



HAL
open science

Analyse intelligente des images pour la surveillance dans une agriculture de précision

Solemane Coulibaly

► **To cite this version:**

Solemane Coulibaly. Analyse intelligente des images pour la surveillance dans une agriculture de précision. Autre. Institut National Polytechnique de Toulouse - INPT; Université des Sciences Techniques et Technologiques de Bamako (Mali), 2021. Français. NNT : 2021INPT0099 . tel-04186483

HAL Id: tel-04186483

<https://theses.hal.science/tel-04186483>

Submitted on 23 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Informatique et Automatique

Présentée et soutenue par :

M. SOLEMANE COULIBALY

le lundi 6 décembre 2021

Titre :

Analyse intelligente des images pour la surveillance dans une agriculture
de précision

Ecole doctorale :

Aéronautique-Astronautique (AA)

Unité de recherche :

Laboratoire Génie de Production de l'ENIT (E.N.I.T-L.G.P.)

Directeurs de Thèse :

M. BERNARD KAMU-FOGUEM

M. DANTOUMA KAMISSOKO

Rapporteurs :

M. DENIS POMORSKI, UNIVERSITE LILLE 1

M. ENGELBERT MEPHU NGUIFO, UNIVERSITE CLERMONT AUVERGNE

Membres du jury :

MME CARMEN GERVET, UNIVERSITE DE MONTPELLIER, Présidente

M. ANNIG LACAYRELLE, UNIVERSITE DE PAU ET DES PAYS DE L'ADOUR, Membre

M. BERNARD KAMU-FOGUEM, ECOLE NATIONALE D'INGENIEUR DE TARBES, Membre

M. DANTOUMA KAMISSOKO, UNIVERSITE DE SEGOU, Membre

M. DAOUA TRAORE, UNIVERSITE DE SEGOU, Membre

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National Polytechnique de Toulouse

Cotutelle internationale : Institut polytechnique rural de formation et de recherche
appliquée de Katibougou

Présentée et soutenue par **Solemane COULIBALY**

Le 6 décembre 2021

Analyse intelligente des images pour la surveillance dans une agriculture de précision

Ecole doctorale : **AA Aéronautique, Astronautique** Spécialité : **Informatique**

Unité de recherche :

LGP - Laboratoire Génie de Production

Thèse dirigée par

Bernard KAMSU FOGUEM et Dantouma KAMISSOKO

Jury

M. Denis POMORSKI, Professeur de l'Université de Lille, Rapporteur

M. Engelbert MEPHU NGUIFO, Professeur de l'Université Clermont Auvergne (UCA),

Rapporteur

Mme Annig LACAYRELLE, Maître de conférences, Université de Pau et des Pays de
l'Adour, Examineur

M. Daouda TRAORE, Maître de conférences, Université de Ségou, Examineur

Mme Carmen GERVET, Professeur Université de Montpellier, Examinatrice

M. Bernard KAMSU FOGUEM, Professeur de l'ENIT, Directeur de thèse

M. Dantouma KAMISSOKO, Professeur Université de Ségou, Directeur de thèse

Résumé

Les avancées technologiques de la vision par ordinateur et l'utilisation des systèmes d'intelligence artificielle (comme l'apprentissage automatique ou profond) ont eu un fort impact dans l'agriculture en la faisant passer à une nouvelle ère digitalisée. Il s'agit d'une agriculture numérique ou de précision dans laquelle on assiste à une forte utilisation des données, de l'informatique intelligente, des drones, et des capteurs pour produire davantage. A cet effet, cette thèse propose de nouvelles réponses aux enjeux de la surveillance des cultures menacées par des attaques fongiques. Ce qui nous a amené à donner de nouvelles contributions visant à relever les défis d'une agriculture de précision.

Partant de là, nous avons effectué une étude bibliométrique sur les apports de l'apprentissage profond en agriculture. Nous avons discuté les contributions des chercheurs principalement impliqués, en vue d'apporter de nouvelles réponses aux défis de l'agriculture de précision. Cette étude a été clôturée par des recommandations essentielles dans la réalisation d'un système agricole intelligent. Il s'agit de (i) considérer la perception des acteurs humains du système; (ii) exiger la prise en compte des tests statistiques et des validations croisées des données d'entraînement lors de la comparaison des performances de plusieurs classificateurs; (iii) analyser la performance d'un classificateur sur les données d'entraînement en faible quantité.

Dans un premier temps, pour consolider cette étude, nous avons étudié la classification de la maladie mildiou au niveau de la culture du mil, une des cultures vivrières des régions d'Afrique et d'Asie. Dans ce travail, un accent a été mis sur la faible quantité de données d'entraînements supervisées, nécessaires pour former de tels classificateurs.

Dans un second temps, nous avons proposé une nouvelle tâche de classification des réseaux de neurones convolutifs en augmentant les espaces de caractéristiques des données d'entraînement. Cette approche se base sur les principes de l'apprentissage multitâches dans l'élaboration d'un modèle de classification multi-labels avec la comparaison de plusieurs classificateurs.

Durant des années, les recherches en apprentissage automatique étaient beaucoup plus concentrées sur la performance des modèles de prédiction plutôt que sur leur compréhension, leur interprétation et leur pouvoir de décision. Si nous comprenions ce que le modèle a appris, il est possible de garantir la qualité des résultats obtenus. Dans un troisième temps, nous avons observé les propriétés de visualisation d'un modèle profond afin d'obtenir des résultats significatifs, explicables par un utilisateur quelconque. Nous avons qualitativement analysé des cartes de visualisation des méthodes d'explicabilité lors de la localisation des insectes ravageurs des cultures. Au-delà d'une prédiction à base du calcul de probabilité, nous avons guidé le processus de la localisation d'insectes en employant la théorie de l'information mutuelle.

L'apprentissage profond requiert assez souvent une grande quantité de données et une puissance de calcul pour entraîner un réseau de neurones profonds. Les résultats obtenus par nos travaux ont prouvé que, l'apprentissage par transfert, l'augmentation des données et le partage des tâches constituent des moyens efficaces pour améliorer la capacité des réseaux de neurones

profonds.

Notons que les systèmes ont pour vocation d'interagir avec des utilisateurs humains. Ils doivent donc être en mesure d'expliquer, de justifier leur comportement et les décisions qu'ils prennent afin que ces utilisateurs puissent comprendre les actions réalisées. Les contributions de cette thèse ont dévoilé que, l'exploration des méthodes de l'explicabilité est un moyen pertinent et utile pour le déploiement des outils d'intelligence artificielle au service de l'agriculture de précision.

Mots clés : agriculture de précision, vision par ordinateur, apprentissage supervisé, réseau de neurones, apprentissage profond, apprentissage par transfert, explicabilité des réseaux de neurones, apprentissage multitâches, détection des maladies, détection des insectes ravageurs

Abstract

Technological advances in computer vision and the use of artificial intelligence systems (such as automatic or deep learning) have had a strong impact on this field, bringing it into a new digitalized era. This is digital or precision agriculture in which there is a strong use of data, intelligent computing, drones, and sensors to produce even more. In this thesis, we propose new answers to the challenges of monitoring crops threatened by fungal attacks. This has led us to make new contributions to the challenges of precision agriculture. Based on this, we conducted a bibliometric study on the contributions of deep learning in agriculture. We discussed the contributions of the researchers mainly involved, in order to provide new answers to the challenges of precision agriculture. This study was closed with essential recommendations in the realization of an intelligent agricultural system. These are (i) to consider the perception of the human actors of the system; (ii) to require the consideration of statistical tests and cross-validations of the training data when comparing the performances of several classifiers; (iii) to analyze the performance of a classifier on the low quantity training data.

First, to consolidate this study, we studied the classification of late blight disease in millet, one of the food crops in African and Asian regions. In this work, emphasis was placed on the small amount of supervised training data needed for such classifiers.

In a second step, we improved the classification tasks of convolutional neural networks by increasing the feature spaces of the training data. This approach is based on the principles of multi-task learning in the development of multi-label classification model the comparison of several classifiers.

For years, research in artificial intelligence was much more focused on the performance of learning models rather than on their understanding, interpretation, and decision power. If we understood what the model learned, we could guarantee the quality of the results obtained. In a third step, we observed the visualization properties of a deep model in order to obtain meaningful results that can be explained by any user. We have qualitatively analyzed the explanatory capacity of deep neural networks when locating crop pests. In addition to a prediction based on a probability value, we have guided the process of insect localization by using the mutual information theory.

Deep learning requires a large amount of data to build a deep neural network and computational power to train it. The results obtained from our work have proven that transfer learning, data augmentation, and task sharing are efficient ways to improve the capacity of deep neural networks.

Note that learning systems based on artificial intelligence are intended to interact with human users. They must therefore be able to explain and justify their behavior and the decisions they make so that these users can understand the actions they take. The contributions of this thesis have revealed that the exploration of the explanation methods is a useful and beneficial tool for farmers. They aim to shed light on the process of scaling up the concerns and challenges of the

new form of agriculture that is precision agriculture.

Keywords : precision agriculture, computer vision, supervised learning, neural network, deep learning, transfer learning, explainable Artificial Intelligence, multi-task learning, disease detection, insect pest detection

Remerciements

Arrivant à l'aboutissement de ce travail, je me trouve dans l'obligation du devoir de présenter mes chaleureux remerciements et de témoigner ma gratitude à tous ceux qui ont contribué à l'élaboration de ce mémoire.

Je tiens profondément à remercier mes encadrants absolument remarquables : le Pr Bernard KAMSU-FOGUEM du Laboratoire Génie de Production (LGP) de Tarbes, le Pr Dantouma KAMISSOKO de l'université de Ségou, Dr Daouda TRAORE de l'université de Ségou. C'était une chance d'apprendre et de travailler à vos côtés. Votre confiance, votre disponibilité et vos différents conseils ont permis la réalisation de ce travail et m'ont donné le goût de la recherche scientifique.

Je tiens à remercier profondément les membres du jury qui m'ont fait l'honneur de juger ce travail. Merci à M. Denis POMORSKI et M. Engelbert MEPHU NGUIFO, qui m'ont fait l'honneur de rapporter sur ma thèse. Merci pour vos commentaires et suggestions en vue d'améliorer ce travail. Merci à Mme Annig LACAYRELLE, Mme Carmen GERVET, d'avoir accepté d'examiner cette thèse.

Je tiens à remercier le personnel du LGP pour leur accueil et leur franche collaboration. Un remerciement particulier à Eliane CASTA pour son aide constante dans les démarches administratives.

Mes remerciements sincères à la coopération SCAC, Campus France pour le financement de cette thèse à travers la convention cotutelle entre l'Institut National Polytechnique de Toulouse (INPT) et l'Institut de Pédagogie Universitaire (IPU) ex-ISFRA.

Enfin mes remerciements vont à ma famille et mes collègues de l'université de Ségou pour m'avoir supporté, soutenu et encouragé durant ces années d'étude.

Table des matières

Résumé	i
Abstract	iii
Remerciements	v
Introduction générale	1
Contexte	1
Objectifs	5
Contributions	6
L'apprentissage profond à la rencontre de l'agriculture de précision : une analyse bibliométrique	6
Partage de réseaux de neurones convolutifs pour la classification multi-labels des images	6
Classification des maladies du mil à l'aide d'un réseau de neurones convolutifs profonds et de l'apprentissage par transfert	7
Explicabilité des réseaux de neurones convolutifs pour la localisation des insectes ravageurs dans les cultures	8
Organisation du mémoire	8
1 Etat de l'art	11
1.1 Introduction	11
1.2 Définitions	11
1.2.1 Apprentissage automatique	11
1.2.2 Vision par ordinateur	12
1.2.3 Classification d'images	13
1.3 Concepts de l'apprentissage profond	14
1.3.1 Apprentissage profond ou Deep Learning	14
1.3.2 Réseau de neurones artificiels	18
1.3.3 Réseaux de neurones convolutifs	19
1.3.4 Architectures des réseaux de neurones profonds	25
1.3.5 Explicabilité des réseaux de neurones profonds	26
1.3.6 Métriques d'évaluation	31
1.3.7 Framework et ensemble de données	34
1.4 Agriculture de précision	36
1.5 Apprentissage profond et limitation des données	41
1.5.1 Augmentation des données	41

1.5.2	Apprentissage par transfert	42
1.6	Les bases de données utilisées	45
1.6.1	Maladie du mildiou du mil	46
1.6.2	IP102	46
1.6.3	PASCAL VOC	46
1.6.4	AMAZON RAINFOREST	47
1.6.5	NUS-WIDE	48
1.7	Conclusion	48
2	Apprentissage profond à la rencontre de l’agriculture de précision	51
2.1	Introduction	51
2.2	Techniques Bibliométriques	52
2.3	Analyse bibliométrique : Vue d’ensemble	53
2.3.1	Sources et auteurs les plus pertinents	53
2.3.2	Affiliation la plus pertinente	56
2.3.3	Top manuscrits par citations	56
2.3.4	Mots-clés les plus pertinents	58
2.3.5	Relations mots-clés, pays et sources avec un diagramme de flux	59
2.3.6	Répartition des travaux de recherche en agriculture de précision	59
2.3.7	Pays et publications des auteurs en correspondance	66
2.4	Discussion	67
2.5	Conclusion	68
3	Réseaux de neurones convolutifs pour la classification multi-labels des images	71
3.1	Introduction	71
3.2	Travaux connexes	72
3.3	Apprentissage multitâches	74
3.4	Méthodologie	75
3.4.1	Vue d’ensemble	75
3.4.2	Réseau de neurones profonds multi-branches pour les classifications multi-labels	76
3.4.3	Réseau de neurones à sorties multiples	77
3.4.4	Réseau de neurones à caractéristiques multiples	79
3.4.5	Algorithme de réseau neurones profonds multi-branches	80
3.5	Expérimentons	81
3.5.1	Jeux de données	81
3.5.2	Paramètres d’expérimentation	81
3.5.3	Métriques d’évaluations	82
3.5.4	Architecture CNN	82
3.6	Résultats	83
3.6.1	Evaluation de la performance prédictive	83
3.6.2	Comparaison statistique des classificateurs	83
3.7	Discussions	86
3.7.1	Analyse technique	87
3.7.2	Analyse statistique	87
3.7.3	Comparaison avec les algorithmes traditionnels et leurs explications	88
3.8	Conclusion	89

