d'échantillons avec la valeur x_i appartenant à la feuille j au nœud t . Le critère de division de l'arbre de décision est basé sur le choix de l'attribut avec l'indice d'impureté de Gini I_G le plus bas.

2.3.1.3 Forêts aléatoires (Random Forest — RF)

La forêt aléatoire est une technique de régression qui combine les performances de nombreux algorithmes d'arbres de décision pour classifier ou prédire la valeur d'une variable [68]. C'est-à-dire que lorsque la forêt aléatoire reçoit un vecteur d'entrée (x), composé des valeurs des différentes caractéristiques probantes analysées pour une zone d'entraînement donnée, RF construit un nombre K d'arbres de régression et fait la

moyenne des résultats. Après que K arbres de ce type $T(X)_1^k$ sont formés, le prédicteur de régression RF est le suivant :

$$\hat{f}_{r,f^K}(x) = \frac{1}{K} \sum_{k=1}^K T(x) \quad (2.6)$$

Pour éviter la corrélation des différents arbres, RF augmente la diversité des arbres en les faisant croître à partir de différents sous-ensembles de données d'entraînement créés par une procédure appelée bagging. Le bagging est une technique utilisée pour la création de données d'entraînement en ré-échantillonner aléatoirement l'ensemble de données d'origine avec remplacement, c'est-à-dire sans suppression des données sélectionnées dans l'échantillon d'entrée pour générer le sous-ensemble suivant $h(x, \Theta k)$, $k = 1, \dots, K$, où Θk sont des vecteurs aléatoires indépendants avec la même distribution. Ainsi, certaines données peuvent être utilisées plus d'une fois dans l'apprentissage, tandis que d'autres peuvent ne jamais être utilisées. Ainsi, une plus grande stabilité est obtenue, car elle le rend plus robuste face à de légères variations des données d'entrée et, en même temps, elle augmente la précision de la prédiction. D'autre part, lorsque la RF fait croître un arbre, elle utilise la meilleure caractéristique/le meilleur point de séparation dans un sous-ensemble de caractéristiques probantes qui a été sélectionné de manière aléatoire dans l'ensemble des caractéristiques probantes d'entrée. Par conséquent, cela peut diminuer la force de chaque arbre, mais cela réduit la corrélation entre les arbres, ce qui réduit l'erreur de généralisation [69].

Une autre caractéristique intéressante est que les arbres d'un RF croissent sans élagage, ce qui les rend légers, d'un point de vue informatique.

En outre, les échantillons qui ne sont pas sélectionnés pour la formation du k -ième arbre dans le processus de mise en sac sont inclus dans un autre sous-ensemble appelé out-of-bag (OOB). Ces éléments OOB peuvent être utilisés par le k -ième arbre pour évaluer les performances. De cette façon, RF peut calculer une estimation non biaisée de l'erreur de généralisation sans utiliser un sous-ensemble externe de données textuelles [70], [69].

L'erreur de généralisation converge au fur et à mesure que le nombre d'arbres augmente ; par conséquent, la RF ne s'ajuste pas trop aux données. La RF fournit

également une évaluation de l'importance relative des différentes caractéristiques. Cet aspect est utile pour les études multi-sources, où la dimensionnalité des données est très élevée, et il est important de savoir comment chaque caractéristique influence le résultat. Il est important de savoir comment chaque caractéristique influence le modèle de prédiction pour pouvoir sélectionner les meilleurs éléments probants [71].

Pour évaluer l'importance de chaque variable (par exemple, la bande de l'image satellite), le RF change l'un des éléments probants d'entrée tout en gardant les autres constants et il mesure la diminution de la précision qui a eu lieu au moyen de l'estimation de l'erreur oob.

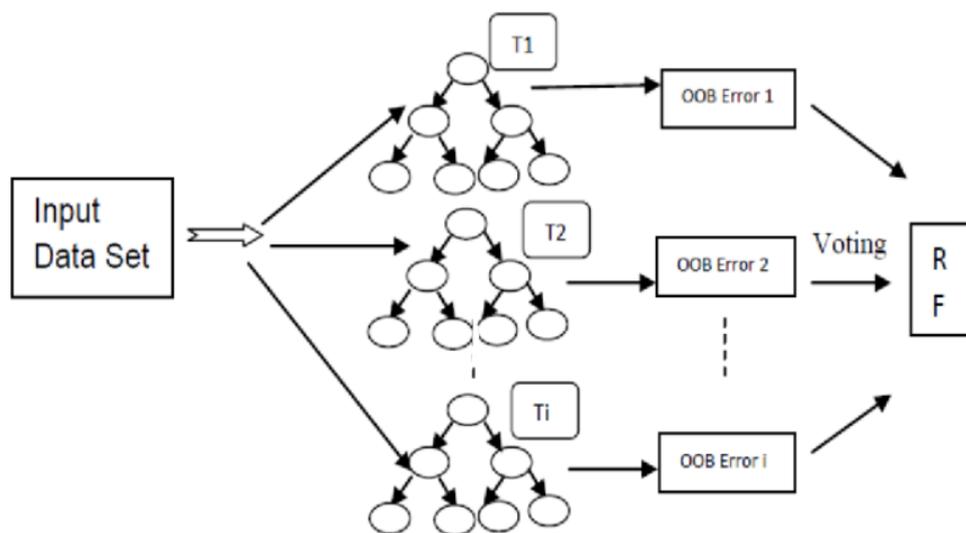


FIGURE 2.3: Diagramme de fonctionnement de la forêt aléatoire.

2.3.1.4 XGBoost

XGBoost a été largement utilisé dans de nombreux domaines pour obtenir des résultats de pointe sur certains problèmes de données. Il s'agit d'un système d'apprentissage automatique évolutif très efficace pour le boosting d'arbres [72], [73].

XGBoost est optimisé dans le cadre du Boosting de Gradient et développé par Chen et Guestrin [74], qui est conçu pour être hautement efficace, flexible et portable. L'idée principale du boosting est de combiner une série de régresseurs faibles avec une faible précision pour construire un régresseur fort avec une meilleure performance de prédiction. Si l'apprenant faible pour chaque étape est basé sur la direction du gradient de la fonction de perte, on peut l'appeler les machines de boosting à gradient.

En supposant qu'un ensemble de données est $D = [(x_i, y_i) : i = 1 \dots n, x_i \in \mathbb{R}^m]$, nous avons n échantillons avec m caractéristiques. Soit y_i la valeur prédite par le modèle :

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2.7)$$

où f_k représente un arbre de régression indépendant et $f_k(x_i)$ désigne le score de prédiction donné par le k - ième arbre au i - ième échantillon. L'ensemble des fonctions f_k dans le modèle d'arbre de régression peut être appris en minimisant la fonction objective :

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2.8)$$

Le terme l est une fonction de perte d'apprentissage, qui mesure la différence entre la prédiction \hat{y} et l'objet y_i . Pour éviter le surajustement, ce terme pénalise la complexité du modèle :

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (2.9)$$

Où γ et λ sont les degrés de régularisation. T et w sont respectivement le nombre de feuilles et les scores de chaque feuille.

Le modèle d'ensemble d'arbres peut être entraîné de manière additive. Si \hat{y}_i^t est la prédiction de la i - ième instance à la t - ième itération, il faut ajouter f_t pour minimiser l'objectif suivant :

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (2.10)$$

en utilisant l'expansion de Taylor du second ordre sur la fonction $Obj^{(t)}$, on obtient une fonction simplifiée et sans terme constant :

$$Obj^{(t)} = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2] + \Omega(f_t) \quad (2.11)$$

Où $g_i = \frac{\partial}{\partial \hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ et $h_i = \frac{\partial^2}{\partial (\hat{y}_i^{(t-1)})^2} l(y_i, \hat{y}_i^{(t-1)})$ Sont les gradients de premier et de second ordre sur l . Ensuite, la fonction objective est réécrite comme suit :

$$Obj^t = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 = \sum_{j=1}^T [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma T \quad (2.12)$$

Où $I_j = [i : q(x_i) = j]$ désigne l'ensemble d'instances de la feuille j . Pour une structure arborescente q fixe, le poids optimal w_j^* de la feuille j et la valeur optimale correspondante peuvent être calculés par :

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (2.13)$$

$$Obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \lambda T \quad (2.14)$$

Où $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$, obj présentent la qualité d'une structure arborescente q . Plus la valeur est petite, meilleure est la structure de l'arbre. Puisqu'il est impossible d'énumérer toutes les structures d'arbre, un algorithme gourmand est utilisé pour ajouter des branches à l'arbre de manière itérative. I_L et I_R sont les ensembles d'instances des nœuds de gauche et de droite après la division. En énumérant les points de segmentation réalisables et en sélectionnant la fonction cible minimale et la partition de gain maximal, la formule de gain est présentée comme suit :

$$G = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} + \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] \quad (2.15)$$

Cette formule est généralement utilisée dans la pratique pour évaluer les candidats au fractionnement.

Le modèle XGBoost produit de nombreux arbres simples, qui sont utilisés pour évaluer un nœud de feuille lors du fractionnement. Les premier, deuxième et troisième termes de l'équation représentent respectivement le score sur la gauche, la droite et la feuille d'origine. De plus, le terme γ est la régularisation sur la feuille supplémentaire. Il sera utilisé dans le processus de formation.

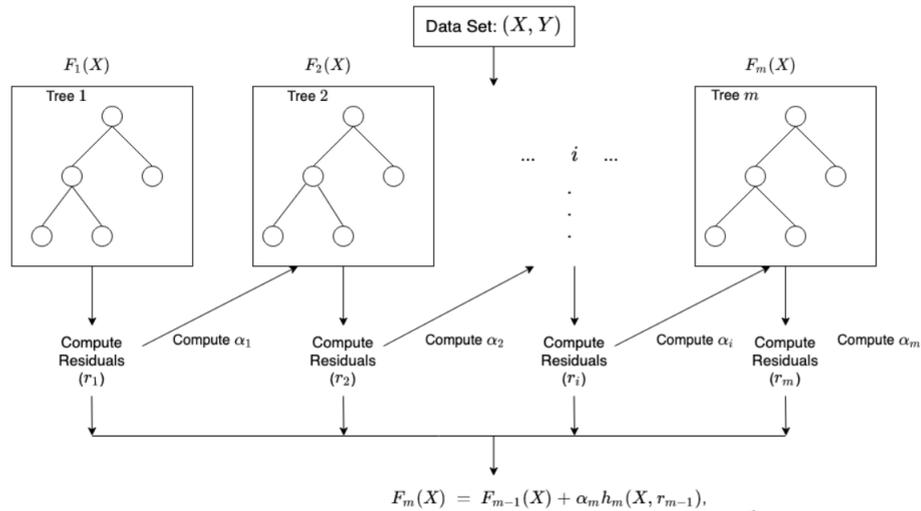


FIGURE 2.4: Diagramme de fonctionnement de XGBoost.