4	Classification des maladies du mil à l'aide d'un réseau de neurones convolutifs profonds et de l'apprentissage par transfert	91
4.1	Introduction	91
4.2	Travaux connexes	92
4.3	Méthodologie proposée	93
4.3.1	Vue d'ensemble du système	94
4.3.2	Approche basée sur l'extraction de caractéristiques	94
4.4	Expérimentation et résultats	95
4.4.1	Détails de l'entraînement	95
4.4.2	Résultats de l'expérimentation	96
4.5	Conclusion	97
5	Réseau de neurones convolutif explicable pour la reconnaissance des insectes nuisibles	99
5.1	Introduction	99
5.2	Travaux connexes	101
5.3	Méthode proposée	103
5.3.1	Extraction des caractéristiques	103
5.3.2	Classification	104
5.3.3	Production des cartes visuelles	105
5.3.4	Combinaison des méthodes de l'explicabilité pour une localisation fine des insectes ravageurs dans les cultures	106
5.4	Expérimentation	108
5.4.1	Paramètres d'apprentissage	108
5.4.2	Implémentation	109
5.5	Résultats et discussions	109
5.5.1	Résultat de la classification	110
5.5.2	Analyse qualitative	110
5.6	Combinaison des méthodes de l'explicabilité	112
5.7	Conclusion	114
	Conclusion et perspectives	117
	Bibliographie	119

Table des figures

1.1	Tâche de reconnaissances d'objet. De gauche à droite : classification, classification + détection, détection d'objets multiples, segmentation	13
1.2	Architecture du neocognitron [Fuk80]	15
1.3	Apprentissage automatique et Apprentissage profond [AZH ⁺ 21]	17
1.4	Réseau de neurones à plusieurs couches cachées [RZ18]	17
1.5	Réseau de neurones artificiels	18
1.6	Algorithme d'optimisation backpropagation (rétropropagation)	19
1.7	LeNet-5, un exemple d'architecture d'un réseau de neurones à convolution [LBBH98]. En entrée le réseau prend une image de taille 32x32 pixels.	20
1.8	Carte de caractéristiques [ZF14]	21
1.9	Opération de convolution [EC21]	21
1.10	Exemple de convolution de deux matrices. Chaque valeur de la matrice d'entrée représente un pixel.	22
1.11	Exemple d'opération de pooling (max-pooling) de taille avec un pas (stride) de 1. La taille du filtre est 2x2.	23
1.12	Couche complètement connectée [RZ18]	23
1.13	Opération dropout [SHK ⁺ 14]	24
1.14	Architecture VGG16	26
1.15	Module inception	27
1.16	Bloc résiduel standard. x est l'entrée d'un bloc résiduel. \mathcal{F} désigne une fonction à apprendre. [HZRS15]	27
1.17	Différentes architectures de ResNet [HZRS15]	28
1.18	Modèle d'apprentissage vu comme une boîte noire	28
1.19	Concept de l'intelligence artificielle explicable [GA19]	29
1.20	Champ de références en Explicabilité [XRvD20]	30
1.21	Visualisation d'image par les techniques de l'explicabilité [ACÖG18].	31
1.22	Validation croisée à 5 plis	35
1.23	Usage de la donnée sur les exploitations agricoles (Source : JEMS Group)	38
1.24	Apprentissage automatique pour des usages agricoles d'après l'étude couvrant une période de 2004 à 2018 [LBM ⁺ 18]	39
1.25	Traitement des données agricoles	40
1.26	Usages du numérique dans l'Agriculture	40
1.27	Augmentation de données	42
1.28	Principe de l'apprentissage par transfert	43
1.29	Sélections des domaines sources.	45
1.30	Exemple d'images de la maladie du mildiou	46

1.31	Exemples d'images tirées de l'ensemble de données IP102	47
1.32	Exemples d'images issues PascalVOC	48
1.33	Exemples d'images issues d'Amazon rainforest	48
1.34	Exemples d'images issues NUSWide	49
2.1	Processus bibliométrique	52
2.2	Courbe des productions scientifiques annuelles	54
2.3	Les 10 principales sources les plus pertinentes	54
2.4	Affiliations les plus pertinentes	56
2.5	Mots-clés les plus pertinents	59
2.6	Visualisation du flux de citations par le diagramme de Sankey	60
2.7	Publication des pays pour les auteurs en correspondance	67
2.8	Développement d'un système agricole intelligent (SSD : Solid-State Drive; UAV : Unmanned Aerial Vehicle)	68
3.1	Architecture de base d'un réseau de neurones convolutifs	76
3.2	Réseau de neurones à sorties multiples.	77
3.3	Réseau de neurones à caractéristiques multiples.	79
3.4	Procédure de classification d'un réseau avec des caractéristiques multiples	80
3.5	Procédure de classification d'un réseau de neurones à sorties multiples	81
3.6	Diagramme critique pour Nemenyi test. NetMF : réseau de neurones à ca- ractéristiques multiples et NetMO : réseau de neurones à sorties multiples.	86
4.1	Vue d'ensemble de la méthodologie proposée [CKKT19].	94
4.2	Approche proposée pour l'extraction de caractéristiques [CKKT19].	95
5.1	Système d'identification des parasites basé sur l'explicabilité du réseau de neu- rones convolutifs	103
5.2	Ajustement de l'architecture d'inception-v3 pour la classification des insectes ravageurs.	104
5.3	Relation entre l'entropie, l'entropie conjointe et l'information mutuelle	106
5.4	Alignement des images en une seule image en utilisant les opérateurs OR, AND, and XOR	107
5.5	Exemples de résultats de classification explicables. « class » désigne la vérité terrain, « pred » classe prédite.	111
5.6	Carte d'explication d'une observation avec Grad-CAM sur les couches pro- fondes mixed0 à mixed10	112
5.7	Paramètres d'apprentissage affectant la visualisation LIME.	112
5.8	Alignement des images avec les mesures de l'information mutuelle (MI) sui- vantes : MI (image 1 et 2) = 0.33 ; MI (image 1 et 3) = 0.56 ; MI (image 2 et 3) = 0.398. Les images image 1, image 2 et image 3 correspondent respectivement aux résultats des méthodes Grad-CAM, LIME et Integrated Gradient	113
5.9	Alignement des images avec les mesures de l'information mutuelle (MI) sui- vantes : MI (image 1 et 2) = 0.20 ; MI (image 1 et 3) = 0.36 ; MI (image 2 et 3) = 0.25. Les images image 1, image 2 et image 3 correspondent respectivement aux résultats des méthodes Grad-CAM, LIME et Integrated Gradient	113
5.10	Un exemple d'alignement négatif avec MI (image 1 et 2) = 0.20	114

Liste des tableaux

1.1	Librairies pour l'explicabilité des réseaux de neurones	32
1.2	Liste des frameworks les plus courants	35
1.3	Jeux de données de référence	36
1.4	Appliquer l'apprentissage par transfert	44
1.5	Quantité d'échantillons de types de parasites provenant de l'ensemble de données IP102	47
2.1	Top-5 des publications des auteurs par rapport à H Index, TC (Total de Citation) et NP (Nombre de Publication)	55
2.2	Documents les plus cités	57
2.3	Présentation des travaux de recherche sur la détection de mauvaises herbes . . .	60
2.3	Présentation des travaux de recherche sur la détection de mauvaises herbes . . .	61
2.3	Présentation des travaux de recherche sur la détection de mauvaises herbes . . .	62
2.4	Présentation des travaux de recherche sur les maladies de culture	62
2.4	Présentation des travaux de recherche sur les maladies de culture	63
2.4	Présentation des travaux de recherche sur les maladies de culture	64
2.5	Présentation des travaux de recherche sur la reconnaissance des parasites	64
2.5	Présentation des travaux de recherche sur la reconnaissance des parasites	65
2.6	Publication des pays pour les auteurs en correspondance.	66
3.1	Liste des hyper-paramètres	82
3.2	Comparaison des performances entre les différents classificateurs avec la mesure « précision ».	84
3.3	Comparaison des performances entre les différents classificateurs avec la mesure « rappel ».	84
3.4	Comparaison des performances entre les différents classificateurs avec la mesure « f1-score ».	85
3.5	Comparaison des performances entre les différents classificateurs avec la mesure « moyenne de precision ».	85
3.6	Rang de classement de tous les classificateurs.	86
4.1	Ensemble de données	96
4.2	Les paramètres d'entraînement (hyper-paramètres)	96
4.3	Résultats d'expérimentations	97
5.1	Valeurs des hyper-paramètres utilisés dans les expériences.	109
5.2	Comparaison de nos résultats avec ceux du travail original	110

Introduction générale

Contexte

L'agriculture est l'ensemble des pratiques ayant pour objet les activités de production végétale et animale. Avec l'augmentation des populations, ce secteur doit faire face aux demandes alimentaires croissantes tout en tenant compte des enjeux sociétaux (comme la disponibilité de la main-d'œuvre), environnementaux (comme la rareté de l'eau, la perte de biodiversité, la dégradation des terres . . .) et économiques (comme la perte de rendement, la faible production). Les contraintes de son développement sont devenues nombreuses compte tenu des variabilités saisonnières et un régime climatique extrême. À cet effet, il est plus que jamais nécessaire de trouver des pratiques innovantes pour le développement du secteur agricole.

Les opportunités offertes par les innovations technologiques ont fortement fait progresser l'agriculture vers une industrialisation. Cette mécanisation a donc fait passer l'agriculture à une nouvelle ère, l'agriculture 4.0 où l'agriculture intensive s'impose désormais en contribuant à l'efficacité des procédés de production agricole. Avec une forte croissance de la population mondiale, l'agriculture intensive a apporté de nouveaux défis dont la résolution conditionne l'amélioration de la qualité et la quantité des produits et du sol tout en respectant l'homme et l'environnement.

Le secteur subit ainsi une forte transformation numérique pour relever les nouveaux défis tout en augmentant la productivité des sols, des cultures et en consommant moins d'intrants. L'intégration du numérique a considérablement modifié les connaissances sur la gestion des parcelles avec l'utilisation des technologies innovantes, notamment le géopositionnement par satellite (Global Positioning System ou GPS), des capteurs embarqués sur les engins agricoles, des robots ou des drones.

L'usage de ces technologies incite les data-scientists et les agronomes à concevoir des outils et des techniques d'analyse pour organiser avec précision la gestion et la surveillance des parcelles, relever les nouveaux défis qui s'imposent (la détection des attaques fongiques, la prédiction du rendement des cultures, la pulvérisation de pointe). Ces nouvelles pratiques nécessitent une assistance technique des agriculteurs pour répondre à leurs besoins et les aider à optimiser (ou maximiser) le rendement de leurs cultures.

Toutefois, cette nouvelle forme de l'agriculture soulève aussi de nombreux problèmes et défis relatifs à la transition écologique du système agricole, tels que la réduction de la consommation d'énergie, la minimisation des intrants et la place de l'humain dans la mise en place des systèmes intelligents. Plusieurs sociétés agricoles proposent des équipements programmables, autonomes, dotés d'un système intelligent. À partir des données collectées sur les exploitations, ces systèmes sont capables de recommander des opérations spécifiques avec parcimonie. Par exemple, la proxy-détection (avec l'Internet des objets, smartphone, casque, lunette virtuelle, etc.) offre un avantage dans la surveillance des champs et des cultures et fournit des données

de précision (zone infectée, espèces de parasites ou d'adventices par exemple) aux agriculteurs ou conseillers agricoles dans les prises de décisions. La démocratisation du secteur et la sophistication des équipements ont produit un grand nombre de volumes de données diversifiées par la qualité (éléments structurés ou brutes, etc.), et la nature (images, valeurs numériques ou textuelles, etc.). Cette explosion quantitative de données massives (big data) peut devenir rapidement une source possible d'erreurs dans la prise de décisions pour le pilotage des travaux d'exploitation agricole. Le déploiement des solutions d'analyse ou de traitement de données contraignent les scientifiques à trouver de nouvelles manières de voir et d'analyser le monde. Ceci est notamment manifeste dans les sous-domaines de l'intelligence artificielle qui s'appuient sur des outils mathématiques, d'électronique, d'informatique pour permettre des progrès dans l'extraction de connaissances sur des ensembles de données.

Les technologies de l'intelligence artificielle (IA) ont eu un impact profond dans plusieurs domaines dans ces dernières décennies. Ces diverses applications ont contribué à une progression significative des solutions basées sur la vision par ordinateur lors du développement des systèmes automatisés, nouveaux et robustes. Au fil des années, les chercheurs ont donc construit plusieurs familles d'algorithmes selon la taille du problème considéré. On peut citer en autres : la méthode de k plus proches voisins (KNN), les machines à vecteur de support (SVM), les arbres de décision et les forêts aléatoires (Random Forest), les réseaux de neurones. Selon [Res17], le marché de l'IA en agriculture était évalué à 518,7 millions USD en 2017. Avec un taux de croissance de 32,7%, il va atteindre 312,4 milliards USD d'ici 2027. Avec une utilisation croissante de la technologie de vision par ordinateur, les applications agricoles ont transformé les pratiques conventionnelles de l'agriculture en opportunités de générations de configurations incroyablement plus productives. Leurs usages ont eu un impact très positif sur la production et la rentabilité du secteur. Parmi ces avantages dans l'agriculture, citons l'utilisation optimisée et précise des produits chimiques (pesticides, insecticides, ou intrants), la planification du travail des agriculteurs, la reconnaissance d'images de plantes, la surveillance continue et l'analyse de la santé des cultures ainsi que la réduction de la dégradation de l'environnement. De ce fait, l'agriculture est en mesure de faire face à de nombreuses contraintes nouvelles poussant les agronomes à chercher de nouveaux outils dans la chaîne de valeurs agricoles avec de meilleures perspectives. Les récents progrès de l'apprentissage profond ont tenté d'apporter des réponses encourageantes en y associant les connaissances expertes des agriculteurs [ALG⁺19, LBM⁺18, KP18, HS15].

En 2012, les réseaux de neurones profonds ont produit un résultat spectaculaire dans les travaux [KSH12] lors de la compétition « ImageNet Large Scale Visual Recognition Challenge (ILSVRC) » a fait renaître l'apprentissage profond. Par la suite, un ensemble de nouvelles méthodes d'analyse de données comme les réseaux de neurones profonds apparaissent. Elles constituent aujourd'hui l'écosystème du « Deep learning ou apprentissage profond ». L'apprentissage profond concerne aujourd'hui tous les secteurs d'activité : l'industrie, la télédétection, le traitement de la langue naturelle, la médecine y compris l'agriculture. Il s'inspire du fonctionnement du cerveau humain pour traiter et analyser des données. Il permet à la machine d'effectuer des opérations complexes sur des données brutes, en fournissant des prédictions avec des résultats prometteurs. Presque toujours alimenté par la disponibilité croissante de la puissance de calcul, l'amélioration de la connectivité et les données massives, l'apprentissage profond offre des moyens novateurs caractérisés par l'automatisation de la gestion des exploi-

tations agricoles.

La transformation numérique de l'agriculture soutenue par l'apprentissage profond constitue une nouvelle carte de méthodes et d'outils qui ont changé la vision des agriculteurs en apportant de nombreux avantages pour gérer les changements fréquents et les conditions extérieures dans les pratiques agricoles. Les solutions réalisées ont permis de compléter les technologies (agronomiques et informatiques) déjà mises en œuvre pour accroître la production et moderniser la surveillance des cultures en temps réel. Cette mise à niveau a stimulé les recherches scientifiques (pluridisciplinaires) pour la revitalisation des secteurs applicatifs avec de meilleures perspectives. Les scientifiques sont en mesure de construire des systèmes agricoles durables et robustes avec un grand potentiel pour l'atteinte de certains objectifs tels que l'identification des maladies, des cultures, de mauvaises herbes ou des insectes ravageurs, l'augmentation de la fertilité des sols ou réduire les faiblesses dans la chaîne d'approvisionnement. Leur succès dépend de la performance et de l'efficacité des modèles d'apprentissage et le développement d'approches sûres, efficaces, transparentes et compréhensibles par les agriculteurs. La bonne mise en œuvre de ces approches augmentera la capacité des agriculteurs à maintenir leur profit et à faire face aux diverses menaces des cultures et des sols. Pour cela, les techniques d'analyse des données avec les réseaux de neurones devront tenir compte des variations climatiques ou environnementales qui peuvent impacter ces mêmes données.

Par exemple, les réseaux de neurones convolutifs (CNN) profonds présentés par [LBBH98] ont eu un grand succès dans le domaine de la vision par ordinateur pour la reconnaissance d'objets. Ils ont obtenu un regain de performance dans les traitements d'images et peuvent être employés pour la reconnaissance des parasites, des maladies de cultures, l'établissement des cartes d'occupation des sols, etc. L'utilisation des CNN ne se limite pas à l'analyse d'image, nous pouvons les exploiter dès lors qu'il est possible d'extraire des sous-éléments caractéristiques des données (données tabulaires, signaux, etc.) comme les pixels d'une image. Les réseaux de neurones convolutifs se composent principalement de couches convolutives et d'autres couches de traitement qui extraient des cartes de caractéristiques sur un ensemble de données (d'images). Ensuite, des catégories d'étiquettes ou de valeurs sont exprimées en sortie du réseau pour désigner le résultat de la classification ou de la régression. Les couches d'un CNN renferment un nombre important de paramètres de calcul à apprendre et une organisation hiérarchique qui sont exploités par les algorithmes d'apprentissage dans la résolution du problème considéré. Mais, tout dépend de l'ampleur des données recueillies pour le fonctionnement des solutions. Plus le problème est compliqué, plus la quantité des données pour entraîner l'algorithme d'apprentissage est cruciale.

Le coût élevé de la collecte de données d'apprentissage (données d'entraînement, de validation et de test) est un facteur majeur qui freine la croissance et le développement de l'agriculture de précision. Néanmoins, certaines pistes d'amélioration concernant la production de données publiques peuvent apporter des contributions pendant les expérimentations ou le passage à l'échelle des algorithmes. Des projets agricoles comme PlantVillage [HS15] PlantNet [ALG⁺19], ou les ensembles de données de références IP102 [WZL⁺19], AgriPest [WLX⁺21] sont des exemples de base de données à accès public. Par ailleurs, nous pouvons avoir recours à des techniques d'augmentation de données, ou des techniques plus sophistiquées comme l'apprentissage par transfert ou les réseaux de neurones de type réseaux adverses génératifs (en anglais Generative Adversarial Networks ou GAN) pour enrichir, augmenter la taille des ensembles de données ou faire appel aux connaissances acquises pendant la résolution d'une nouvelle tâche.

L'utilisation des méthodes de l'apprentissage profond a apporté de nouveaux outils et a bouleversé les pratiques existantes de l'agriculture. Elles sont encouragées par les efforts de collaboration et de synchronisation des industries, des scientifiques, des institutions et des médias. Les potentialités de l'apprentissage profond en agriculture ont été prouvées dans de nombreux travaux de recherches [ASM17, CR20, CKKT19, KP18, LBM⁺18, ZLL⁺18], mais plusieurs aspects peuvent être ajustés :

(i) Améliorer les tâches de classification pour lesquelles nous avons des petits jeux de données d'entraînement.

(ii) Jusqu'à présent, les réseaux de neurones ont la capacité de réaliser des tâches complexes comme la classification d'images grâce à des modèles profonds dont les résultats ne sont pas bien compris. Alors qu'ils obtiennent de meilleurs scores en s'attaquant à des données de grandes variétés, il est nécessaire d'interpréter et d'expliquer les résultats.

(iii) Dans la plupart des travaux de recherches, l'échantillonnage des données d'apprentissage est aléatoire. Nous trouvons généralement les subdivisions suivantes : 80/20 (respectivement 70/30, 60/40) qui représente 80% de l'ensemble de données pour l'entraînement et la validation (respectivement 70%, 60%) et 20% sont pour l'ensemble de tests (respectivement 30%, 40%). Afin d'apporter des éléments de réponses à ces analyses, il est essentiel lors de la réalisation d'une approche basée sur l'apprentissage profond de : (i) considérer la perception des acteurs humains dans le processus d'apprentissage des modèles profonds, (ii) prendre en compte les tests statistiques pour expliciter l'échantillonnage des données et établir un schéma clair sur la comparaison des performances des classificateurs issus du processus d'apprentissage. En conséquence, dans cette thèse, nous proposons des approches et des techniques à destination des data-scientists et agriculteurs dans leurs tâches de surveillance des cultures. A cet effet, nos contributions ont porté sur deux aspects : améliorer le processus d'apprentissage des réseaux de neurones convolutifs, classifier et détecter des maladies et des insectes ravageurs dans les cultures qui menacent la production. Ces contributions nous ont conduit à présenter une optimisation du processus d'apprentissage automatique des données agricoles dans un premier temps. Le but de cette tâche est de permettre une meilleure prise en compte des exigences induites par les contraintes du domaine et d'évaluer les performances réelles (et l'impact) des nouvelles méthodes.

Dans un second lieu, nous avons identifié la maladie du mildiou (une attaque fongique) au niveau du mil, une des cultures vivrières des régions tropicales d'Afrique. Nous avons mis l'accent sur la faible quantité de données d'entraînement supervisé nécessaire pour former de tels classificateurs. La détection de maladies joue un rôle clé dans la détermination du rendement des cultures et des champs. Durant cette thèse, les réseaux de neurones convolutifs ont permis d'effectuer une analyse efficace afin de prédire la présence ou l'absence de la maladie à partir des données images du mil (tige, feuille ou épi).

Dans une proposition suivante, nous avons amélioré les tâches de classification des réseaux de neurones convolutifs en augmentant l'espace de caractéristiques des données d'apprentissage partagées entre les tâches. Cette étude présente deux nouvelles approches de classification multi-labels, axées sur les principes de l'apprentissage multitâches. Nos résultats ont été comparés à d'autres classificateurs standards pour démontrer la contribution (la pertinence) des méthodes proposées. Des tests statistiques sont effectués pour analyser la signification des résultats et établir une certaine cohérence de l'expérimentation en appuyant/indiquant l'importance des tests statistiques lors de la comparaison des classificateurs.

Les systèmes qui ont une vocation d’interagir avec des utilisateurs humains doivent être en mesure d’expliquer, de justifier leur comportement et les décisions qu’ils prennent afin que les utilisateurs puissent comprendre les actions suggérées. Durant de nombreuses années, les nombreuses recherches en apprentissage machine se sont beaucoup plus concentrées sur la performance des modèles d’apprentissage plutôt que sur leur compréhension, leur interprétation et leur capacité d’aide à la décision. Si nous comprenions ce que le modèle a appris, il est possible de garantir la qualité des résultats obtenus. Cette capacité est cruciale et importante pour les utilisateurs du domaine de l’agriculture. Dans cette thèse, nous nous intéressons aux propriétés de visualisation des prédictions d’un modèle profond afin d’obtenir des résultats significatifs. Plusieurs méthodes pour l’explicabilité des réseaux de neurones (comme Grad-CAM, LIME, Integrated Gradient, etc.) ont été explorées pour démontrer leurs complémentarités dans la localisation des objets. Nous avons ainsi analysé qualitativement la capacité d’explicabilité des réseaux de neurones lors de la localisation des insectes ravageurs des cultures. Au-delà d’une prédiction à base du calcul de probabilité, nous avons guidé le processus de la localisation d’insectes en employant la théorie de l’information mutuelle.

Les résultats ainsi obtenus dans nos travaux ont prouvé que, l’apprentissage par transfert, l’augmentation ou le partage des tâches constituent des moyens efficaces pour améliorer les performances des réseaux de neurones. Les différentes contributions de thèse ont aussi dévoilé que, l’exploration des techniques de l’explicabilité des réseaux de neurones est un moyen utile et bénéfique pour les agriculteurs afin d’éclairer le processus de résolution de problèmes liés aux défis de la nouvelle forme de l’agriculture qui est l’agriculture de précision. Enfin l’intérêt des tests statistiques constitue un enjeu dans l’évaluation de plusieurs modèles profonds.

Objectifs

Dans ce travail, nous nous sommes intéressés aux réseaux de neurones convolutifs, qui ont démontré une efficacité pour les tâches de reconnaissance d’objets. Nous avons étudié les différents apports scientifiques sur les réseaux de neurones profonds au travers d’une étude bibliométrique. L’intérêt a été d’apporter de nouvelles réponses aux défis de l’agriculture de précision, ce qui constitue notre motivation pour les travaux qui seront dans cette thèse. Ce travail s’intéresse donc à de nouvelles possibilités d’exploration des données (images) agricoles avec des méthodes ou des outils de l’apprentissage profond pour consolider la surveillance de la santé des cultures. Il s’agit de proposer de nouvelles approches qui pourront rendre les réseaux de neurones convolutifs efficaces et transparents durant une tâche de classification. Ces réseaux de neurones profonds utilisent une approche d’apprentissage supervisé où les données en entrées sont des images et la sortie des classes d’objets. Rappelons que l’utilisation des CNN exige l’utilisation de bases de données d’entraînement et de test. Pour cela, quatre bases de données publiques PascalVOC [EVGW⁺07, EVGW⁺12], NusWide [CTH⁺09], Amazon rainforest [Pla17], IP102 [WZL⁺19] et une base de donnée Mildiou du Mil [CKKT19] construite par nous-mêmes, ont été utilisées.

Contributions

Cette thèse apporte quatre contributions dans le domaine agricole en utilisant des modèles de classifications bâtis sur l'apprentissage profond. Un accent particulier est mis sur les réseaux de neurones convolutifs profonds (Deep CNN) à travers les méthodes de l'apprentissage multitâches ou le phénomène de l'explicabilité de ces réseaux de neurones. Nous passons en revue ces contributions ci-après.

L'apprentissage profond à la rencontre de l'agriculture de précision : une analyse bibliométrique

Les récentes avancées des technologies de la communication avec l'avènement des objets connectés ont révolutionné le domaine agricole. Dans cette nouvelle ère numérique, le développement de l'intelligence artificielle notamment l'apprentissage profond a permis des accélérations et des progrès dans le traitement des données collectées. Pour mettre en lumière l'évolution et les avancées observées de l'apprentissage profond face à l'agriculture, nous avons mené une étude bibliométrique sur 440 travaux recherches datant de 2016 à 2021. Les analyses effectuées sur ces travaux de recherches suggèrent que l'apprentissage profond est fortement impliqué dans la digitalisation du monde agricole avec une grande précision et dépasse les techniques standards de traitement de données (images). La grande majorité des travaux se focalisent sur les problèmes de classification des cultures, l'identification de mauvaises herbes et des insectes nuisibles. Les techniques adoptées utilisent principalement l'architecture des réseaux de neurones convolutifs et de manière marginale l'architecture des réseaux de neurones récurrents et l'apprentissage par renforcement. À partir de ce cas d'étude, nous avons relevé trois défis essentiels qui sont indispensables dans la mise en œuvre des méthodologies basées sur l'apprentissage profond, à savoir :

(i) la nécessité de **considérer la perception des acteurs du domaine**, leur appropriation ou leur interaction avec les outils existants ; (ii) l'exigence d'effectuer **des calculs statistiques** pour l'échantillonnage des données d'apprentissage et l'analyser des performances des classificateurs issus du processus d'apprentissage ; (iii) l'**analyse et l'amélioration du processus d'entraînement d'un classificateur** sur les données expérimentales en faible quantité.

À la fin, nous avons proposé une optimisation du processus d'apprentissage constitué de plusieurs phases, afin de faciliter une meilleure prise en compte des exigences induites par les contraintes du domaine.

Partage de réseaux de neurones convolutifs pour la classification multi-labels des images

La résolution des problèmes par la classification multi-labels (MLC) est pertinente dans de nombreux domaines d'application : le diagnostic médical, l'annotation de textes, etc. Elle consiste à apprendre simultanément plusieurs étiquettes pour une donnée quelconque. Dans ce travail, nous utilisons la classification multi-labels pour construire un cadre générique d'analyse des images qui possèdent plusieurs étiquettes (catégories, tags ou labels).

Nous avons construit une architecture de classification multi-labels à partir d'un modèle de réseau de neurones multi-branches (MBNN). Le MBNN permet d'encoder des données images à partir de plusieurs sous-réseaux ou couches semi-parallèles, séparément. Il effectue une décomposition

de la tâche de classification avec une analyse appropriée des caractéristiques de l'image. Cependant, différents types de réseaux de neurones profonds peuvent être utilisés pour concevoir le MBNN, mais notre proposition se compose de réseaux de neurones convolutifs (CNN), formés et joints pour prédire les sorties. Le travail proposé permet donc d'effectuer des changements incrémentiels sur les architectures d'apprentissage multitâches (MTL) pour une adaptation au scénario multi-labels. Ces séries de transformation ou d'adaptation de données et d'architecture de réseau de neurones nous amènent à proposer deux méthodes d'apprentissage alternatives pour la classification multi-labels :

1. Une première méthode appelée « réseau de neurones à sorties multiples » où l'architecture est constituée d'un ensemble de classificateurs binaires disposés parallèlement, dont un classificateur est mis pour chaque label. Dans cette méthode, un CNN pré-entraînés est utilisé pour extraire les caractéristiques des images.
2. Une deuxième méthode appelée « réseau de neurones à caractéristiques multiples » où il y a plusieurs CNN parallèles pour extraire différentes propriétés caractéristiques des images. Ces CNN sont ensuite fusionnés en un réseau de neurones unique pour former le classificateur final.