2.4 Conclusion

Nous venons à travers ce chapitre d'introduire l'IA et son rôle inéluctable dans l'AP, ainsi que les contraintes actuelles à son adoption. On a ensuite introduit les différents algorithmes d'apprentissage automatique supervisé en se focalisant sur les trois algorithmes de régression qui vont être utilisés afin de résoudre notre problématique.

Chapitre 3

Implémentation et résultats

3.1 Introduction

La production agricole suit de fortes tendances saisonnières liées au cycle de vie biologique des cultures. Il faut néanmoins mettre en place des systèmes de surveillance nous permettant d'avoir des données fréquentes sur les différents paramètres physiques ayant une influence sur cette production cela dans le but de pouvoir agir en temps voulu pour maximiser la récolte.

Pour pouvoir réagir au moment opportun, l'utilisation des RCSFs est indéniable, puisqu'ils nous permettent de collecter des données sur les paramètres physiques de la surface agricole à de courts intervalles réguliers. Le nombre de capteurs utilisés reste toutefois un problème notable à ne pas négliger pour minimiser les couts et ainsi faciliter l'adoption de cette solution dans les pays en voie de développement.

Comme vu précédemment (chapitre 1, tableau 1.4) différents capteurs avec différentes caractéristiques sont en commercialisation.

Lors de l'acquisition et du déploiement de ces capteurs à différents points de la surface d'intérêt, ce qui laisse la grande majorité de la surface non échantillonnée, ne nous donnant pas ainsi les informations nécessaire pour agir en temps voulu.

Il faut donc construire un modèle capable d'extrapoler les données du sol et fournir des prévisions précises dans ces emplacements non échantillonnés, ainsi que donner une incertitude concernant ces prédictions. Pour répondre à ce problème, nous suggérons l'utilisation d'une modélisation à base d'algorithmes d'apprentissage automatique, à savoir : la régression linéaire, les forêts aléatoires ainsi que XGBoost.

Ainsi, dans ce chapitre, nous exposant la modélisation faite à travers ces différents algorithmes, avec une comparaison entre eux, le gain obtenu à travers l'ajustement de leurs hyper paramètres ainsi que l'influence du nombre de points échantillonnés sur l'efficacité du modèle.

3.2 Description de la base de données

La validation des modèles se fera à travers des données disponibles en ligne qui sont fournies par le département des sciences des cultures et des sols, université de Washington [75]. Ces données sont acquises à partir de 42 capteurs éparpillées sur une

parcelle agricole de 1000*600 m², chaque capteur contient des informations sur l'eau volumétrique et la température au point échantillonné à différentes profondeurs (30 cm, 60 cm, 90 cm, 120 cm et 150 cm), de plus, la localisation du capteur est aussi donnée.

Data set	Unit	Moyenne	Min	Max	Ecart type	Asymétrie(skewness)
VW_30cm	m ³ /m ³	0.1825	0.112	0.238	0.03	-0.3
VW_60cm	m ³ /m ³	0.1913	0.108	0.316	0.04	0.3836
VW_90cm	m ³ /m ³	0.217	0.103	0.402	0.055	0.32
VW_120cm	m ³ /m ³	0.252	0.117	0.409	0.065	0.468
VW_150cm	m ³ /m ³	0.272	0.116	0.481	0.082	0.513
T_30cm	C°	13.54	11.2	15.36	1.05	-0.28
T_60cm	C°	14.5	12.74	17.19	1.177	0.36
T_90cm	C°	14.81	13.3	17.06	1.03	0.49
T_120cm	C°	14.83	12.96	17.56	1.16	0.6
T_150cm	C°	14.6	13	18.88	1.32	1.26

TABLE 3.1: Résumé de quelques données statistiques du data set

Le tableau 3.1 résume les propriétés statistiques des données d'entrées qui sont la température et l'humidité à différentes profondeurs (30cm, 60cm, 90cm, 120cm et 150cm), il est apparent que les valeurs montrent une distribution normale (les valeurs d'asymétrie(skewness) sont proches de 0). La variation de l'humidité (VW) augmente à chaque profondeur(valeurs de l'écart-type), tandis que la variation de la température a une variation qui est peu dépendante de la profondeur. Les valeurs maximales de la température et de l'humidité ont leurs plus grandes valeurs pour les profondeurs maximales.

3.3 Critères d'évaluation

Afin d'évaluer et avoir une idée de la performance d'un modèle, il faut utiliser des critères d'évaluations adaptés, davantage en Agriculture de Précision où la variable

à prédire est coûteuse et chronophage à acquérir. Les indicateurs choisis et qui sont largement utilisés dans les problèmes de régression sont mentionnés ci-dessous.

3.3.1 Erreur moyenne absolue — MAE

L'erreur absolue moyenne est une mesure d'évaluation de modèle utilisée avec les modèles de régression. L'erreur absolue moyenne d'un modèle par rapport à un ensemble de tests est la moyenne des valeurs absolues des erreurs de prédiction individuelles sur toutes les instances de l'ensemble de tests. Chaque erreur de prédiction est la différence entre la valeur réelle et la valeur prédite pour l'instance. L'erreur absolue moyenne utilise la même échelle que les données mesurées. Il s'agit d'une mesure de précision qui dépend de l'échelle et qui ne peut donc pas être utilisée pour faire des comparaisons entre des séries utilisant des échelles différentes.

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (3.1)$$

Où y_i est la valeur de la i – ème observation du jeu de données de validation et \hat{y}_i est la valeur prédite pour la i – ème observation.

3.3.2 Erreur quadratique moyenne — RMSE

L'erreur quadratique moyenne $RMSE$ est l'écart type des résidus (erreurs de prédiction). Les résidus sont une mesure de l'éloignement des points de données par rapport à la ligne de régression ; $RMSE$ est une mesure de la dispersion de ces résidus. En d'autres termes, elle indique à quel point les données sont concentrées autour de la ligne de meilleur ajustement. L'erreur quadratique moyenne est couramment utilisée en climatologie, en prévision et en analyse de régression pour vérifier les résultats expérimentaux. Souvent, les valeurs de $RMSE$ sont difficiles à interpréter parce que l'on n'est pas en mesure de dire si une valeur de variance est faible ou forte.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3.2)$$

Où y_i est la valeur de la i – ème observation du jeu de données de validation et \hat{y}_i est la valeur prédite pour la i – ème observation.

3.3.3 Erreur maximale — Max error

La *MaxError* fonction calcule l'erreur résiduelle maximale, une métrique qui capture l'erreur la plus défavorable entre la valeur prédite et la valeur vraie. Dans un modèle de régression à sortie unique parfaitement ajusté, *MaxError* serait 0 sur l'ensemble d'apprentissage et bien que cela soit hautement improbable dans le monde réel, cette métrique montre l'étendue de l'erreur du modèle lors de l'ajustement.

$$Max_Error(y, \hat{y}) = \max(|y_i - \hat{y}_i|) \quad (3.3)$$

Où y_i est la valeur de la i -ème observation du jeu de données de validation et \hat{y}_i est la valeur prédite pour la i -ème observation.

3.3.4 Variance expliquée — Explained variance score

La variance expliquée mesure la proportion dans laquelle un modèle mathématique tient compte de la variation (dispersion) d'un ensemble de données. Souvent, la variation est quantifiée comme la variance; ensuite, le terme plus spécifique de variance expliquée peut-être utiliser.

$$explained_variance(y, \hat{y}) = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}} \quad (3.4)$$

\hat{y} est la cible estimée, y la sortie cible correspondante (correcte), et Var est la variance, le carré de l'écart type.

3.4 Méthodologie de comparaison

La variable étudiée ici est la température à la profondeur de 30 cm, et les variables caractéristiques sont : les coordonnées spatiales, les températures et l'humidité du sol à différentes profondeurs. En premier lieu, nous allons prouver que les modèles génèrent de bonnes estimations des valeurs de la variable étudiée. Pour ce faire, il est nécessaire de travailler au moins avec un jeu de données d'apprentissage et un jeu de données de validation. Assez simplement, les données d'apprentissage sont utilisées pour étalonner le modèle tandis que le jeu de validation sert à montrer que le modèle est fiable et

pertinent. Pour être le plus objectif possible, les jeux de données d'apprentissage et de validation devraient être issus de population indépendante (pour être sûr de ne pas biaiser le résultat). D'une manière plus générale, ce qui est fait est de séparer un ensemble d'échantillons initial de base en un jeu de données d'apprentissage et un jeu de données de validation. Il est souvent conseillé d'utiliser entre 60% et 80% du jeu de données initial comme jeu d'apprentissage et les 20% à 40% restants comme jeu de validation. Néanmoins, ces pourcentages ne sont pas fixes. Après une comparaison initiale des différents modèles, nous allons essayer d'optimiser les résultats obtenus à travers un ajustement des hyper-paramètres de ces modèles. Nous verrons ensuite l'effet du nombre de capteurs (compte tenu de la contrainte énoncée en introduction) sur l'efficacité des prévisions.

3.5 Implémentation

3.5.1 Logiciels utilisés

PyCharm est un environnement de développement intégré utilisé pour programmer en Python.

Il permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web.

Développé par l'entreprise tchèque JetBrains, c'est un logiciel multi-plateforme qui fonctionne sous Windows, Mac OS X et Linux. Il est décliné en édition professionnelle, diffusé sous licence propriétaire, et en édition communautaire diffusé sous licence Apache.

Pour l'implémentation des modèles de machine learning choisis à savoir, Régression linéaire multiple, Forêts aléatoires et Extreme gradient boosting, plusieurs packages sont disponibles comme Keras, H2O, Scikit learn et XGBoost. Il est aussi possible d'implémenter les fonctions nécessaires sans avoir recours aux bibliothèques disponibles (efficace pour une implémentation sur carte FPGA par exemple pour optimiser les temps d'exécution). Dans notre cas nous avons opté pour l'utilisation de scikit-learn

et de XGBoost. Ces kits offrent des classes facilitant l'implémentation des différents algorithmes, mais aussi l'ajustement des hyper-paramètres.

Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique. Elle est développée par de nombreux contributeurs notamment dans le monde académique par des instituts français d'enseignement supérieur et de recherche comme Inria.

Elle propose dans son framework de nombreuses bibliothèques d'algorithmes à implémenter. Ces bibliothèques sont à disposition notamment des data scientists.

Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions logistiques, des algorithmes de régression. Elle est conçue pour s'harmoniser avec d'autres bibliothèques libres Python, notamment NumPy et SciPy.

XGBoost est une bibliothèque distribuée optimisée de boosting de gradient conçue pour être hautement efficace, flexible et portable. Elle implémente des algorithmes d'apprentissage automatique dans le cadre du Gradient Boosting. XGBoost fournit un boosting d'arbre parallèle (également connu sous le nom de GBDT, GBM) qui résout de nombreux problèmes de science des données de manière rapide et précise. Le même code fonctionne sur les principaux environnements distribués (Hadoop, SGE, MPI) et peut résoudre des problèmes au-delà de milliards d'exemples.

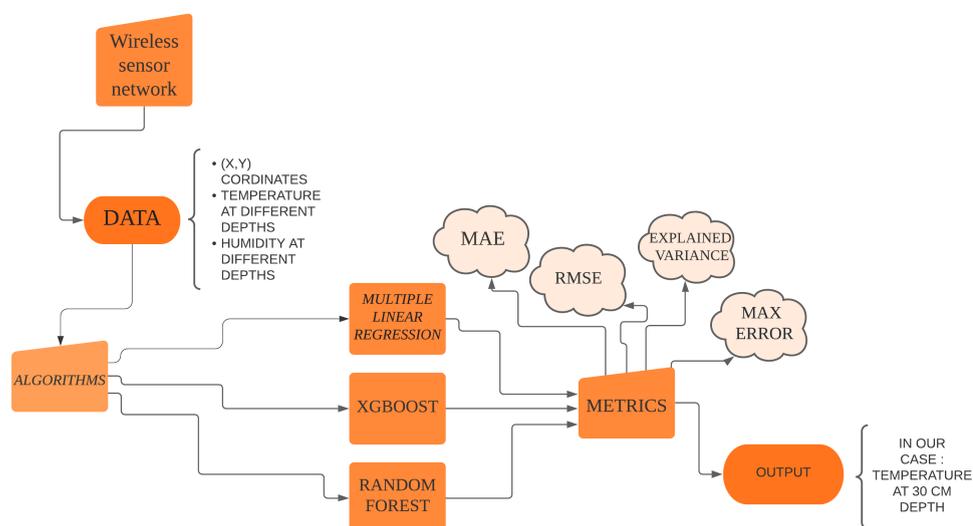


FIGURE 3.1: Diagramme explicatif du processus de traitement de données

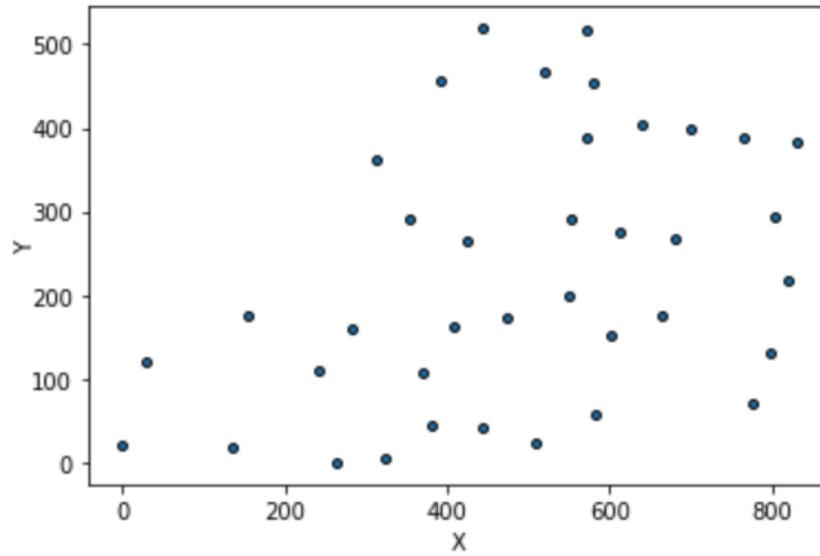


FIGURE 3.2: Emplacement des capteurs.

Dans ce qui suit, les données de calibrage et de validation seront ordonnées de la sorte :

- 80% des observations pour le calibrage.
- 20% pour la validation.

Avant de continuer dans l'implémentation, nous allons visualiser la corrélation entre nos variables caractéristiques.

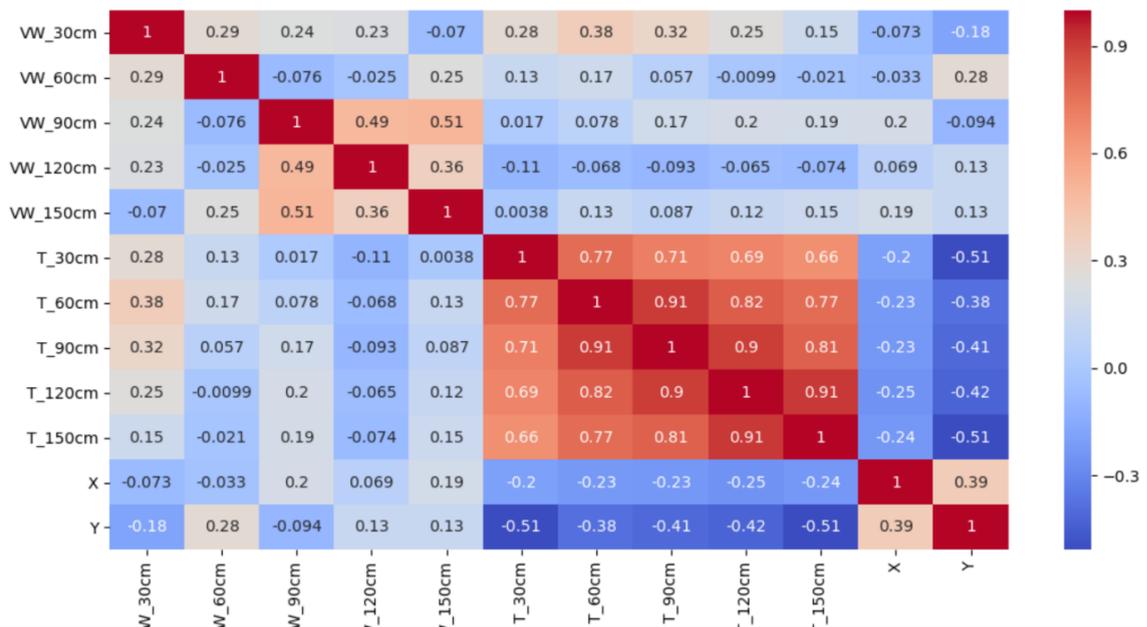


FIGURE 3.3: Matrice de corrélation.

On peut voir de prime abord que notre variable cible (température à 30 *cm*) est en corrélation avec nos variables caractéristiques, notamment les températures aux autres profondeurs, mais aussi à la localisation et l'humidité du sol pour les profondeurs basses (30 *cm* et 60 *cm*).

3.5.2 Régression linéaire multiple

Hyper paramètres par défaut :

- `fit_intercept` — Par défaut : `True` \Rightarrow Calcul ou non de l'intercepte pour ce modèle. Si la valeur est `False`, aucune ordonnée à l'origine ne sera utilisée dans les calculs (c'est-à-dire que les données sont censées être centrées).
- `Normalize` — par défaut : `False` \Rightarrow Ce paramètre est ignoré lorsque `fit_intercept` est défini sur `False`. Si `True`, les régresseurs X seront normalisés avant la régression en soustrayant la moyenne et en divisant par la $l_2 - norm$.
- `N_jobs` — par défaut : `None`
- `Positive` — par défaut : `False` \Rightarrow Lorsqu'elle est définie sur `True`, elle force les coefficients à être positifs.

MAE	RMSE	Max error	Explained variance
0.9627	1.12	1.79	0.05

TABLE 3.2: Critère d'évaluation pour la régression linéaire multiple

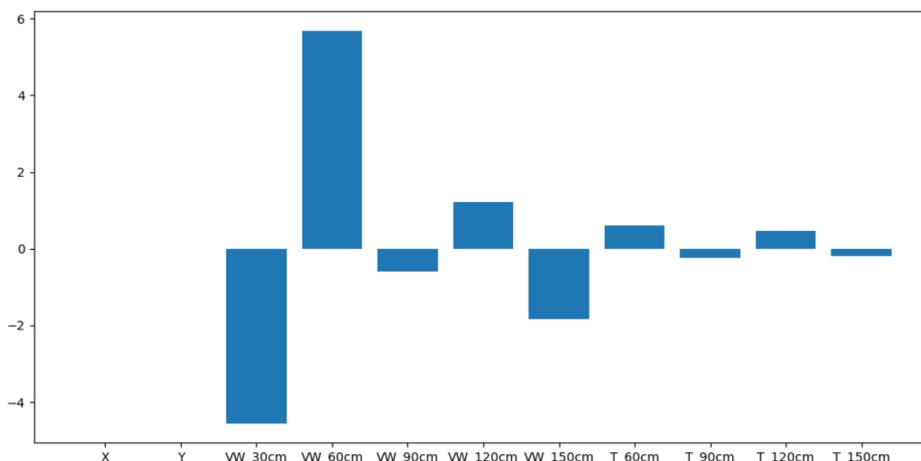


FIGURE 3.4: Histogramme de l'importance des caractéristiques pour la régression linéaire.

D'après la figure 3.4 on remarque que les paramètres ayant le plus d'impact sur la prédiction de la température à une profondeur de 30cm pour la régression linéaire sont l'humidité dans les profondeurs de 30 et 60 cm suivi de celles à 150cm et 120cm, l'influence des températures quant à elle est négligeable alors que celle de la position est nulle.