Une contribution concerne les possibilités de généralité de l'approche proposée permettant la réutilisation dans d'autres systèmes d'apprentissage. Par ailleurs, la méthode d'apprentissage par transfert est engagée dans la prédiction des classes dont le but est d'obtenir un modèle performant et efficient. L'expérimentation a été validée avec les réseaux pré-entraînés sur l'ensemble de données ImageNet : VGG16 [SZ14], ResNet [HZRS15] et Inception-v3 [SVI⁺15]. Les démonstrations empiriques sont effectuées sur les ensembles de données de références (Amazon rainforest, Pascal VOC, et NusWide) et elles ont prouvé que la modélisation architecturale conduit à un apprentissage accru et à une meilleure performance du réseau de neurones.

Classification des maladies du mil à l'aide d'un réseau de neurones convolutifs profonds et de l'apprentissage par transfert

Les maladies des plantes ou des cultures sont des éléments importants de la réduction (du rendement) des qualités et des quantités de la production. Par conséquent, la détection et le diagnostic des maladies sont nécessaires dans les cultures pour répondre à cette préoccupation. En apprentissage profond, l'entraînement d'un modèle avec de petits ensembles de données est un défi majeur pour la communauté data-scientiste (scientifique). En outre, il est difficile et coûteux de générer manuellement des données étiquetées selon certains critères de sélection. Les approches qui utilisent les techniques de l'apprentissage par transfert visent à résoudre ce problème en découvrant et en appliquant les connaissances et les capacités apprises des tâches auxiliaires à de nouvelles tâches (d'un nouveau domaine par exemple).

Dans cette étude, nous proposons un apprentissage par transfert avec l'extraction de caractéristiques pour construire un système d'identification de la maladie du mildiou chez le mil. Notre approche consiste en trois étapes principales : (i) l'acquisition d'images (ii) la définition d'un réseau de neurones profonds composé d'un modèle pré-entraîné pour l'extraction de caractéristiques des images et (iii) la classification des maladies.

Le réseau de neurones pré-entraîné considère le CNN VGG16 qui est entraîné sur l'ensemble de données ImageNet. Nous avons un petit jeu de données de 126 images de mil malades et saines pour identifier le mildiou. L'application de l'apprentissage profond en agriculture permet

une analyse des données rapide et intéressante (de manière pertinente et efficace). L'avantage attendu de la proposition est de fournir un soutien aux parties prenantes (chercheurs et agriculteurs) grâce aux informations et aux connaissances générées par le processus de raisonnement d'une agriculture de précision. Ce travail a conduit à une publication dans une revue internationale.

Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., Traore, D., 2019. Deep neural networks with transfer learning in millet crop images. *Computers in Industry* 108, 115 – 120. <https://doi.org/10.1016/j.compind.2019.02.003>

Explicabilité des réseaux de neurones convolutifs pour la localisation des insectes ravageurs dans les cultures

Les attaques fongiques des cultures sont déterminantes pour la sécurité alimentaire puisqu'elles diminuent le rendement et la qualité de la production. Un des éléments responsables de ces ravages des cultures est relatif aux insectes nuisibles. Le dépistage à temps à partir de l'image des cultures constitue un véritable défi dans le contexte de l'agriculture de précision et de la vision par ordinateur. De nos jours, les avancées techniques des réseaux de neurones profonds dits convolutifs ont permis d'avoir de meilleurs résultats dans tous les domaines y compris en agriculture dans la gestion ou dans la surveillance de la santé des cultures. Malgré ces résultats satisfaisants des modèles obtenus par l'entraînement de ces réseaux lors des tâches de classification des images, l'un des inconvénients est la difficulté de déchiffrer ce que le modèle a appris ou d'interpréter les décisions qu'il prend.

La méthode proposée dans ce travail consiste à identifier et localiser les insectes ravageurs dans les cultures à l'aide d'un réseau de neurones convolutifs (CNN) tout en intégrant des techniques de visualisation pour comprendre les décisions des modèles. Notre approche de classification et de visualisation fournit aux agriculteurs une méthode et un outil d'aide à la décision pour relever le défi lié à la surveillance des cultures dans la lutte contre les insectes nuisibles.

Pour construire notre modèle de classification et de localisation d'insectes et évaluer ses performances, notre étude utilise plus de 75 000 images pour 102 catégories différentes de ravageurs issues du jeu de données de référence IP102. L'expérimentation a utilisé la combinaison des techniques de visualisation comme LIME (Local Interpretable Model-agnostic Explanations), Integrated Gradient ou Gradient-weighted Class Activation Map (Grad-CAM) pour interpréter formellement la localisation des insectes. Le processus de la combinaison est guidé par la théorie de l'information mutuelle. Les résultats obtenus offrent des opportunités aux utilisateurs du domaine des sciences des plantes pour améliorer la transparence du raisonnement produit par le CNN et utiliser plus efficacement l'apprentissage profond dans le processus de diagnostic des cultures.

Organisation du manuscrit

La structure de cette thèse est organisée en cinq (5) chapitres. Le chapitre 1 fournit un état de l'art sur les techniques de l'intelligence artificielle. Il introduit les concepts et mots-clés permettant de situer le cadre de notre travail. Ensuite, nous présentons en détail les concepts de l'agriculture de précision, de l'apprentissage profond, de l'architecture des réseaux de neurones

convolutifs et de l'apprentissage par transfert.

Dans le chapitre 2, nous examinons les travaux de recherches de l'apprentissage profond dans le secteur agricole. À travers une étude bibliométrique depuis Web of science, nous avons classé plus de quatre cents (400) travaux de recherches selon les défis à relever dans l'agriculture de précision, puis dégager les forces et les faiblesses de ces travaux. Ce chapitre présente, dans les différentes sections, un nouveau cadre méthodologique pour les futurs systèmes d'apprentissage agricole, en tenant compte de l'environnement extérieur au système tel que l'utilisateur.

Les chapitres 3 à 5 discutent de nos autres contributions. Chaque chapitre commence par une introduction du contexte suivi de la description des travaux connexes. Ensuite, nous décrivons et évaluons la méthode proposée sur un ou plusieurs ensembles de données.

Le chapitre 3 décrit une nouvelle approche de l'apprentissage profond, la classification multi-labels appliquée à la classification d'images. Dans ce contexte, deux méthodes d'apprentissage profond multi-branches et multitâches ont été développées pour résoudre le problème de classification multi-labels. Nous utilisons des réseaux de neurones convolutifs pré-entraînés bien connus pour développer ces approches.

Le chapitre 4 présente une méthode de classification pour reconnaître la maladie du mildiou du mil. Le mil est une culture des régions d'Afrique et d'Asie. Cette culture est favorable à l'attaque du mildiou qui conduit une perte énorme de productions en termes de rendement. Ce chapitre fournit une description détaillée de la méthode proposée ainsi que les résultats expérimentaux obtenus.

Le fonctionnement des réseaux de neurones profonds permet d'apprendre à partir d'un jeu de données certains comportements, afin d'associer une entrée à une sortie. Cependant, leur interprétation reste floue, car les calculs qui se produisent entre les deux ne sont pas faciles à expliquer aux humains qui les utilisent. L'étude de ce comportement qui qualifie de « boîte noire » les réseaux de neurones profonds permet d'augmenter la confiance dans les résultats produits. Cette préoccupation est étudiée dans le chapitre 5 qui présente des cartes de visualisation pour comprendre les décisions du réseau de neurones profonds. Il décrit également dans ses différentes sections comment identifier et localiser les insectes ravageurs dans les cultures à l'aide d'un réseau de neurones convolutifs.

Le dernier chapitre conclut cette thèse avec les perspectives.

Chapitre 1

Etat de l'art

1.1 Introduction

Dans ce chapitre, nous présentons un état de l'état de l'art sur les concepts qui sont pertinents pour ce travail. Nous commençons par définir les terminologies de l'apprentissage automatique et ses types qui sont l'apprentissage supervisé et non supervisé dans la section 1.2. Nous passons en revue dans la section 1.3, l'apprentissage profond. Nous présentons particulièrement les réseaux de neurones convolutifs, l'apprentissage par transfert et les méthodes pouvant donner un aperçu des prédictions d'un modèle. Nous accordons une attention particulière à l'agriculture de précision qui constitue la base de cette thèse dans la section 1.4. Au fil des années grâce à des dispositifs automatiques intelligents et à des modèles prédictifs qui utilisent des algorithmes d'apprentissage automatique, les pratiques agricoles subissent d'énormes changements. Ils sont à l'origine de la révolution agricole expliquée dans cette section. Les techniques de contournement du problème de la limitation des données sont expliquées dans la section 1.5. Les algorithmes d'apprentissage automatique ont besoin de données pour apprendre. Dans la section 1.6, nous présentons les jeux de données utilisés dans ce travail de thèse pour l'entraînement des réseaux de neurones conçus. La section 1.7 conclut ce chapitre.

1.2 Définitions

Dans cette section, nous définissons quelques concepts pertinents pour ce travail, à savoir l'apprentissage automatique, l'apprentissage supervisé/non supervisé, la vision par ordinateur et la classification des images.

1.2.1 Apprentissage automatique

L'apprentissage automatique est un sous-ensemble de l'intelligence artificielle. Elle correspond au fait de donner à la machine la capacité d'apprendre avec des données sans être explicitement programmée. En général, un problème d'apprentissage considère un ensemble composé de plusieurs observations de données et tente ensuite de prédire les propriétés des données inconnues. Cependant, les problèmes d'apprentissage se divisent en plusieurs catégories :

— **Apprentissage supervisé**

Lors d'un apprentissage supervisé, les données d'entrées sont associées à des sorties que nous voulons prédire. L'objectif est d'apprendre une règle générale liant les entrées aux sorties. Une sortie peut être une valeur réelle, discrète ou catégorique (ou étiquette). Par exemple, les modèles de régression estiment les sorties à valeurs réelles, tandis que les modèles de classification estiment les sorties à valeurs discrètes. Un algorithme d'apprentissage supervisé analyse les données et produit une fonction inférée qui peut être utilisée pour mapper de nouveaux exemples (non étiquetés) à la sortie désirée.

— **Apprentissage non supervisé**

En apprentissage non supervisé, l'algorithme d'apprentissage automatique reçoit un ensemble de données d'entrée sans aucune valeur de sortie (cible) correspondante. On le laisse seul pour qu'il trouve les groupes d'exemples similaires dans les données en entrée (groupement) ou déterminer la distribution des données dans l'espace d'entrée (estimation de la densité) ou projeter les données d'un espace à grande dimension vers un espace à dimension réduite (e.g, deux ou trois) pour la visualisation.

1.2.2 Vision par ordinateur

Aujourd'hui, il est difficile de dissocier le progrès de l'agriculture et la vision par ordinateur qui est une application de la science de l'informatique. Elle combine les caméras, les logiciels et l'intelligence artificielle pour permettre à un système d'acquies une compréhension de haut niveau à partir d'image ou de vidéos¹. En plus, elle utilise l'apprentissage profond pour former des réseaux de neurones qui guident le traitement et l'analyse des images [LWW⁺21, MCDG21, ALY⁺21, RHGS15]. Une fois qu'ils sont totalement entraînés, les modèles sont capables de reconnaître des objets, de détecter, de reconnaître des personnes ou suivre leurs mouvements par exemple. Cependant, les tâches de reconnaissances d'objets à partir d'une image (Figure 1.1) peuvent être séparées en trois domaines avec une complexité croissante :

- **Classification** : c'est le problème le plus connu en détection d'image. Elle consiste à détecter la présence ou non d'un objet dans une image, c'est-à-dire classer une image dans une catégorie.
- **Détection d'objet** : elle vise à donner la position de l'objet dans l'image en l'encadrant d'une boîte appelée bounding box qui englobe l'objet. C'est un problème de classification et de recherche d'une ou plusieurs instances.
- **Segmentation** : elle permet de représenter les instances d'objets dans une image. Ainsi on parle de segmentation sémantique lorsqu'on ne peut faire la différence entre deux instances d'un même objet alors que la segmentation d'instances va distinguer les différentes instances d'un même objet.

En agriculture, l'usage de la vision par ordinateur et l'apprentissage profond permettent d'imaginer un grand nombre d'applications : le guidage automatique des engins agricoles, la réduction de l'usage des pesticides grâce à la détection automatique des maladies et adventices ou des insectes dans les cultures.

1. <https://www.intel.com/content/www/us/en/internet-of-things/computer-vision/overview.html>

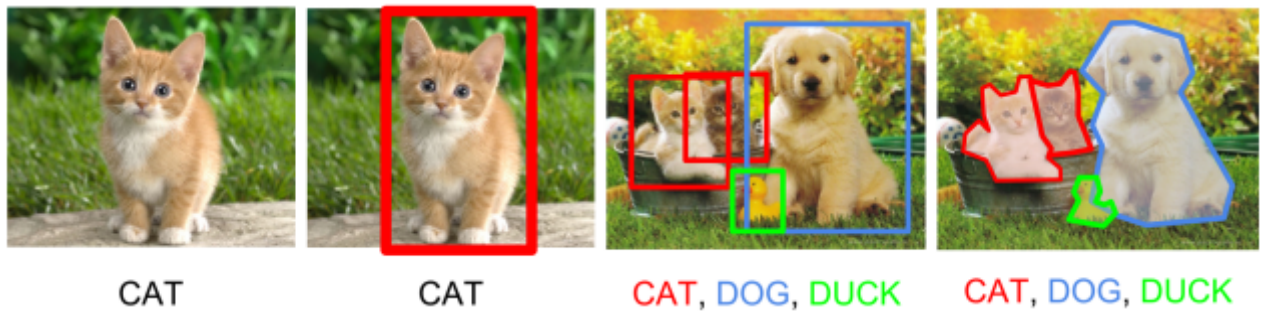


FIGURE 1.1 – Tâche de reconnaissances d’objet. De gauche à droite : classification, classification + détection, détection d’objets multiples, segmentation

1.2.3 Classification d’images

La classification automatique d’images est une méthode de reconnaissance d’images qui consiste à attribuer automatiquement une classe (catégorie ou étiquette) à une image à l’aide d’un système de classification. Dans le cadre d’un apprentissage supervisé, une tâche de classification utilise des données annotées (ou données d’entraînement) pour construire un modèle capable de classer ou d’annoter des images non vues ou non annotées appartenant à l’ensemble de données de test. Chaque image des données d’entraînement est associée à une classe correspondant à la catégorie où elle appartient. Dans l’apprentissage non supervisé, les images n’ont pas d’étiquettes. Il revient au modèle de trouver une règle qui fait correspondre chaque image à un groupe donné. Cependant, il existe différents types de classifications d’images qui sont les suivants :

- **Classification multi-classes** : la tâche de classification comporte plus de deux classes, mais chaque image est associée à une et une seule classe.
- **Classification binaire** : la classification comporte deux et uniquement deux classes disjointes (notée 1 et 0), soit l’image est de la classe 1 ou ne l’est pas, un 0.
- **Classification multi-labels** : une tâche de classification où un ensemble d’étiquettes (plus d’une classe) est attribué à une image.