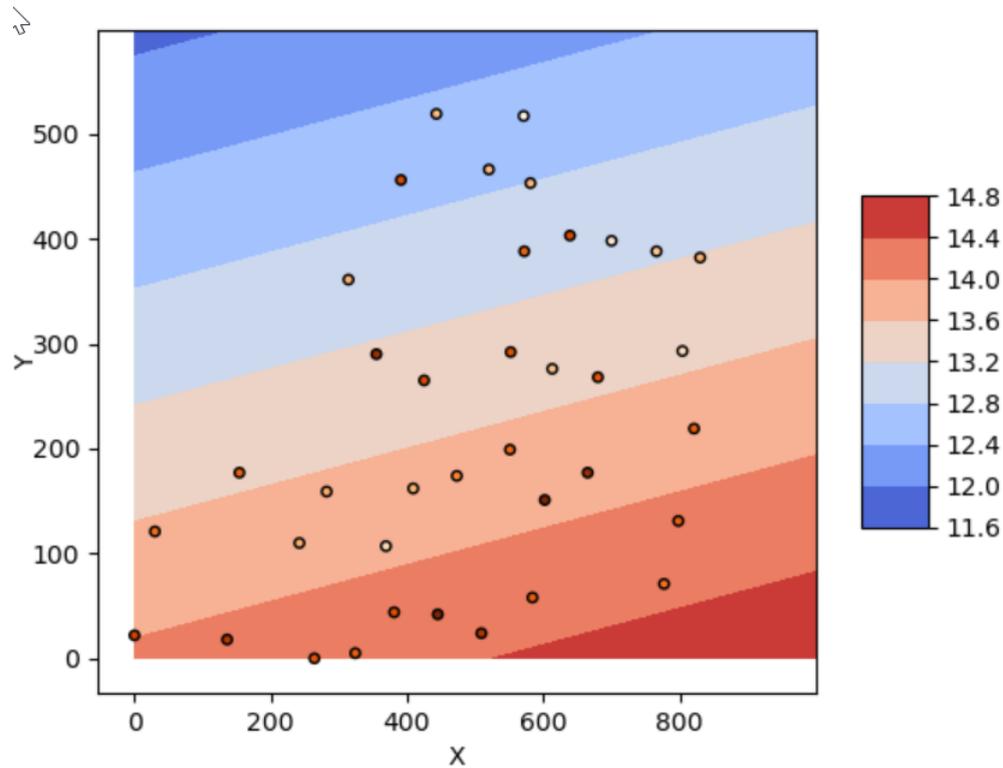


FIGURE 3.5: Distribution de la température sur la surface d'intérêt pour la régression linéaire.

La figure 3.5 représente la distribution de la température à 30 cm sur la surface d'intérêt obtenu à travers l'algorithme de régression linéaire. Elle contient aussi l'emplacement des capteurs avec leurs températures réelles (non prédites). La température estimée à partir de ce modèle est distribuée linéairement, et ceci est dû au fait que notre modèle a établi une relation linéaire entre la variable dépendante (Y : la température à 30cm) et les variables indépendantes (X : Les températures aux autres profondeurs, l'humidité ainsi que la localisation du capteur).

3.5.3 Forêts aléatoires

Hyper paramètres par défaut :

- `n_estimators` — par défaut : 100 \Rightarrow Le nombre d'arbres dans la Forêt.
- `Criterion` — `{'mse', 'mae'}`, par défaut = `'mse'` \Rightarrow fonction permettant de mesurer la qualité d'un fractionnement. Les critères supportés sont `"mse"` pour l'erreur quadratique moyenne, qui est égale à la réduction de la variance comme critère de sélection des caractéristiques, et `"mae"` pour l'erreur absolue moyenne.
- `Max_depth` — par défaut : `None` \Rightarrow La profondeur maximale de l'arbre. Si `None`, alors les noeuds sont développés jusqu'à ce que toutes les feuilles soient pures ou jusqu'à ce que toutes les feuilles contiennent moins de `min_samples_split` échantillons.
- `min_samples_split` — par défaut = 2 \Rightarrow Le nombre minimum d'échantillons requis pour diviser un noeud interne.
- `min_samples_leaf` — par défaut = 1 \Rightarrow Le nombre minimum d'échantillons requis pour être à un noeud feuille. Un point de séparation à n'importe quelle profondeur ne sera considéré que s'il laisse au moins `min_samples_leaf` échantillons d'entraînement dans chacune des branches gauche et droite. Cela peut avoir pour effet de lisser le modèle, en particulier dans la régression.

MAE	RMSE	Max error	Explained variance
0.7667	0.8482	1.332	0.4

TABLE 3.3: Critère d'évaluation pour les Forêts aléatoires

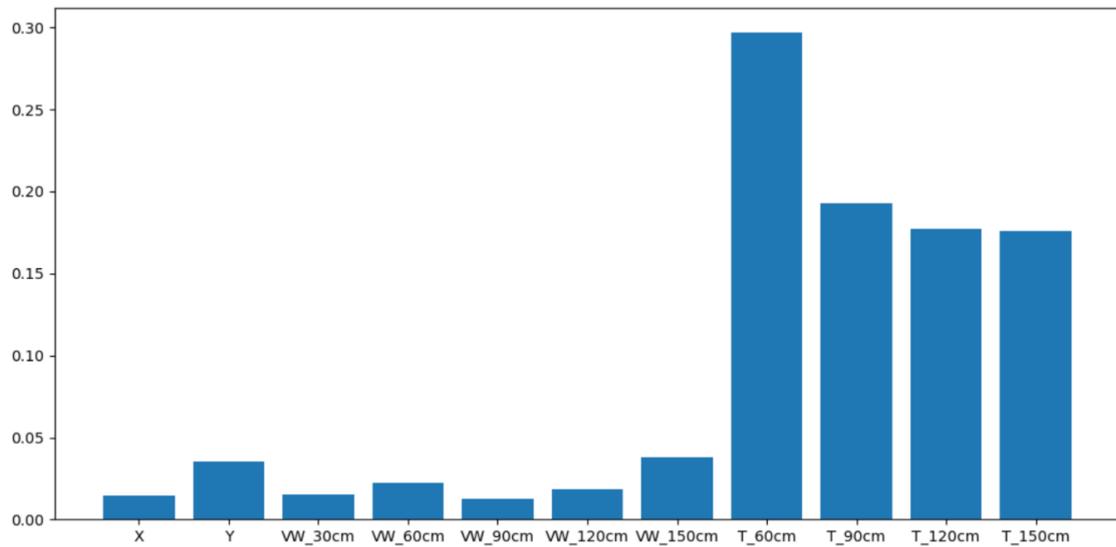


FIGURE 3.6: Histogramme de l'importance des caractéristiques pour les forêts aléatoires.

D'après la figure 3.6 on remarque que les paramètres ayant le plus d'impact sur la prédiction de la température à une profondeur de 30cm pour les forêts aléatoires sont d'abord la température à 60cm suivi des températures aux autres profondeurs, l'influence des autres paramètres quant à elle est minime.

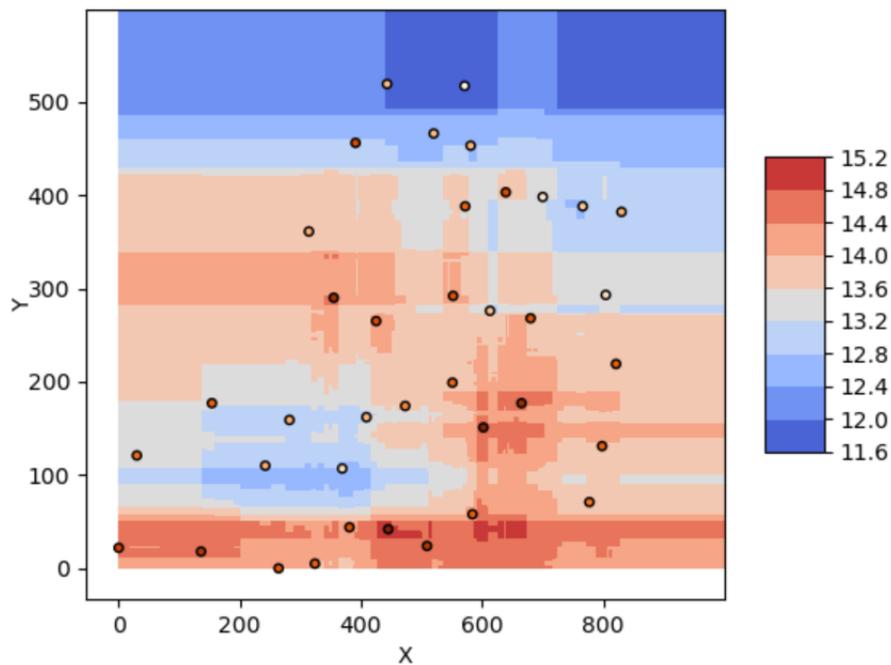


FIGURE 3.7: Distribution de la température sur la surface d'intérêt pour les forêts aléatoires

La figure 3.7 représente la distribution de la température à 30 cm sur la surface d'intérêt obtenu à travers l'algorithme des forêts aléatoires. Elle contient aussi l'emplacement des capteurs avec leurs températures réelles (non prédites). On remarque que ce modèle nous donne une meilleure estimation sur les températures réelles et a une distribution plus homogène (pas de variation brusque de la distribution de la température).

3.5.4 Extreme gradient boosting — XGBoost

Hyper paramètres par défaut :

- Booster — par défaut : gbtree \Rightarrow Le booster à utiliser, peut être gbtree, gblinear ou dart ; gbtree et dart utilisent des modèles basés sur des arbres tandis que gblinear utilise des fonctions linéaires.
- Nthred : par défaut : nombre maximum de threads disponibles.
- Eta — par défaut = 0.3 \Rightarrow Le rétrécissement de la taille du pas est utilisé dans la mise à jour pour éviter l'overfitting. Après chaque étape de boosting, nous pouvons obtenir directement les poids des nouvelles caractéristiques, et eta réduit les poids des caractéristiques pour rendre le processus de boosting plus conservateur. Intervalle : $[0,1]$
- Gamma(Γ) — par défaut = 0 \Rightarrow Réduction minimale de la perte requise pour effectuer une nouvelle partition sur un nœud feuille de l'arbre. Plus gamma est grand, plus l'algorithme sera conservateur.
- max_depth — par défaut = 6 \Rightarrow Profondeur maximale d'un arbre. L'augmentation de cette valeur rendra le modèle plus complexe et plus susceptible d'être sur ajusté.
- Min_child_weight — par défaut = 1 \Rightarrow Somme minimale de poids d'instance (hessian) nécessaire dans un enfant. Si l'étape de partition de l'arbre aboutit à un nœud feuille dont la somme des poids d'instance est inférieure à min_child_weight, le processus de construction renonce à poursuivre le partitionnement. Dans la tâche de régression linéaire, cela correspond simplement au nombre minimum d'instances nécessaires dans chaque nœud. Plus min_child_weight est grand, plus l'algorithme sera conservateur.

MAE	RMSE	Max error	Explained variance
0.7366	0.8956	1.71	0.359

TABLE 3.4: Critère d'évaluation pour Extreme gradient boosting

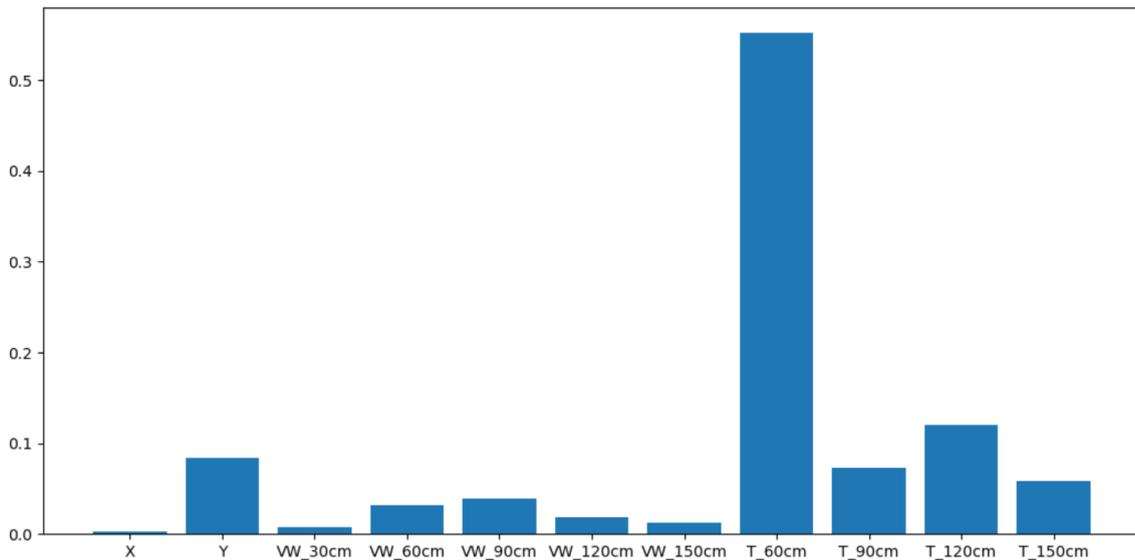


FIGURE 3.8: Histogramme de l'importance des caractéristiques pour XGBoost

D'après la figure 3.8 on remarque que les paramètres ayant le plus d'impact sur la prédiction de la température à une profondeur de 30cm pour XGBoost est principalement la température à 60cm, suivi des températures aux autres profondeurs et la localisation verticale du capteur, l'humidité ainsi que la position horizontale ont une influence très négligeable.

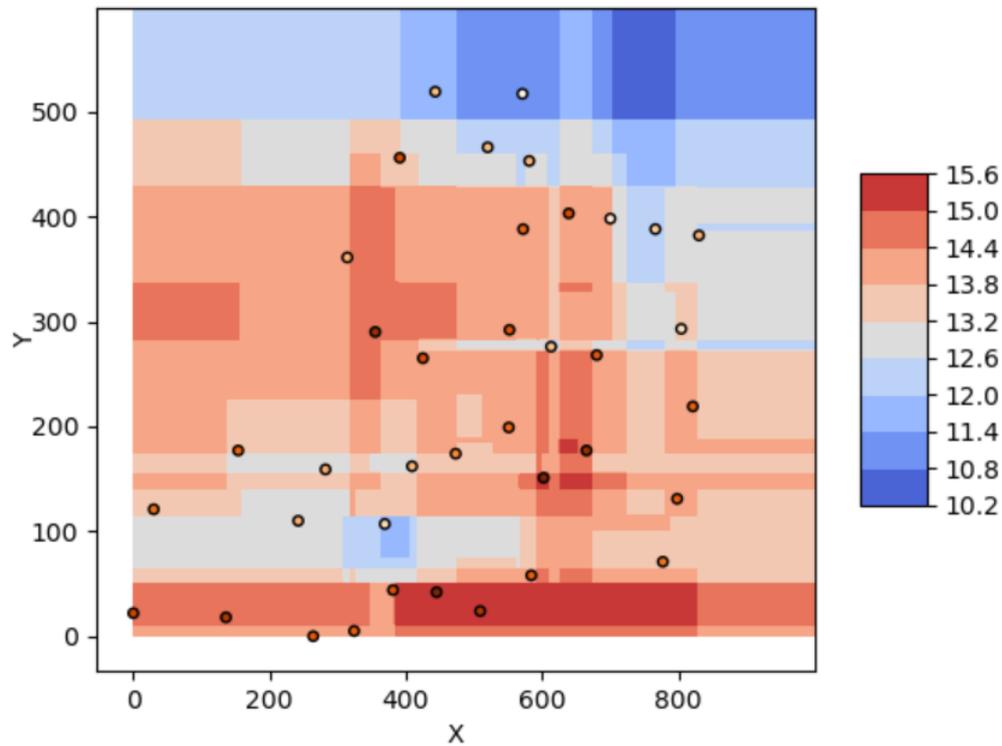


FIGURE 3.9: Distribution de la température sur la surface d'intérêt pour XGBoost

La figure 3.9 représente la distribution de la température à 30 cm sur la surface d'intérêt obtenu à travers l'algorithme XGBoost. Elle contient aussi l'emplacement des capteurs avec leurs températures réelles (non prédites). On remarque que ce modèle nous donne aussi une bonne estimation sur les températures réelles, cependant, la distribution de cette dernière est globalement homogène.

3.5.5 Comparaison entre les trois méthodes

Modèle	Précision
Régression linéaire	92.48%
Forêts aléatoires	94.38%
XGBoost	94.2%

TABLE 3.5: Précision des différents modèles dans la phase de validation

D'après les tableaux 3.4, 3.3 et 3.2 XGBoost donne le meilleur score de la moyenne des erreurs absolues avec un score de 0.7366 suivi de près des forêts aléatoires et enfin

la régression linéaire multiple avec respectivement des scores de 0.7667 et 0.9627. Néanmoins, pour les autres critères, à savoir la *RMSE*, *MaxError* et la variance expliquée, les forêts aléatoires se démarquent des 2 autres modèles et donne un score très satisfaisant. En comparant aussi les précisions des modèles dans le tableau 3.5, bien que les trois modèles donnent une grande précision, les forêts aléatoires restent en première place. Pour l'importance des caractéristiques pour les forêts aléatoires et XGBoost, les températures dans les autres profondeurs sont les plus influents comme envisagé logiquement, par contre pour la régression linéaire, c'est l'humidité du sol qui a le plus d'influence sur le modèle.

3.5.6 Ajustement des hyper paramètres

Les modèles d'ensemble d'arbres à savoir Forêts aléatoires et XGBoost ayant donné des résultats satisfaisants, on veut voir si on peut améliorer davantage leur performance à travers un tuning d'hyper paramètres.

Validation croisée à k-blocs Avant de rentrer dans le vif du sujet, il est nécessaire de définir un processus impératif au tuning des hyper paramètres qui est la validation croisée à *k – blocs* : Supposons posséder un modèle statistique avec un ou plusieurs paramètres inconnus, et un ensemble de données d'apprentissage sur lequel on peut apprendre (ou « entraîner ») le modèle. Le processus d'apprentissage optimise les paramètres du modèle afin que celui-ci corresponde le mieux possible aux données d'apprentissage. Si on prend ensuite un échantillon de validation indépendant, supposément issu de la même population que l'échantillon d'apprentissage, il s'avérera en général que le modèle ne modélise pas aussi bien les données de validation que les données d'apprentissage : on parle de surapprentissage. Néanmoins, un échantillon de validation indépendant n'est pas toujours disponible. De plus, d'un échantillon de validation à un autre, la performance de validation du modèle peut varier. La validation croisée permet de tirer plusieurs ensembles de validation d'une même base de données et ainsi d'obtenir une estimation plus robuste, avec biais et variance, de la performance de validation du modèle. La validation croisée à *k – blocs*, « *k – foldcross – validation* » : on divise l'échantillon original en *k* échantillons (ou « blocs »), puis on sélectionne

un des k échantillons comme ensemble de validation pendant que les $k - 1$ autres échantillons constituent l'ensemble d'apprentissage. Après apprentissage, on peut calculer une performance de validation. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les blocs prédéfinis. À l'issue de la procédure nous obtenons ainsi k scores de performances, un par bloc. La moyenne et l'écart type des k scores de performances peuvent être calculés pour estimer le biais et la variance de la performance de validation.