Dans ce travail, nous ne traitons que le cas de l’apprentissage supervisé dont le principe suppose que chaque image possède une étiquette. Les trois types de classification sont tous utilisés dans cette thèse pour les tâches de classification d’images qui comprend plusieurs étapes :

Préparation des données d’apprentissage : elle divise l’ensemble des données en trois sous-ensembles de données : entraînement, validation et test sur la base des étiquettes. Au préalable, les images peuvent subir un prétraitement qui consiste à les « nettoyer » afin d’obtenir des éléments représentatifs du problème considéré où un encodage des noms de classes d’une chaîne de caractères en numérique est souvent nécessaire.

Construction du réseau de neurones : elle permet de construire un classificateur ou un modèle profond. Durant cette construction, l’architecture du réseau est définie, il s’agit de la taille de l’entrée du réseau, les couches et leurs relations, la sortie et les paramètres d’entraînement comme l’optimiseur, la fonction perte, une métrique pour observer les étapes d’entraînement ou de test.

Apprentissage : elle comprend une étape d'extraction des caractéristiques qui décrit les informations (caractéristiques) pertinentes contenues dans l'image. L'apprentissage utilise le classificateur conçu précédemment. Le classificateur détermine la classe d'une image en entrée du système tout en minimisant l'erreur de classification, grâce aux données de validation. Ce classificateur est ensuite utilisé pour classer les images non vues (appartenant aux données de test ou du terrain) durant son entraînement. Les données de validation sont donc bien utiles pour trouver une bonne configuration du modèle de classification. En effet, l'entraînement d'un modèle profond implique toujours des réglages comme le choix du nombre de couches, leurs tailles ou le taux d'apprentissage. Ces réglages sont possibles en utilisant les retours d'informations de l'ensemble des données de validation.

Evaluation de la performance du modèle de classification : détermine sa capacité à classer correctement les images des données test. La phase d'évaluation va déterminer ce qui est correct dans le modèle, c'est-à-dire à quel point ces prédictions sont justes. Les performances sont calculées sur l'ensemble des données d'entraînement, de validation et de test avec des métriques incluant la précision (precision), le rappel (recall), la justesse ou l'exactitude (accuracy).

Lors de la classification d'images, toutes ces étapes sont essentielles pour la construction et l'évaluation d'un système de classification.

1.3 Concepts de l'apprentissage profond

Nous donnons le résumé d'un noyau de cette thèse, l'apprentissage profond, en section 1.3.1. Nous présentons dans les sections 1.3.2, 1.3.3 et 1.3.4 une brève revue des réseaux de neurones artificiels, puis convolutifs et des exemples d'architectures standards définies pour les réseaux de neurones dits profonds. L'interprétation, la compréhension du fonctionnement et de la prédiction des réseaux de neurones sont des sujets brûlants au cœur de l'actualité de recherche. Ces mécanismes trouveront leurs explications dans la section 1.3.5 qui parle de l'explicitabilité d'un réseau de neurones. Comment peut-on mesurer la performance d'un réseau de neurones ? Les métriques employées pour y parvenir sont formulées dans la section 1.3.6. Enfin, les références en termes de logiciels et ensembles de données que nous pouvons exploiter pour développer les réseaux de neurones profonds passent dans la section 1.3.7.

1.3.1 Apprentissage profond ou Deep Learning

L'apprentissage profond est une branche de l'apprentissage automatique fondé sur les réseaux de neurones qui traite les données en s'inspirant du fonctionnement des réseaux de neurones des cerveaux humains. Il s'agit d'un ensemble d'outils et de méthode d'apprentissage automatique basés sur une utilisation avancée des réseaux de neurones. Il a permis de progrès important dans de nombreux secteurs d'activité : le jeu, l'industrie, la médecine, l'enseignement y compris l'agriculture pour le traitement et l'analyse des données. Il permet à la machine d'effectuer des calculs complexes sur des données volumineuses, brutes et produit des résultats prometteurs. Il peut utiliser l'apprentissage supervisé et non supervisé et réalise des tâches de type humain sans intervention humaine. Pour y parvenir, les réseaux de neurones profonds requièrent une large

1.3. Concepts de l'apprentissage profond

quantité de données pour former un bon prédicateur.

Dans l'apprentissage profond, les représentations sont en couches, structurées, empilées les unes sur les autres et manipulées par des réseaux neurones. En conséquence, il utilise des modélisations architecturales formées à partir de ces nombreuses couches plus ou moins profondes. Une couche contient des unités qui transforment les données d'entrée en informations que la couche supérieure peut l'utiliser. Grâce à cette structure un réseau de neurones apprend au travers de son propre traitement de données.

Les réseaux de neurones profonds (réseaux de neurones à plusieurs couches) sont le fruit d'une longue histoire de développement des réseaux de neurones, qui remonte au perceptron de Frank Rosenblatt [Ros57] qui crée un algorithme de reconnaissance de formes. Les premiers systèmes d'apprentissage profond du type perceptron multi-couches commencent avec les travaux de Ivakhnenko [Iva68, Iva71, IL65]. Ils utilisent des modèles de fonctions d'activation polynomiales. En dehors de ces réseaux, le neocognitron de Kunihiko Fukushima [Fuk80] est peut-être le premier réseau de neurones artificiels qui mérite l'attribut "profond". Fukushima a introduit les premiers réseaux de neurones convolutifs (appelés CNN ou convenet) conçus avec plusieurs couches de convolution et de couches d'abandon ou pooling (Figure 1.2). Il a été utilisé pour la reconnaissance des caractères manuscrits et des formes. Les neocognitrons modernes sont capables de mieux faire grâce aux rétropropagations ou descente du gradient pour minimiser le taux d'erreur de l'apprentissage.

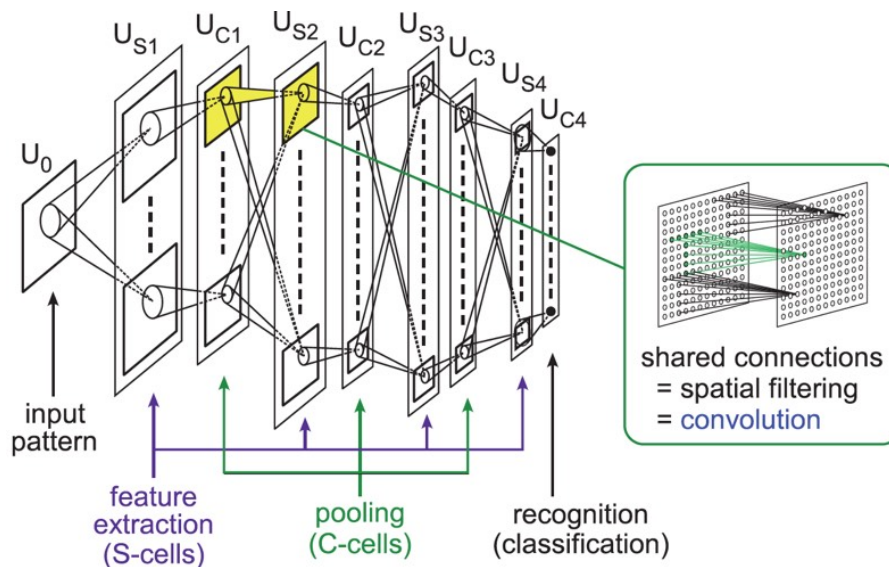


FIGURE 1.2 – Architecture du neocognitron [Fuk80]

Les réseaux de neurones convolutifs sont vus comme une forme particulière de réseau de neurones artificiels multi-couches (Multi Layer Perceptron ou MLP) dont l'architecture des connexions est inspirée de celle du cortex visuel des mammifères. Dans les années 1990, Yan LeCun [LBD⁺89] a présenté la première démonstration des réseaux de neurones convolutifs, qui utilisent la théorie de rétropropagation sur la reconnaissance de chiffres manuscrits. Ces réseaux n'admettent que des connexions locales, de plus, les paramètres de ces connexions sont partagés au sein d'une couche donnée. Cela réduit considérablement le nombre de paramètres. Les modèles prédictifs (modèle de classification ou de régression) sont entraînés sur de données

historiques pour prédire de futurs résultats. L'extraction de caractéristiques est cruciale dans le processus de modélisation prédictive. Dans certaines situations, la transformation des données devient essentielle avant de construire un tel modèle. Elle implique la transformation des données brutes en caractéristiques pertinentes afin de réduire les erreurs de modélisation et d'améliorer les performances prédictives du modèle. Dans une démarche d'analyse d'image classique par exemple, si on cherche à reconnaître une attaque fongique, les méthodes utilisées étaient basées sur la réflectance, la morphologie, la texture. Cette extraction des caractéristiques demande une bonne connaissance de la tâche de prédiction et un travail d'ingénierie permet d'adapter la méthode d'extraction. Une mauvaise extraction conduit donc à une mauvaise performance de l'algorithme d'apprentissage. Cette méthode d'extraction a été largement simplifiée par la capacité des réseaux de neurones profonds qui extraient eux-mêmes les caractéristiques dans les couches du réseau (Figure 1.3). Chaque couche apprend une représentation des données en son entrée qui doit être utile à la couche suivante. Ainsi les informations obtenues dans chacune des couches sont de plus en plus riches (voir la Figure 1.4). Grâce à cette capacité de généralisation, un réseau de neurones entraîné pour une tâche peut être réutilisé pour une tâche similaire, dans un autre contexte. En ce moment, on parle d'apprentissage par transfert 1.5.2 où il est possible de changer certaines couches pour l'adapter à la nouvelle tâche.

Un processus d'apprentissage automatique est très coûteux en temps et en calcul, notamment pour tout ce qui concerne la préparation des données. Depuis l'arrivée des unités de traitement graphique avec le processeur graphique ou Graphics Processing Unit (GPU), la disponibilité des données massives et de calcul parallèle les réseaux de neurones profonds ont connu un retour avec succès. Alors qu'ils pouvaient être lents, les GPU permettent d'accélérer les routines et constituent un élément central des réseaux de neurones. Ainsi, ils ont commencé à dépasser les algorithmes classiques de l'apprentissage automatique comme les machines à vecteurs de support. Avec les données massives, les réseaux peuvent continuer à améliorer leur qualité d'apprentissage en donnant de meilleurs résultats. Leur premier grand succès de cette nouvelle ère fut probablement le travail de [KSH12]. Ils ont développé et entraîné le réseau de neurones profonds AlexNet sur de grandes quantités de données (ImageNet) d'entraînement. AlexNet est une amélioration de l'architecture LeNet5 construite par Yan LeCun des années auparavant. Les expériences menées chez Google et Microsoft pour la reconnaissance d'images ont produit des résultats notables, comme la victoire éclatante de l'architecture de [RDS⁺15] lors de la compétition de reconnaissance d'objets.

Grâce aux réseaux de neurones profonds, les modèles d'apprentissage profond sont maintenant au cœur de nombreuses réussites récentes sur des problèmes dont les solutions font appel à des techniques de l'intelligence artificielle. Parmi ces tâches, on peut citer la reconnaissance d'objets ou la segmentation de scènes dans les images, la reconnaissance des objets, de la parole dans les signaux, la compréhension du langage naturel.

Les réseaux de neurones profonds permettent un traitement hiérarchique multi-échelles des données et sont capables de catégoriser les informations des plus simples aux plus complexes. L'augmentation du nombre de couches permet au réseau de neurones de détecter de légères variations des données d'entrées qui n'influencent pas le résultat du réseau. Pour un objet par exemple, la première couche détecte de petits contours élémentaires, la seconde assemble ces contours en motifs, puis les motifs en parties d'objets, et enfin ces parties en objets.

Une phase d'apprentissage du réseau sur des objets connus est nécessaire pour trouver les

1.3. Concepts de l'apprentissage profond

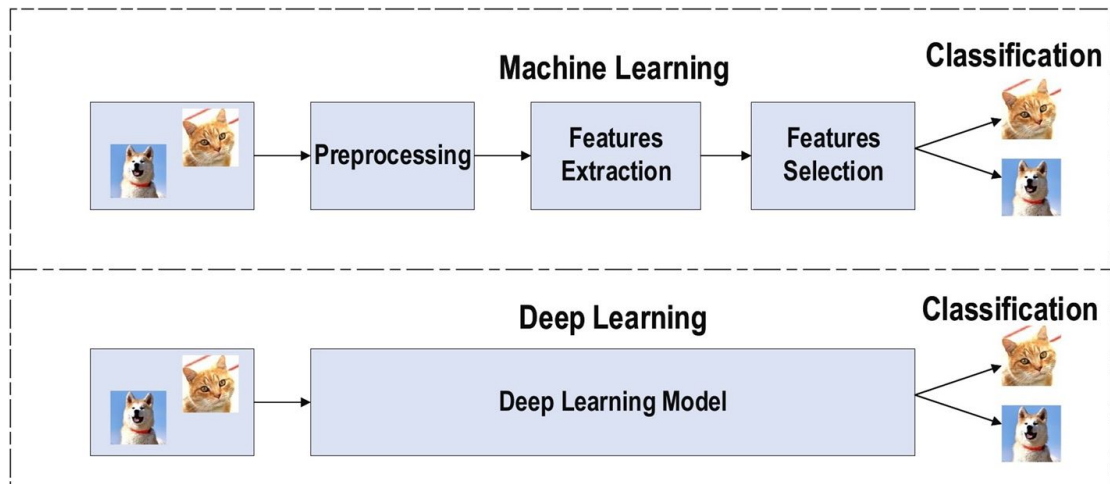


FIGURE 1.3 – Apprentissage automatique et Apprentissage profond [AZH⁺21]

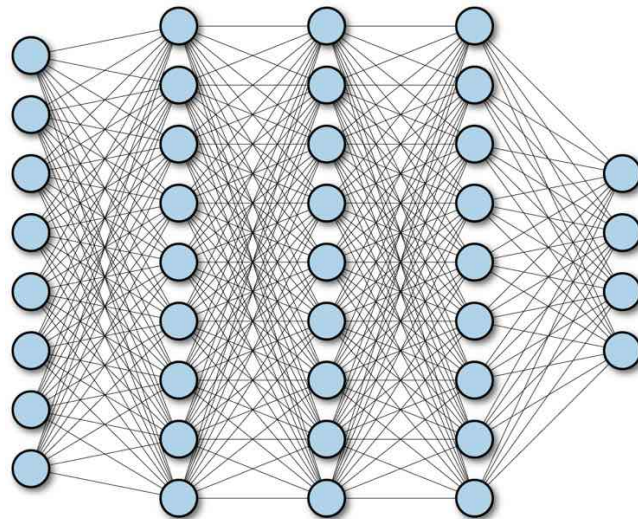


FIGURE 1.4 – Réseau de neurones à plusieurs couches cachées [RZ18]

meilleurs paramètres d'entraînement. L'un des enjeux étant de trouver des méthodes pour ajuster ces paramètres le plus rapidement et le plus efficacement possible en tirant donc parti des symétries permettant la détermination des invariants. Cette symétrie se retrouve dans l'architecture des réseaux de neurones convolutifs dans lesquels les poids des neurones sont eux invariants par translation. L'entraînement des modèles d'apprentissage fait appel à des algorithmes d'optimisation dont un des plus importants est le principe de la descente de gradient qui permet de trouver le minimum de n'importe quelle fonction convexe en convergeant progressivement vers celui-ci.