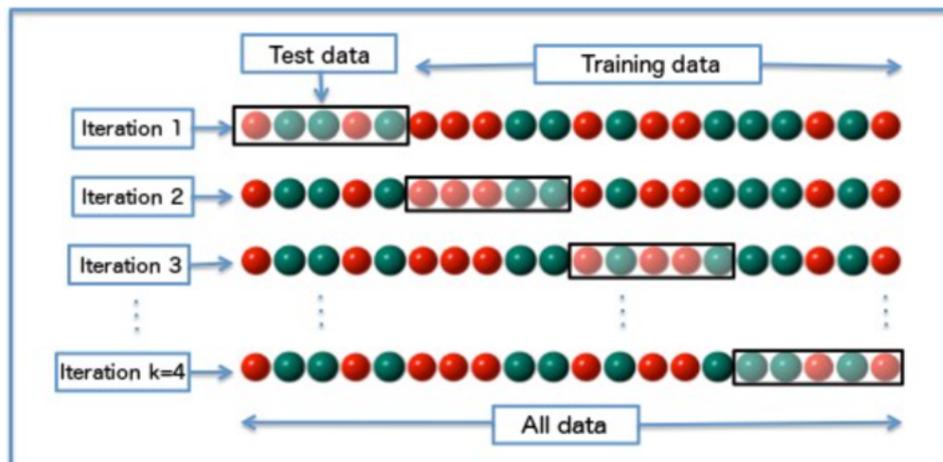


FIGURE 3.10: Diagramme explicatif de la validation croisée à k -blocs

Les algorithmes de tuning les plus simples et répandus sont GridSearch et RandomSearch :

GridSearch GridSearch est un algorithme exhaustif qui englobe toutes les combinaisons, de sorte qu'il peut réellement trouver le meilleur point du domaine. Le grand inconvénient est qu'il est très lent. Vérifier chaque combinaison de l'espace demande beaucoup de temps qui, parfois, n'est pas disponible. Chaque point de la grille nécessite une validation croisée k -fold, ce qui requiert k étapes de formation. Ainsi, le réglage des hyperparamètres d'un modèle de cette manière peut être assez complexe et coûteux. Heureusement, son alternative, Random Search est similaire à Grid Search, mais au lieu d'utiliser tous les points de la grille, elle ne teste qu'un sous-ensemble de ces points choisis au hasard. Plus ce sous-ensemble est petit, plus l'optimisation est rapide mais moins précise. Plus ce sous-ensemble est grand, plus l'optimisation est précise mais plus elle se rapproche d'une recherche sur grille. Nous permettons ainsi

de diminuer considérablement le temps d'exécution tant précieux en agriculture de précision.

Les hyper paramètres optimaux pour les forêts aléatoires sont :

- N_estimators : 50
- Max_depth : 10
- Max_features : 0.45

Ces hyper paramètres ont augmenté la précision du modèle de 1.31% à savoir une nouvelle précision de 95.61%.

Les hyper paramètres optimaux pour XGBoost :

- Eta : 0.75
- Gamma : 0.45
- Max_depth : 6
- Min_child_weight : 1

Ces hyper paramètres ont augmenté la précision du modèle de 0.93% à savoir une nouvelle précision de 95.07%.

3.5.7 Influence du nombre de données

Pour ce faire, on visualisera la courbe d'apprentissage qui détermine les scores de formation et de test validés par recouplement pour différentes tailles d'ensemble de formation. Un générateur de validation croisée divise l'ensemble de données k fois en données de formation et de test. Des sous-ensembles de l'ensemble d'apprentissage de différentes tailles seront utilisés pour entraîner l'estimateur et un score pour chaque taille de sous-ensemble d'apprentissage et l'ensemble de test sera calculé. Ensuite, la moyenne des scores sera calculée sur l'ensemble des k exécutions pour chaque taille de sous-ensemble d'entraînement.

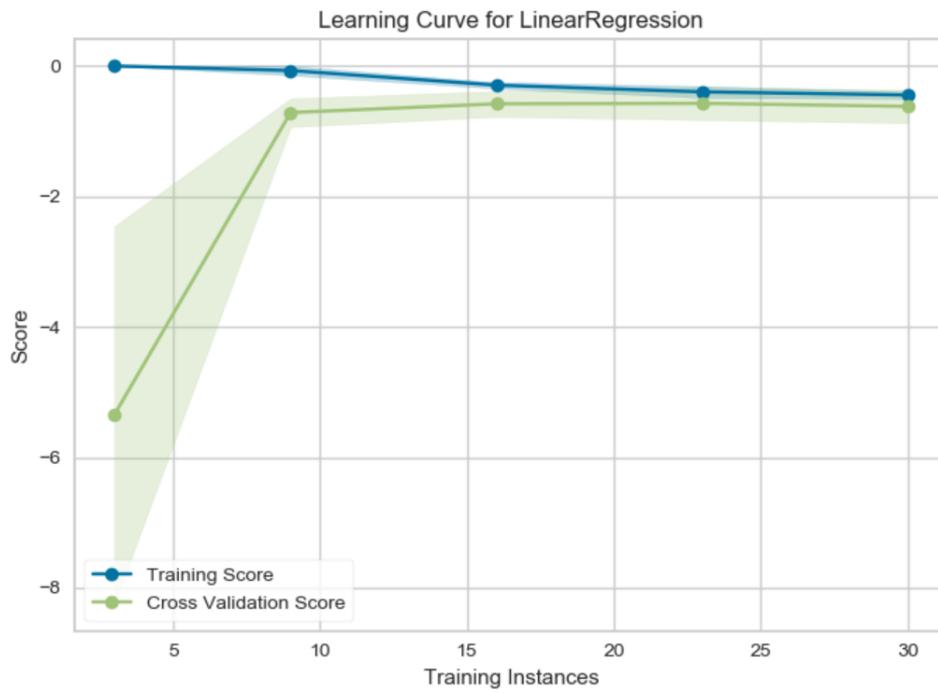


FIGURE 3.11: Courbe d'apprentissage pour la régression linéaire

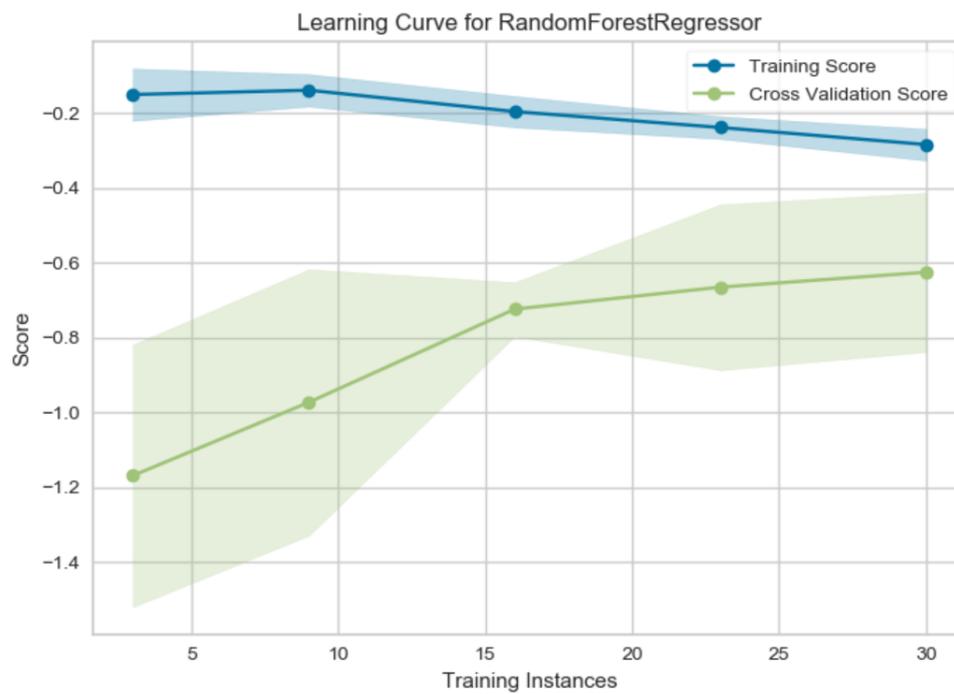


FIGURE 3.12: Courbe d'apprentissage pour les forêts aléatoires

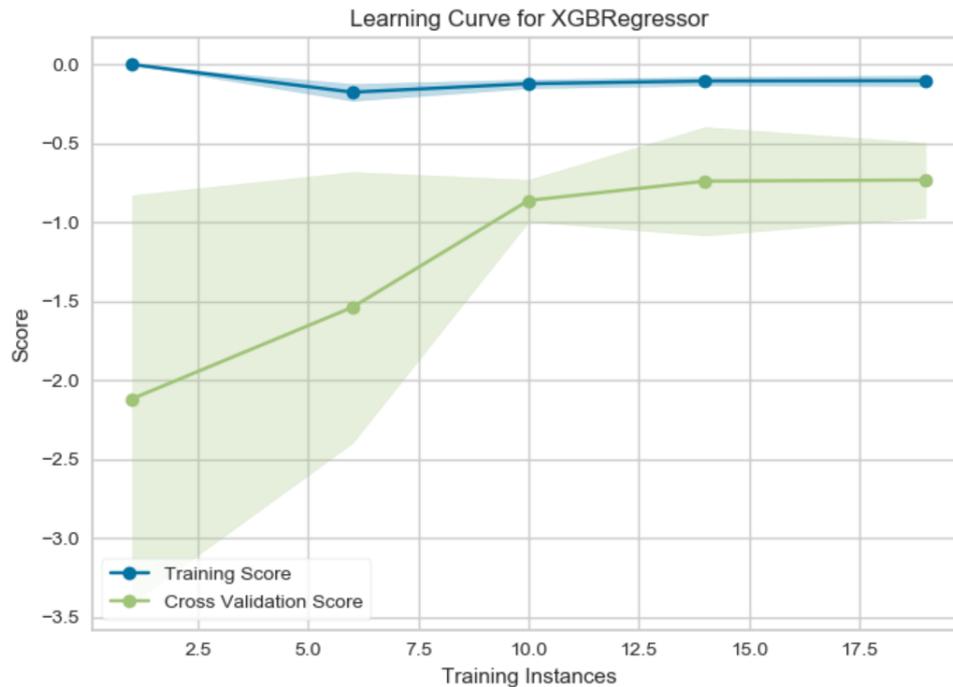


FIGURE 3.13: Courbe d'apprentissage pour XGBoost

On peut remarquer depuis les figures 3.10, 3.11 et 3.12 qui montrent les courbes d'apprentissage de la régression linéaire, forêts aléatoires et XGBoost respectivement que le nombre de données nécessaires pour arriver à la performance du modèle varie d'un modèle à un autre. La régression linéaire ne nécessite que 10 points d'échantillons, suivi de XGBoost avec approximativement 14 points ensuite les forêts aléatoires qui nécessitent 30 points d'échantillons. Ces résultats peuvent changer la donne quant aux choix du modèle, puisque malgré le fait que les forêts aléatoires donnent les meilleurs résultats, XGBoost reste néanmoins aussi performant mais en plus de cela nécessite moins de points d'échantillons, c'est-à-dire moins de capteurs, et davantage d'économie, encore plus si la parcelle ciblée est d'une plus grande superficie.

3.6 Conclusion

D'après les résultats obtenus précédemment par les trois méthodes on constate que les forêts aléatoires donnent les meilleurs résultats bien que XGBoost nécessite moins de capteurs pour un résultat presque similaire à celui des forêts aléatoires qui eux nécessitent presque le double des capteurs pour un gain minime, de plus

bien que la régression linéaire donne un résultat relativement plus faible, elle reste néanmoins robuste et le choix final du modèle reste ouvert aux problèmes potentiels d'implémentation sur carte pour un traitement en temps réel qui nécessiterait un compromis entre simplicité du modèle et performance.

Conclusion générale et perspectives

Les récents développements technologiques notamment en électronique et en informatique ont permis aux industries d'arborer de nouvelles stratégies en accord avec ces avancées.

L'agriculture étant un domaine de fort intérêt puisqu'elle apporte nourriture aux populations, il est tout à fait rationnel qu'elle soit de plus en plus touchée par ces innovations technologiques.

L'agriculture de précision, qui permet de mieux comprendre et analyser les besoins physiologiques des cultures et de développer des outils d'aide à la décision pour l'utilisateur, est l'un des domaines les plus en expansion, et de par la disponibilité de capteurs performants à prix réduit, les réseaux de capteurs sans fils sont de plus en plus utilisés pour la collecte de données et l'actionnement des différents actionneurs. De ce fait nous proposons un système à base de réseau de capteurs et actionneurs sans fil dédié à l'agriculture de précision.

Dans le premier chapitre, nous étudions l'état de l'art de l'agriculture de précision, tout en énonçant les différentes technologies utilisées dans ce domaine, avec différents exemples d'applications.

Dans le deuxième chapitre, on traite l'intelligence artificielle, en se focalisant sur l'aspect régression, tout en citant les différents modèles algorithmiques de régression et en s'approfondissant sur les modèles choisis pour l'implémentation qui sont : la régression linéaire, les forêts aléatoires et XGBoost.

Il s'ensuit au troisième chapitre l'implémentation des différents algorithmes et la comparaison entre eux avec pour objectif l'interpolation des données sur les parties

non-échantillonnées, permettant ainsi un meilleur usage des capteurs disponibles qui comme dit précédemment sont souvent limités en nombre.

La comparaison entre les différents algorithmes démontre que les forêts aléatoires sont les plus performants, suivi de près du XGBoost puis la régression linéaire. XGBoost reste toutefois très souple puisqu'il a un comportement assez bon en cas de faibles quantités de données.

Les résultats et commentaires de cette étude sont basés sur seulement un jeu de données issu d'une région lointaine, il serait donc intéressant de tester ces approches sur d'autres données notamment d'autres régions d'Algérie.

Néanmoins, ce travail peut être amélioré en répondant aux perspectives suivantes :

- Utiliser les résultats de ce mémoire et l'implémenter sur des cartes pour l'automatisation des tâches telle que l'irrigation.
- Etudier l'influence de l'emplacement des capteurs sur l'efficacité des modèles proposés dans ce mémoire.
- Utiliser les paramètres extrapolés comme nouvelle entrée pour prédire les autres paramètres.

Bibliographie

- [1] D. Jared, “Farmers and their languages : The first expansions,” vol. 300, pp 597-603.
- [2] “Concise report on the world population situation in 2014,” 2014. [Online]. Available : <https://www.un.org/en/development/desa/population/publications/pdf/trends/Concise%20Report%20on%20the%20World%20Population%20Situation%202014/en.pdf>
- [3] M. José, García-Ruiz, J. Ignacio, López-Moreno, S. M. Vicente-Serrano, T. Lasanta, and S. Beguería, “Mediterranean water resources in a global change scenario,” 2009. [Online]. Available : <https://core.ac.uk/download/pdf/36045074.pdf>
- [4] P. Thenkabail, M. Hanjra, V. Dheeravath, and M. Gumma, “Global croplands and their water use from remote sensing and nonremote sensing perspectives,” *Advances in environmental remote sensing : sensors, algorithms and applications*, pp. 383–420, 01 2010.
- [5] N.Wang and Z.Li, *Robotics and Automation in the Food Industry, Chapter 8 : Wireless sensor networks (WSNs) in the agricultural and food industries*. Elsevier, pp. 171-199, 2013.
- [6] Groupe-MadeInFutura, “futura-sciences, developpement-durable-agriculture,” 2021. [Online]. Available : <https://www.futura-sciences.com/planete/definitions/developpement-durable-agriculture-11492/>
- [7] M. N. BOUHDJAR, “Etude et implémentation d’un système dédié à l’agriculture de précision,” 2018.

- [8] Parlement-Européen, “L’agriculture de précision et l’avenir de l’agriculture en europe,” 2016. [Online]. Available : [https://www.europarl.europa.eu/RegData/etudes/STUD/2016/581892/EPRS_STU\(2016\)581892_FR.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2016/581892/EPRS_STU(2016)581892_FR.pdf)
- [9] OZ, “L’assistant agricole pour les tâches chronophages et pénibles,” 2013. [Online]. Available : <https://www.naio-technologies.com/oz/>
- [10] R. Lal, “The plow and agricultural sustainability,” *Journal of Sustainable Agriculture - J SUSTAINABLE AGR*, vol. 33, pp. 66–84, 01 2009.
- [11] S. A. Amrita, E. Abirami, A. Ankita, R. Praveena, and R. Srimeena, “Agricultural robot for automatic ploughing and seeding,” in *2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR)*, 2015, pp. 17–23.
- [12] V. Kumar, “agriculture soil analysis for suitable crop prediction,” 2012. [Online]. Available : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3852055
- [13] Purdue.Edu, “Smartcore autonomous robot,” 2018. [Online]. Available : <https://www.purdue.edu/newsroom/releases/2019/Q2/autonomous-robots-in-agricultural-fields-advance-through-200,000-investment-to-purdue-startup.html5>
- [14] K. Sampooram, T. Dinesh, and P. Jegannathan, “Agriculture robot (agribot) for harvesting underground plants (rhizomes),” *Agricultural Engineering International : CIGR Journal*, vol. 19, pp. 62–67, 01 2017.
- [15] J. K. Yongtae Do, “Infrared range sensor array for 3d sensing in robotic applications,” vol. 10, 2013. [Online]. Available : <https://journals.sagepub.com/doi/10.5772/55896?icid=int.sj-full-text.similar-articles.2&>
- [16] S. Water, “Soil monitoring,” 2021. [Online]. Available : <https://stevenswater.com/applications/soil-monitoring/>
- [17] W. Holland, “Terrasentia : the automated crop monitoring robot,” 2018. [Online]. Available : <https://blog.plantwise.org/2018/07/17/terrasentia-the-automated-crop-monitoring-robot/>

- [18] Y. Xiong, Y. Ge, L. Grimstad, and P. From, “An autonomous strawberry-harvesting robot : Design, development, integration, and field evaluation,” *Journal of Field Robotics*, vol. 37, 08 2019.
- [19] G. Kootstra, X. Wang, P. M. Blok, J. Hemming, and E. van Henten, “Selective harvesting robotics : Current research, trends, and future directions,” 2021. [Online]. Available : <https://link.springer.com/article/10.1007/s43154-020-00034-1>
- [20] X. You, J. Meng, M. Zhang, and T. Dong, “Remote sensing based detection of crop phenology for agricultural zones in china using a new threshold method,” vol. 5, no. 7, pp. 3190–3211. [Online]. Available : <https://www.mdpi.com/2072-4292/5/7/3190>
- [21] Y. S. J.-s. C. T. H. H.-z. P. Chang-LIU, Zhong-xin CHEN, “Research advances of sar remote sensing for agriculture applications : A review,” vol. 18, no. 3, pp. 506–525, 2019.
- [22] “Landsat 1, usgs.gov,” 2018. [Online]. Available : https://www.usgs.gov/core-science-systems/nli/landsat/landsat-1?qt-science_support_page_related_con=0#qt-science_support_page_related_con
- [23] J. Xue and B. Su, “Significant remote sensing vegetation indices : A review of developments and applications,” vol. 2017. [Online]. Available : <https://www.hindawi.com/journals/js/2017/1353691/>
- [24] D. Patón, “Normalized difference vegetation index determination in urban areas by full-spectrum photography,” no. 1, pp. 22–35, 2020.
- [25] L. Fan, Y. Gao, H. Brueck, and C. Bernhofer, “Investigating the relationship between NDVI and LAI in semi-arid grassland in inner mongolia using in - situ measurements,” vol. 95, pp. 151–156.
- [26] R. P. Sishodia, R. L. Ray, and S. K. Singh, “Applications of remote sensing in precision agriculture : A review,” vol. 12, no. 19, p. 3136. [Online]. Available : <https://www.mdpi.com/2072-4292/12/19/3136>

- [27] H. Navarro-Hellín, J. Martínez-del Rincon, R. Domingo-Miguel, F. Soto-Valles, and R. Torres-Sánchez, “A decision support system for managing irrigation in agriculture,” vol. 124, pp. 121–131. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S016816991630117X>
- [28] F. Li, D. Li, S. Elsayed, Y. Hu, and U. Schmidhalter, “Using optimized three-band spectral indices to assess canopy n uptake in corn and wheat,” vol. 127, p. 126286. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S1161030121000587>
- [29] L. Zhang, W. R. Dawes, and G. R. Walker, “Response of mean annual evapotranspiration to vegetation changes at catchment scale,” vol. 37, no. 3, pp. 701–708. [Online]. Available : <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000WR900325>
- [30] A.-K. Mahlein, T. Rumpf, P. Welke, H.-W. Dehne, L. Plümer, U. Steiner, and E.-C. Oerke, “Development of spectral indices for detecting and identifying plant diseases,” vol. 128, pp. 21–30. [Online]. Available : <https://linkinghub.elsevier.com/retrieve/pii/S0034425712003793>
- [31] S. F. D. GENNARO, S. D. M. Enrico BATTISTON, O. FACINI, A. MATESE, M. NOCENTINI, A. PALLIOTTI, and L. MUGNAI, “Unmanned aerial vehicle (uav)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex,” vol. 55, no. 2, pp. 262–275, 2016. [Online]. Available : <https://www.jstor.org/stable/44809332>
- [32] D. Moshou, C. Bravo, R. Oberti, J. West, L. Bodria, A. McCartney, and H. Ramon, “Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using kohonen maps,” vol. 11, pp. 75–83, 2005. [Online]. Available : <https://linkinghub.elsevier.com/retrieve/pii/S1077201405000136>
- [33] A.-K. Mahlein, T. Rumpf, P. Welke, H.-W. Dehne, L. Plümer, U. Steiner, and E.-C. Oerke, “Development of spectral indices for detecting and