L'organisation et le principe de fonctionnement des réseaux de neurones leur permettent de traiter des problèmes non linéaires d'une manière plus efficace que le type de modèle standard comme celui de l'apprentissage automatique. Cependant, ces interprétations restent floues dans de nombreux cas d'où la nécessité des méthodes de l'explicabilité pour comprendre ce que les réseaux ont appris. Cette problématique est expliquée dans la section 1.3.5.

1.3.2 Réseau de neurones artificiels

Le neurone artificiel représente l'unité de calcul des réseaux de neurones. Il prend en entrée plusieurs signaux ou valeurs numériques x_i d'un autre neurone ou d'une source externe et génère une valeur en sortie y (comme représentée sur la Figure 1.5). A chaque entrée x_i , un poids w_i est associé en fonction de son importance par rapport aux autres entrées. La sortie est calculée en appliquant une fonction d'activation f ou de seuillage qui active la sortie lorsque la somme des poids associés aux entrées augmentées d'une autre entrée appelée biais qui a un poids b dépasse un certain seuil. La valeur de la sortie y est calculée par la formule 1.1.

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (1.1)$$

où f est une fonction d'activation.

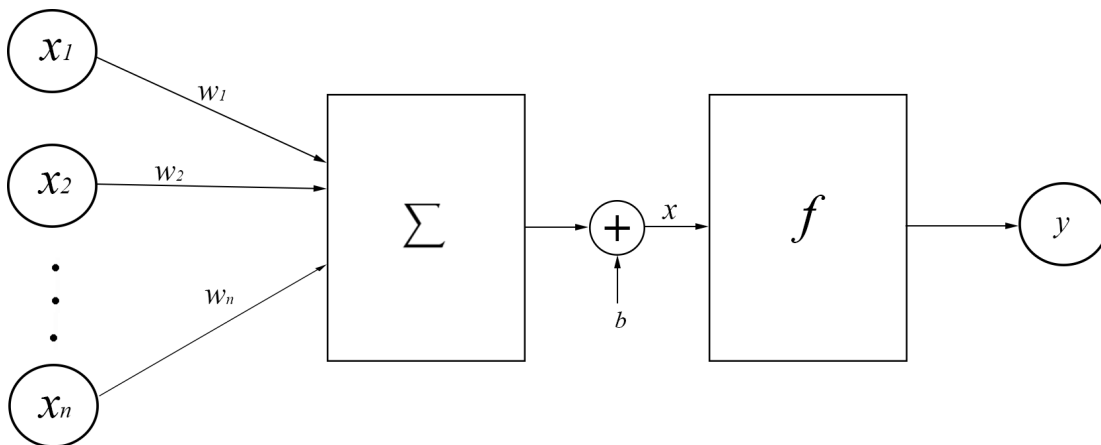


FIGURE 1.5 – Réseau de neurones artificiels

En général les réseaux de neurones apprennent à minimiser l'erreur entre la vérité terrain (valeur réelle notée y_i) et celle prédite (valeur réalisée notée \widehat{y}_i) lorsqu'un échantillon est fourni au modèle en mettant à jour les poids. Ce processus est appelé rétropropagation et repose sur l'hypothèse qu'une fonction de perte (Loss function) ou fonction objective $j(\theta)$ peut-être minimisée par la descente du gradient afin de trouver un minimum (Figure 1.6). Plus sa valeur diminue, plus le réseau est performant.

La fonction objective $j(\theta)$ est minimisée en modifiant la valeur des paramètres θ du réseau sur un ensemble de données d'entraînement. Les paramètres d'apprentissags θ représentent les différents poids du réseau.

Dans le cas d'un problème de régression ou une certaine valeur est prédite, il est courant d'utiliser la fonction de perte de l'erreur quadratique moyenne (mean squared error ou MSE), définie par la formule 1.2.

$$MSE = \frac{1}{n} \sum_{i=1}^n E_i^2 \quad (1.2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{y}_i)^2$$

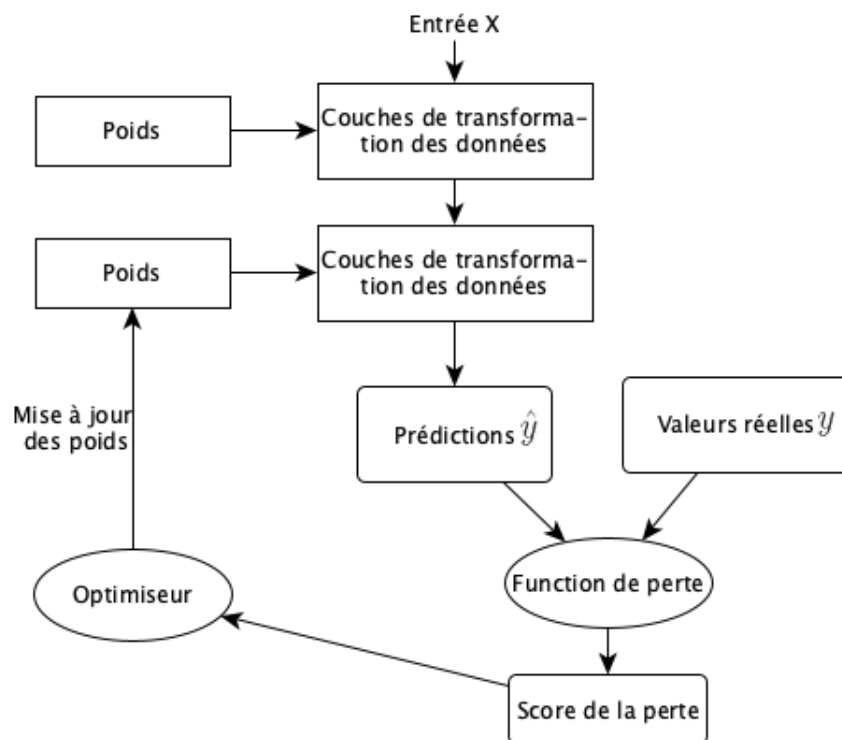


FIGURE 1.6 – Algorithme d’optimisation backpropagation (rétropropagation)

Où n est le nombre d’observations, E étant l’erreur locale se focalisant sur une observation donnée en comparant la valeur réelle et sa valeur prédite : $E_i = y_i - \hat{y}_i$. Pour les tâches de classification binaire ou multi-classification, la fonction de perte est basée sur le calcul de l’entropie croisée. Elle quantifie la différence entre deux distributions probabilistes. Une petite valeur représente un meilleur modèle contrairement à d’une grande valeur. Les fonctions 1.3 et 1.4 définissent l’entropie croisée binaire (binary cross-entropy) et catégorique (categorical cross-entropy).

$$L = \frac{1}{n} \sum_{i=1}^n y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (1.3)$$

$$L = - \sum_{i=1}^n y_i \log(\hat{y}_i) \quad (1.4)$$

La couche de sortie du réseau de neurones est définie avec n nœud (un pour chaque classe) et plus la valeur de la fonction perte est proche de 0, plus elle est fiable est de 0 et meilleur est le réseau de neurones.

1.3.3 Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs (CNN) sont introduits à la fin des années 80 par LeCun [LBBH98]. Un CNN est une cascade d’opérations linéaires et de non-linéaire. Chaque couche

est calculée à partir de la couche précédente. La Figure 1.7 présente le modèle architectural CNN.

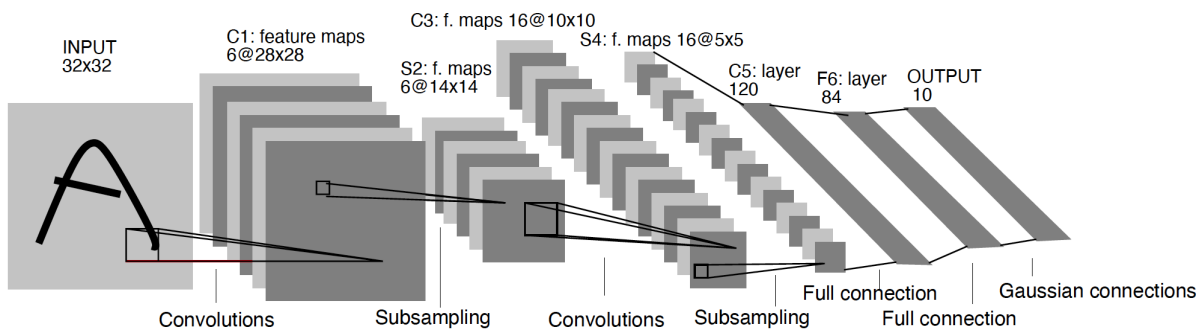


FIGURE 1.7 – LeNet-5, un exemple d'architecture d'un réseau de neurones à convolution [LBBH98]. En entrée le réseau prend une image de taille 32×32 pixels.

Aujourd'hui, ces réseaux sont devenus un standard pour résoudre les tâches de vision telles que la reconnaissance d'images. Techniquement, un réseau de neurones convolutifs se compose généralement d'une combinaison des couches suivantes :

- Couche de convolution (convolutional layers),
- Couche de mise en commun (pooling layer),
- Couche d'activations (activation layer),
- Couche complètement ou entièrement connectée (fully-connected ou dense layer).

1.3.3.1 Couche de convolution

Un CNN repose sur des calculs mathématiques appelés convolutions ou produits de convolution. Il s'agit de la couche la plus importante. En entrée d'une couche de convolution, nous avons une image ou une carte de caractéristiques et en sortie, une carte de caractéristiques extraites grâce à des convolutions (Figure 1.8). A l'origine, la convolution est un filtrage de l'image qui consiste à traiter l'image en un bloc pour mettre en évidence des motifs ou réduire un type de bruit. Il suffit de la diviser en différents petits carrés à analyser séparément (Figure 1.9). Chaque bloc de convolution contient au moins une couche de convolution définie par :

- La « taille du noyau de convolution », généralement de dimension 2 avec un noyau carré
- Le « nombre de filtres de convolution » définie sous forme tenseur de dimension $3(h, w, c)$ avec h la hauteur, w la largeur et c la profondeur de l'image. Un filtre sur une image en couleur a une profondeur de 3.
- Le « pas de convolution ou stride », c'est le décalage du noyau de convolution. Par exemple un stride de $(1, 1)$ signifie un déplacement horizontal vers la droite de 1 pixel, verticalement vers le bas de 1 pixel.
- Le « padding », ajoute des pixels autour de l'image. Par exemple, le zero-padding signifie la valeur 0, autour de l'image.

Une couche de convolution correspond à une combinaison de plusieurs convolutions appliquées sur les cartes de caractéristiques. Cela va permettre aux couches d'extraire des caractéristiques de plus en plus complexes. Par exemple, pour le visage d'une personne, les

1.3. Concepts de l'apprentissage profond

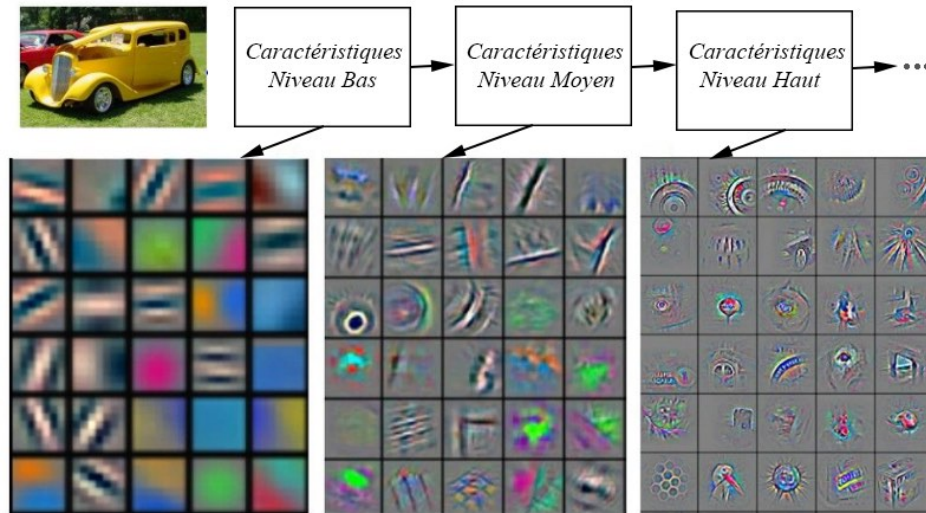


FIGURE 1.8 – Carte de caractéristiques [ZF14]

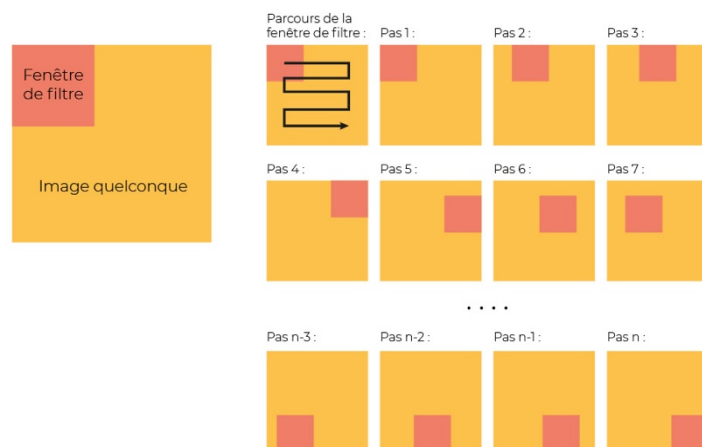


FIGURE 1.9 – Opération de convolution [EC21]

premières couches de convolution peuvent extraire les traits ou les arrondis de l'image tandis que les couches qui suivent peuvent être utilisées pour reconnaître le nez ou la bouche et reconnaître le visage entier par la suite.

En mathématique, une opération de convolution entre deux fonctions, f et g est définie comme suit :

$$(f * g)(x) = \int_{-\infty}^{\infty} f(z)g(x-z)dz$$

Dès que nous avons des objets discrets, l'intégrale se transforme en une somme. Ainsi nous obtenons la formule suivante :

$$(f * g)(i) = \sum_a f(a)g(i-a)$$

Dans le cas des images, les deux fonctions sont en général définies sur un espace à deux dimensions correspondant à une convolution 2D. Dans ce cas la convolution discrète correspond à une somme pour une image f et un noyau g

$$(f * g)(i, j) = \sum_a \sum_b f(a, b)g(i-a, j-b)$$

L'exemple suivant illustre le fonctionnement de l'opérateur de convolution qui se déroule en deux étapes (Figure 1.10) :

- Multiplication d'une partie de l'entrée (matrice) par le filtre de convolution ;
- Sommation de toutes les valeurs du produit de matrice résultante.

128	97	x	1	0	=>	128	0	=	128+22=150
35	22		0	1		0	22		

FIGURE 1.10 – Exemple de convolution de deux matrices. Chaque valeur de la matrice d'entrée représente un pixel.

1.3.3.2 Couche de mise en commun (pooling)

Les couches à convolution sont empilées pour former le réseau de neurones profonds. Elles sont accompagnées de couches dites de pooling, qui permettent de réduire la taille des cartes de caractéristiques. Celle-ci peut se faire suivant plusieurs opérateurs dont les plus courants sont la moyenne (Avg-Pooling) et le maximum (Max-Pooling). La couche pooling est un ré-échantillonnage des données, qui est appliqué à un noyau de taille $n \times n$ sur la carte de caractéristiques. Un Max-Pooling (agrégation par valeur maximale) consiste à agréger les valeurs d'une carte de caractéristiques en retournant la valeur maximale de la partie d'image fournie par le noyau (comme montré par la Figure 1.11). De manière équivalente, le AvgPooling (agrégation par la valeur moyenne) agrège les valeurs d'une carte de caractéristiques en récupérant la

1.3. Concepts de l'apprentissage profond

moyenne des valeurs. Les couches de mise en commun ne sont pas en réalité des couches de neurones, elles mettent en corrélation les motifs détectés par les couches convoluées dans les couches précédentes. Le pooling offre l'avantage de réduire le volume du calcul pour analyser les données par la réduction des dimensions et le contrôle ainsi que le sur-apprentissage.

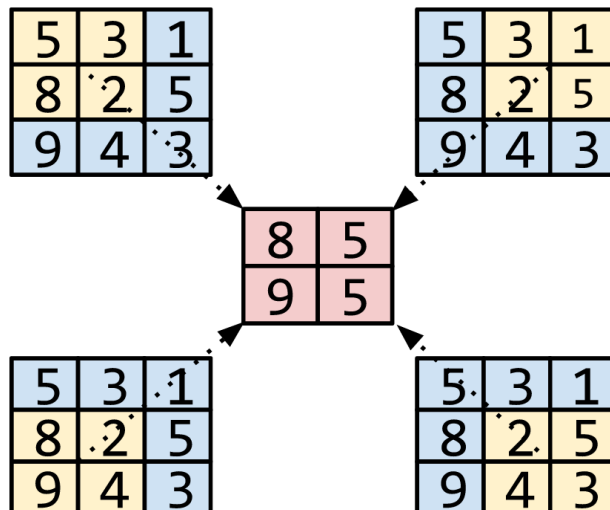


FIGURE 1.11 – Exemple d'opération de pooling (max-pooling) de taille avec un pas (stride) de 1. La taille du filtre est 2x2.

1.3.3.3 Couche complètement connectée

Dans le cadre d'une tâche de classification, un réseau de neurones profonds se termine par une ou plusieurs couches complètement connectées. C'est une couche cachée qui est connectée à tous les nœuds des couches qui le précèdent (Figure 1.12) dont le but est d'utiliser toutes les caractéristiques extraites par les couches précédentes pour former la sortie finale (e.g, classer une image d'entrée). En effet, elle renvoie un vecteur de taille N défini lors de sa construction. Par exemple, sur la dernière couche ou la couche de sortie, N représente le nombre de classes.

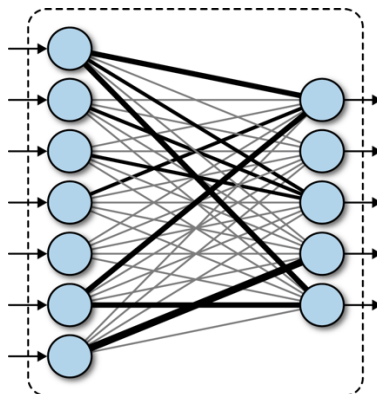


FIGURE 1.12 – Couche complètement connectée [RZ18]

Les couches entièrement connectées sont généralement suivies d'une couche d'abandon ou **Dropout** [SHK⁺14] qui agit sur le poids des couches du réseau en désactivant certains neurones

(Figure 1.13). C'est une technique de régularisation couramment utilisée dans les réseaux de neurones pour contrôler le surapprentissage pouvant être causé par un nombre important de paramètres.

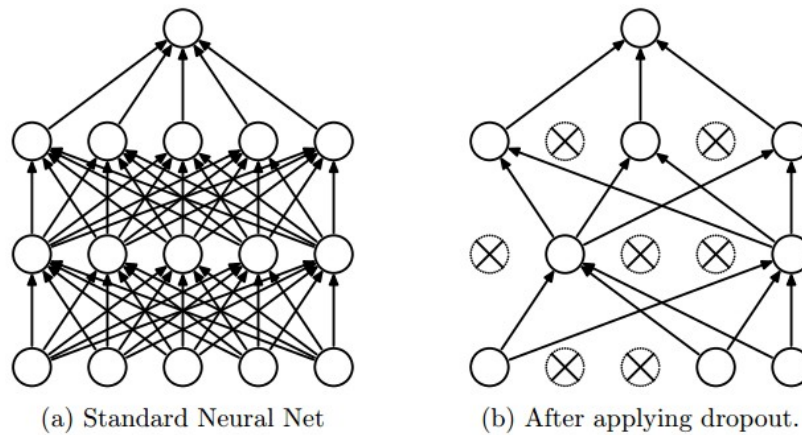


FIGURE 1.13 – Opération dropout [SHK⁺14]

1.3.3.4 Couche d'activation

Les couches d'activation sont généralement placées après chaque couche de convolution et couche complètement connectée. Les couches d'activation ne sont pas des couches de neurones, mais des couches définies à l'aide de fonctions d'activation ou de transfert qui propagent l'information de couche en couche. Elle prend la somme pondérée de toutes les entrées de la couche précédente, puis génère et transmet une valeur de sortie (généralement non linéaire) à la couche suivante. Les fonctions d'activation les plus courantes sont les suivantes :

La fonction Rectified Linear Units ou ReLU : définie si l'entrée est négative ou nulle, la sortie est 0. Si l'entrée est positive, la sortie est égale à l'entrée. Elle est calculée comme suit :

$$f(x) = \max(0, x)$$

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & \text{sinon} \end{cases}$$

La fonction sigmoïde : elle utilise une exponentielle en réduisant l'entrée entre 0 et 1. Sa formule est la suivante :

$$f(x) = \frac{1}{1 + e^{-x}}$$

La fonction tangente hyperbolique : elle prend une entrée de valeurs et réduit à une valeur entre $[-1, 1]$ comme suit :

$$f(x) = \tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

La fonction softmax : elle transforme un vecteur d'entrée en une distribution de probabilité dont leur somme est de $\sum_i \hat{y}_i = 1$. On l'utilise souvent dans la couche finale d'un modèle de classification, notamment pour les problèmes multi-classes. Sa formule est la suivante :

$$\hat{y}_i = \text{softmax}(x_i) = \frac{e^{x_i}}{\sum_j^N e^{x_j}}$$

1.3.4 Architectures des réseaux de neurones profonds

Dans le contexte des réseaux de neurones profonds, l'architecture d'un réseau a été reconstruite comme utile pour rendre l'apprentissage plus efficace. Par exemple, les architectures des réseaux de neurones convolutifs sont reconnues efficaces sur des données d'images [LBBH98, LKF10, SLJ⁺14]. Les architectures LeNet [LBBH98] et AlexNet [KSH12] sont introduites plus d'une décennie avant. Les deux réseaux étaient relativement peu profonds et sont composés respectivement de deux (2) et cinq (5) couches de convolutions. Ils employaient des noyaux de grands champs réceptifs dans les couches proches de l'entrée et des noyaux plus petits dans les couches qui sont plus proches de la sortie.

Après 2012, grâce au challenge de classification d'images sur la base de données ImageNet (ILSRV), nous avons assisté à une explosion des nouvelles architectures de plus en plus profondes avec des noyaux de petite taille. De plus en plus, ces architectures peuvent se déployer sur des appareils embarqués tels que les téléphones intelligents (smartphones) ou les drones. Les premiers [SZ14] à explorer les couches profondes pour répondre à ce challenge ont repris des travaux de [RDS⁺14] en construisant les architectures dites VGGNet (VGG16 et VGG19). Après cela, plusieurs autres architectures ont émergé. Elles sont constituées de plusieurs blocs de convolutions : GoogLeNet [SLJ⁺14], ResNet [HZRS15], DenseNet [HLvW18], MobileNet [HZC⁺17]. Ces architectures existent en plusieurs versions qui marquent la profondeur ou le nombre de couches possédées. Par exemple, ResNet existe en version ResNet18, 34, 50, 101 et ResNet152. GoogLeNet a été améliorée à plusieurs reprises en Inception v1, v2, v3, v4 et Inception-ResNet. Nous présentons ci-dessous, les CNN les plus courantes et qui sont utilisées dans cette thèse à savoir VGG16, Inception-v3 et ResNet152.

1.3.4.1 Visual geometry group (VGG)

Ce réseau est développé par Karen Simonyan et Andrew Zisserman de l'université d'Oxford en 2014 [SZ14]. Ce CNN s'est classé en deuxième position lors du challenge ILSVRC 2014. Il existe deux versions plus connues du VGGNet, **VGG16** et **VGG19** contenant respectivement 16 et 19 couches de convolutions. Ils sont considérés comme une référence pour l'extraction des caractéristiques dans l'ImageNet avec 138 millions de paramètres. Le réseau de neurones VGG16 (Figure 1.14) avec ses treize couches convolutives dispose des filtres homogènes de 3x3 avec un pas ou stride de 1px (voir la section 1.1.9), cinq couches de pooling et 3 couches complètement connectées. La taille d'entrée par défaut est de 224x224 pixels et la sortie fournit une classification dans une des 1000 classes.

1.3.4.2 GoogLeNet (Inception)

Lors de la compétition ILSVRC 2014, GoogLeNet [SLJ⁺14], aussi appelé inception-v1 a été le gagnant. Il a proposé le concept de bloc d'inception où l'entrée est transformée par plusieurs branches de convolutions parallèles (Figure 1.15) avec des filtres de tailles multiples (1x1 ; 3x3) sur le même niveau. Les branches peuvent être subdivisées en plusieurs autres branches parallèles dont les sorties seront ensuite fusionnées et envoyées au module suivant.

L'objectif principal est d'atteindre un haut niveau de précision tout en réduisant la quantité de calcul, mais aussi, le nombre de paramètres d'entraînement. Une autre factorisation proposée par Inception consiste à remplacer les convolutions de taille $n \times n$ par une combinaison de convolutions $1 \times n$ et $n \times 1$ pour limiter la complexité du calcul.

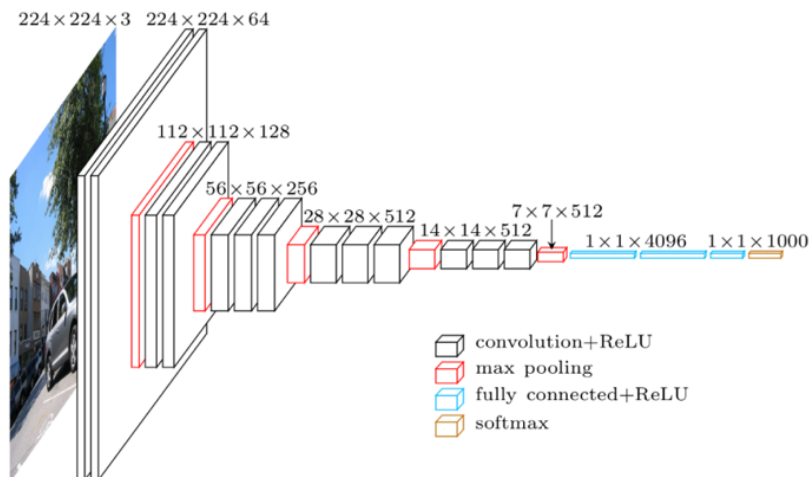


FIGURE 1.14 – Architecture VGG16

De très bons résultats sont obtenus en fixant n à 7 lors de la construction de Inception-v3. Ce CNN contient 48 couches de convolutions, 5 couches de type pooling pour réduire les paramètres d'entraînement à 23 millions. Inception-v3 introduit des filtres convolutifs parallèles en concaténant leurs sorties. Le réseau a une taille d'entrée d'image de 299×299 pixels.

1.3.4.3 Réseau de neurones résiduels (ResNet)

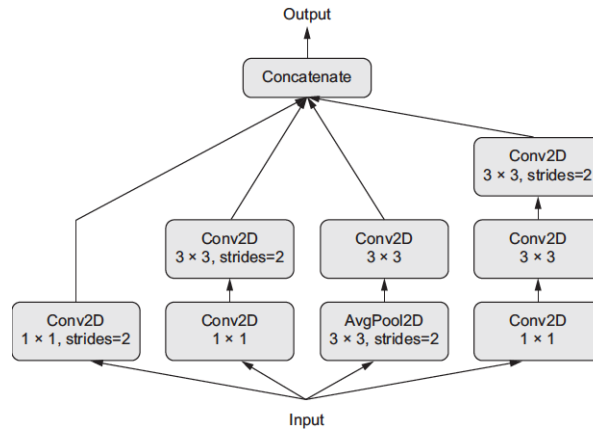
Ce réseau a été développé par Kaiming He et ses collègues [HZRS15] de Microsoft. Il a été le gagnant du concours ILSVRC-2015. La nouveauté de ResNet est la conception des blocs résiduels (Figure 1.16) pour réduire l'erreur d'apprentissage. Ces blocs résiduels sont utilisés pour construire des réseaux très profonds en s'attaquant au problème de la disparition du gradient dans les réseaux de neurones convolutifs. Ainsi, le gradient peut relier l'entrée d'une couche n à la sortie de la couche $(n + i)^{eme}$ où le bloc résiduel contient de deux couches de poids et d'une fonction d'identité (Figure 1.16).

Plusieurs types d'architectures ResNet sont développés en fonction du nombre de couches, de 18 à 152 couches. Par exemple, la version ResNet152 est composée de 152 couches avec 60 millions de paramètres. Le réseau a une taille d'entrée d'image de 224×224 . La différence entre ces architectures se situe au niveau des couches. Leurs détails sont mentionnés dans la Table 1.17 où chaque type est représenté en colonne. Chaque architecture contient une couche de convolution $conv1$ avec un noyau carré $k = 7$, un pas de 2. Cette couche est suivie d'un max-pooling de 3×3 avec un pas $s = 2$. Le reste se compose de 4 couches de convolutions ($conv2_x$, $conv3_x$, $conv4_x$ et $conv5_x$) suivies d'une couche de pooling et d'une couche entièrement connectée de dimension 1000 pour la classification avec l'ensemble des données ImageNet.