- identifying plant diseases,” vol. 128, pp. 21–30, 2013. [Online]. Available : <https://linkinghub.elsevier.com/retrieve/pii/S0034425712003793>
- [34] M. Weiss, F. Baret, G. Smith, I. Jonckheere, and P. Coppin, “Review of methods for in situ leaf area index (LAI) determination,” vol. 121, pp. 37–53, 2004. [Online]. Available : <https://linkinghub.elsevier.com/retrieve/pii/S0168192303001631>
- [35] R. P. Sishodia, R. L. Ray, and S. K. Singh, “Applications of remote sensing in precision agriculture : A review,” vol. 12, no. 19, p. 3136. [Online]. Available : <https://www.mdpi.com/2072-4292/12/19/3136>
- [36] A. G. Schut, P. C. S. Traore, X. Blaes, and R. A. de By, “Assessing yield and fertilizer response in heterogeneous smallholder fields with UAVs and satellites,” vol. 221, pp. 98–107. [Online]. Available : <https://linkinghub.elsevier.com/retrieve/pii/S037842901731016X>
- [37] M. Arias, M. Campo-Bescós, and J. Álvarez Mozos, “Crop classification based on temporal signatures of sentinel-1 observations over navarre province, spain,” vol. 12, no. 2, p. 278. [Online]. Available : <https://www.mdpi.com/2072-4292/12/2/278>
- [38] W. Qian, Y. Huang, Q. Liu, W. Fan, Z. Sun, H. Dong, F. Wan, and X. Qiao, “UAV and a deep convolutional neural network for monitoring invasive alien plants in the wild,” vol. 174, 2020. [Online]. Available : <https://linkinghub.elsevier.com/retrieve/pii/S0168169920302921>
- [39] V. Partel, J. Kim, L. Costa, P. Pardalos, and Y. Ampatzidis, “Smart sprayer for precision weed control using artificial intelligence : Comparison of deep learning frameworks,” in *ISAIM*.
- [40] I. Akyildiz, S. WY, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks : A survey,” vol. 38, pp. 393–422.
- [41] O. Mendoza-Cano, R. Aquino-Santos, J. López-de la Cruz, R. M. Edwards, A. Khouakhi, I. Pattison, V. Rangel-Licea, E. Castellanos-Berjan, M. A.

- Martinez-Preciado, P. Rincón-Avalos, P. Lepper, A. Gutiérrez-Gómez, J. M. Uribe-Ramos, J. Ibarreche, and I. Perez, “Experiments of an IoT-based wireless sensor network for flood monitoring in colima, mexico,” vol. 23, no. 3, pp. 385–401. [Online]. Available : <https://doi.org/10.2166/hydro.2021.126>
- [42] X. Yu, P. Wu, W. Han, and Z. Zhang, “Overview of wireless underground sensor networks for agriculture,” vol. 11, no. 17, pp. 3942–3948. [Online]. Available : <https://www.ajol.info/index.php/ajb/article/view/101035>
- [43] —, “A survey on wireless sensor network infrastructure for agriculture,” vol. 35, no. 1, pp. 59–64. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0920548912000608>
- [44] S. Adamala, N. S. Raghuwanshi, and A. Mishra, “Development of surface irrigation systems design and evaluation software (SIDES),” vol. 100, pp. 100–109. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0168169913002743>
- [45] S. von Westarp, S. Chieng, and H. Schreier, “A comparison between low-cost drip irrigation, conventional drip irrigation, and hand watering in nepal,” vol. 64, no. 2, pp. 143–160. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S0378377403002063>
- [46] J. Gutiérrez, J. F. Villa-Medina, A. Nieto-Garibay, and M. Porta-Gándara, “Automated irrigation system using a wireless sensor network and GPRS module,” vol. 63, no. 1, pp. 166–176.
- [47] L. B. L. Gonçalves, F. G. Costa, L. A. Neves, J. Ueyama, G. F. D. Zafalon, C. Montez, and A. S. R. Pinto, “Influence of mobility models in precision spray aided by wireless sensor networks,” vol. 574, p. 012153. [Online]. Available : <https://doi.org/10.1088/1742-6596/574/1/012153>
- [48] A. K. Tripathy, J. Adinarayana, D. Sudharsan, S. N. Merchant, U. B. Desai, K. Vijayalakshmi, D. Raji Reddy, G. Sreenivas, S. Ninomiya, M. Hirafuji,

- T. Kiura, and K. Tanaka, "Data mining and wireless sensor network for agriculture pest/disease predictions," in *2011 World Congress on Information and Communication Technologies*, pp. 1229–1234.
- [49] United-Nations, "Concise report on the world population situation in 2014," 2014. [Online]. Available : <https://www.un.org/en/development/desa/population/publications/pdf/trends/Concise%20Report%20on%20the%20World%20Population%20Situation%202014/en.pdf>
- [50] J. N. Kok, "Artificial intelligence : Definition, trends, techniques and cases," p. 5.
- [51] S. Das, I. Ghosh, G. Banerjee, and U. Sarkar, "Artificial intelligence in agriculture : A literature survey."
- [52] A. Savla, N. Israni, P. Dhawan, A. Mandholia, H. Bhadada, and S. Bhardwaj, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1–7.
- [53] M. A. I.-L. S. T. O. Aulia Rizkiana, Andri Prima Nugroho, "Crop growth prediction model at vegetative phase to support the precision agriculture application in plant factory," 2019. [Online]. Available : <https://aip.scitation.org/doi/pdf/10.1063/1.5141717>
- [54] N. Jamal and M. Rind, "Ridge regression : A tool to forecast wheat area and production," *Pakistan Journal of Statistics and Operation Research ; Vol 3. No. 2, July 2007*, vol. 3, 07 2007.
- [55] T. B. H.-G. L. Alexander Erler, Daniel Riebe and R. Gebbers, "Soil nutrient detection for precision agriculture using handheld laser-induced breakdown spectroscopy (libs) and multivariate regression methods (pls-r, lasso and gpr)," vol. 418, no. 2, 2020. [Online]. Available : <https://www.mdpi.com/1424-8220/20/2/418>
- [56] R. T.Waheed, S.O.Prasher, and E.Paulet, "Measuring performance in precision agriculture : Cart—a decision tree approach," pp. 173–185, 2006. [Online].

- Available : <https://www.sciencedirect.com/science/article/abs/pii/S0378377405003860>
- [57] D. S. . G. I.-B. Yvette Everingham, Justin Sexton, “Accurate prediction of sugarcane yield using a random forest algorithm,” 2016. [Online]. Available : <https://link.springer.com/article/10.1007/s13593-016-0364-z>
- [58] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, “Crop yield prediction using machine learning algorithms,” in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, 2019, pp. 125–130.
- [59] V. Sirisha and G. Sahitya, “Smart irrigation system for the reinforcement of precision agriculture using prediction algorithm : Svr based smart irrigation,” in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1059–1066.
- [60] H. A.M.Mouazena, B.Kuanga J.De Baerdemaeker, “Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy,” vol. 158, pp. 23–31, 2010. [Online]. Available : <https://www.sciencedirect.com/science/article/abs/pii/S0016706110000753>
- [61] L. P. Suresh, S. S. Dash, and B. K. Panigrahi, “Artificial intelligence and evolutionary algorithms in engineering systems,” 2015. [Online]. Available : <https://link.springer.com/book/10.1007/978-81-322-2126-5>
- [62]
- [63] K. Wessels, R. D. Fries, J. Dempewolf, L. Anderson, A. Hansend, S. Powell, and E. Moran, “Mapping regional land cover with modis data for biological conservation : Examples from the greater yellowstone ecosystem, usa and para´ state, brazil,” pp. 67–83, 2004. [Online]. Available : http://www2.geog.ucl.ac.uk/~mdisney/teaching/msc.cons/papers/wessels_et_al.pdf
- [64] “C4.5 : Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993.” [Online]. Available : <https://link.springer.com/article/10.1007/BF00993309>

- [65] J. Quinlan, "Induction of decision trees," 1986. [Online]. Available : <https://link.springer.com/content/pdf/10.1007/BF00116251.pdf>
- [66] F. J. O. R. Breiman, L. and Stone, "Classification and regression trees," 1984. [Online]. Available : <https://link.springer.com/content/pdf/10.1007/BF00116251.pdf>
- [67] J. Mingers, "An empirical comparison of pruning methods for decision tree induction," *Machine Learning*, vol. 4, pp. 227–243, 01 1989.
- [68] J.-O. J.-S. V.F.Rodriguez-Galiano, B.Ghimire, "An assessment of the effectiveness of a random forest classifier for land-cover classification," vol. 67, pp. 93–104, 2012. [Online]. Available : <https://www.sciencedirect.com/science/article/abs/pii/S0924271611001304>
- [69] L. Breiman, "Random forests," pp. 5–32, 2001. [Online]. Available : <https://link.springer.com/article/10.1023/A:1010933404324>
- [70] J. Peters, B. De Baets, N. Verhoest, R. Samson, S. Degroeve, P. De Becker, and W. Huybrechts, "Random forests as a tool for ecohydrological distribution modelling," *Ecological Modelling*, vol. 207, pp. 304–318, 10 2007.
- [71] J. R. Pall OskarGislason, Jon Atli Benediktsson, "Random forests for land cover classification," vol. 27, pp. 294–300, 2006. [Online]. Available : <https://www.sciencedirect.com/science/article/abs/pii/S0167865505002242>
- [72] I. B. Mustapha and F. Saeed, "Bioactive molecule prediction using extreme gradient boosting," vol. 21, 2016. [Online]. Available : <https://pubmed.ncbi.nlm.nih.gov/27483216/>
- [73] Y. L. Y. Xia, C. Liu and N. Liu, "A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring," 2017.
- [74] T. Chen and C. Guestrin, "Xgboost : A scalable tree boosting system," vol. 78, pp. 225–241, 2016. [Online]. Available : <https://www.sciencedirect.com/science/article/abs/pii/S0957417417301008>

-
- [75] “<https://data.nal.usda.gov/dataset/data-field-scale-sensor-network-data-set-monitoring-and-modeling-spatial-and-temporal-variation-soil-moisture-dryland-agricultural-field>.”