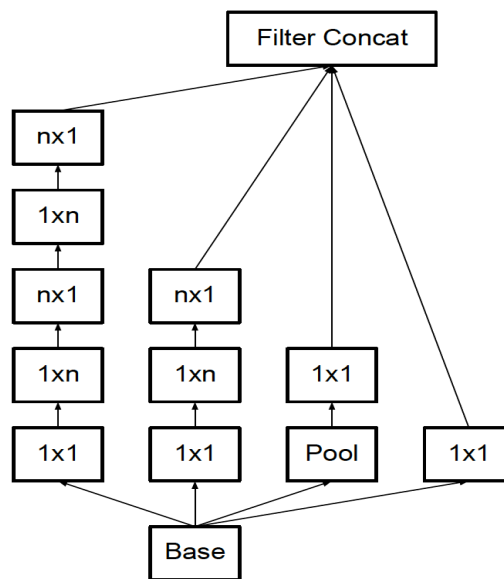
1.3.5 Explicabilité des réseaux de neurones profonds

Les réseaux de neurones profonds sont constitués d'entités connectées les unes aux autres de manière à former des couches plus ou moins profondes. Pour les tâches d'apprentissage, les résultats sont certes spectaculaires, mais la modélisation envisagée est complexe en raison

1.3. Concepts de l'apprentissage profond



(a) avec extension des filtres



(b) Après la factorisation de convolutions $n \times n$, avec $n = 7$

FIGURE 1.15 – Module inception

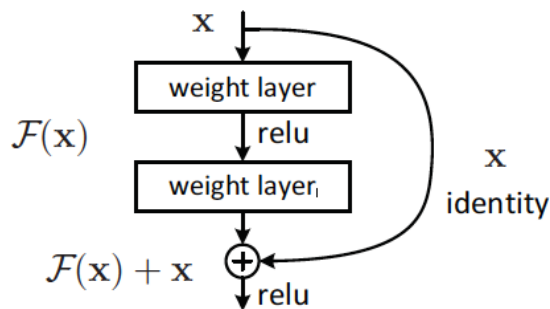


FIGURE 1.16 – Bloc résiduel standard. x est l'entrée d'un bloc résiduel. \mathcal{F} désigne une fonction à apprendre. [HZRS15]

du nombre élevé de paramètres d'entraînement, des relations entre les couches et d'opérations de calculs. Il est donc difficile de comprendre mathématiquement les raisons de ce succès, et

Nom de la couche	Taille de sortie	18 couches	34 couches	50 couches	101 couches	152 couches
conv1	112x112	7x7, 64, pas 2				
		3x3 max pooling, pas 2				
conv2_x	56x56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28x28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4_x	14x14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7x7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1x1	Moy pooling, 1000-d fc, softmax				

FIGURE 1.17 – Différentes architectures de ResNet [HZRS15]

encore plus difficile d'expliquer la prise de décisions pour les utilisateurs humains. Selon Doshi-Velez et Kim, les modèles (ou algorithmes) d'apprentissage automatique doivent être capables de fournir des preuves accessibles à un utilisateur humain dans un format sémantique complet, suffisant en soi [DK17]. Cependant, la nature en boîte noire (Figure 1.18) des modèles profonds leur confère une difficulté fondamentale pour comprendre les détails techniques et les prédictions qui sont critiques dans certaines tâches (diagnostic médical, détection des attaques fongiques des cultures). Cette propriété prouve que la performance des modèles est placée au centre des préoccupations au détriment de sa compréhension alors qu'elle a un rôle essentiel dans le futur des systèmes basés sur l'IA. Pour utiliser efficacement les réseaux de neurones profonds, il est important que les utilisateurs puissent être capables d'expliquer et de justifier leurs réponses.

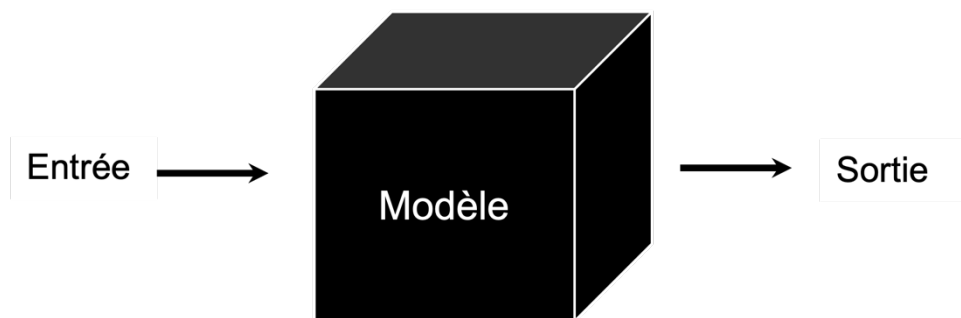


FIGURE 1.18 – Modèle d'apprentissage vu comme une boîte noire

L'intelligence artificielle explicable (eXplainable XAI) est un brûlant sujet de recherche dans le domaine de l'apprentissage profond visant à améliorer la compréhension de la boîte noire des systèmes d'IA, i.e comment ils font leur choix. Il s'agit de débloquent le fonctionnement en boîte noire des modèles profonds pour permettre leurs intégrations et leurs mises en œuvre dans les situations pratiques.

Ce domaine de recherche explore les mesures et modèles impliqués dans les prédictions et produit des informations de sortie (comme l'intensité des pixels) qui (Figure 1.19) : (a) aident à identifier les biais dans les modèles et fournissent des idées pour améliorer les modèles et les données d'apprentissage (b) améliorent la confiance, la robustesse et la fiabilité du modèle d'apprentissage, et (c) caractérisent la capacité du modèle d'apprentissage à expliquer ou représentent

1.3. Concepts de l'apprentissage profond

le résultat d'apprentissage avec des concepts qu'un humain (un utilisateur final ou un non-expert en apprentissage automatique) peut utiliser pour éclaircir l'analyse d'un système prédictif.

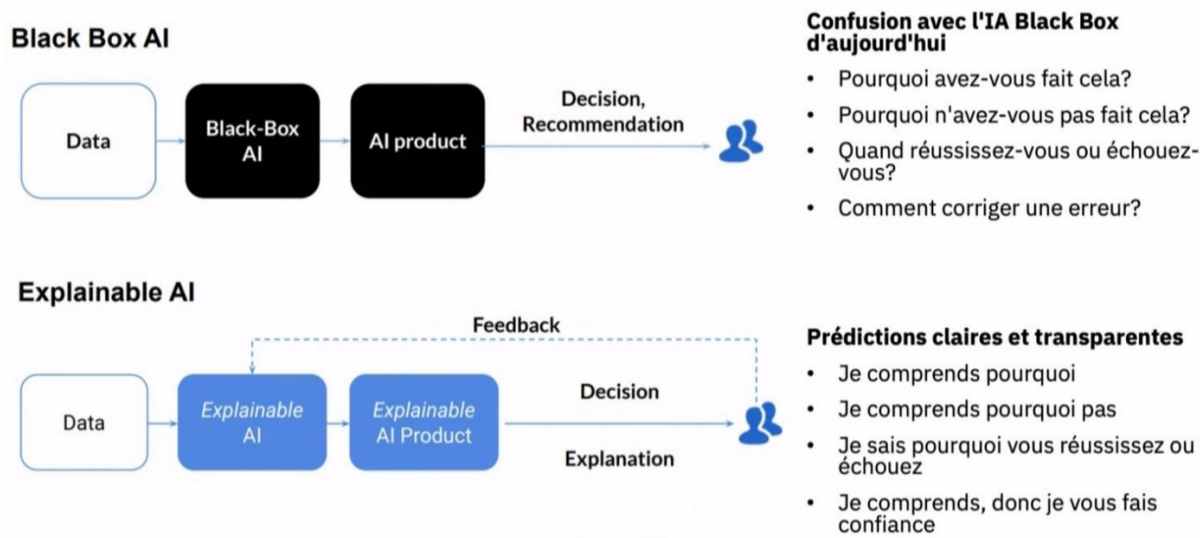


FIGURE 1.19 – Concept de l'intelligence artificielle explicable [GA19]

En conséquence, le développement des méthodes explicables est bénéfique pour atteindre le but recherché avec des garanties suffisantes dans le futur des systèmes issus de l'IA. Le défi de l'IA explicable consiste donc à créer des solutions à la fois compréhensibles et interprétables à des niveaux d'abstractions suivants [GBY⁺18, SM19] :

Explicabilité des données (pre-modelling explainability) : il s'agit d'identifier quelles données peuvent être utiles dans l'apprentissage du modèle et pourquoi. L'objectif est de déterminer les biais dans les données d'entraînement pour construire un modèle robuste.

Explicabilité des prédictions (Post-modelling explainability) : il s'agit de communiquer des informations caractéristiques lors des prédictions individuelles à travers des cartes thermiques pour visualiser l'importance de chaque pixel.

Explicabilité des modèles (Explainable modelling) : il s'agit d'expliquer les stratégies de prédiction par le modèle qui fournit des informations sur les couches et les paramètres d'apprentissage utilisés dans la prédiction.

En outre, de nombreux chercheurs proposent de discuter la classification des techniques de l'explicabilité [ABC⁺19, BADDS⁺20, DR20, DSB17, DK17, SWM17, SSL20, TG20, XRvD20]. Le plus récent [XRvD20], montre une description détaillée des nouveaux champs de recherche de l'IA explicable. Ils proposent une nouvelle taxonomie (Figure 1.20) pour organiser les nombreuses méthodes existantes et particulièrement les méthodes qui expliquent les réseaux de neurones profonds (Methods for Explaining Deep Neural Networks).

Les méthodes de visualisation (Visualization methods) permettent d'expliquer certaines parties du modèle d'apprentissage par des moyens de compréhensions visuelles tels que la rétropropagation ou les perturbations. Nous distinguons les catégories suivantes pour ce cas :

- Méthodes basées sur la rétropropagation (Backpropagation-based methods) : pour identifier la pertinence (saliency) des caractéristiques d'entrée sur la base de certaines exploitations des signaux (bruits) de gradient transmis de l'entrée à la sortie pendant l'entraînement du réseau. Les méthodes de Gradient Input, DeConvNet, Saliency

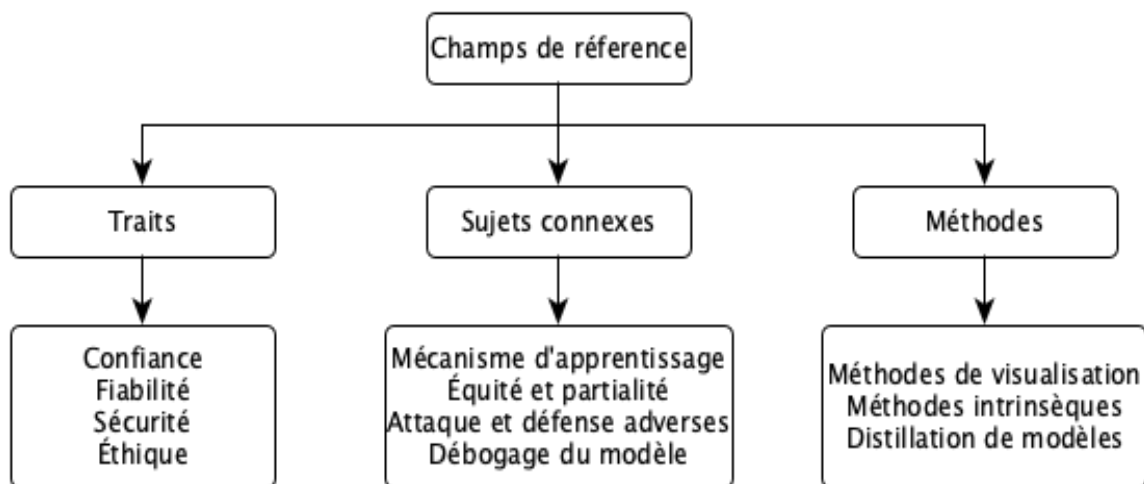


FIGURE 1.20 – Champ de références en Explicabilité [XRvD20]

Maps, Class Activation Mapping (CAM) ou Gradient-weighted Class Activation Mapping (GradCAM,) DeepLIFT, Deep Taylor Decomposition sont des exemples.

- **Méthodes basées sur la perturbation** : elles examinent l'influence de la modification des caractéristiques d'entrée sur la sortie. Ceci consiste à générer une visualisation en perturbant l'image d'entrée x_i . Il s'agit d'identifier les zones sensibles de x_i par rapport à la sortie y en supprimant ou en modifiant des informations sur cette image. La méthode Occlusion Sensitivity [ZF14] est type d'exemple de cette méthode.

Le modèle de distillation (Model Distillation) définit un modèle distillé (ou substitut) g qui apprend à imiter les actions (comportement entrée-sortie) d'un modèle original f sur un même espace de données. Le modèle distillé g , ou modèle en boîte blanche (glass-box) peut atteindre des performances raisonnables tout en fournissant une explication du modèle original f [RSG18, RSG16]. En d'autres termes, une "distillation de modèle" utilise le résultat d'un modèle d'apprentissage entraîné comme base de référence pour un modèle explicable. Cela nous permet d'interpréter la raison pour laquelle un DNN a fourni une étiquette de classe à une donnée d'entrée.

L'explicabilité intrinsèque (Intrinsic Methods) vise à concevoir des DNN qui fournissent une explication en même temps que leurs sorties. Ainsi, il est possible d'optimiser simultanément les performances et certaines qualités souhaitées des explications générées par le modèle. Les techniques d'explicabilité intrinsèque sont organisées en deux catégories :

- Les mécanismes d'attention (ou Attention Mechanisms) sont inspirés de la façon dont les humains prêtent attention à différentes parties d'une image ou d'autres sources de données complexes pour les analyser. Le module d'attention évalue la pertinence de chaque pixel en entrée étant donné une représentation de celui-ci et le contexte de prédiction.
- L'entraînement conjoint (ou Joint Training) fonctionne en ajoutant une tâche d'explication supplémentaire à la tâche initiale du modèle, et en entraînant simultanément les deux tâches.

1.3. Concepts de l'apprentissage profond

Les explications apportées par les méthodes de visualisation ou d'approximation locale (Figure 1.21) sont des outils d'aide à la décision qui peuvent aider à mieux comprendre le comportement des réseaux et donner plus de confiance dans les prédictions des modèles de classification.

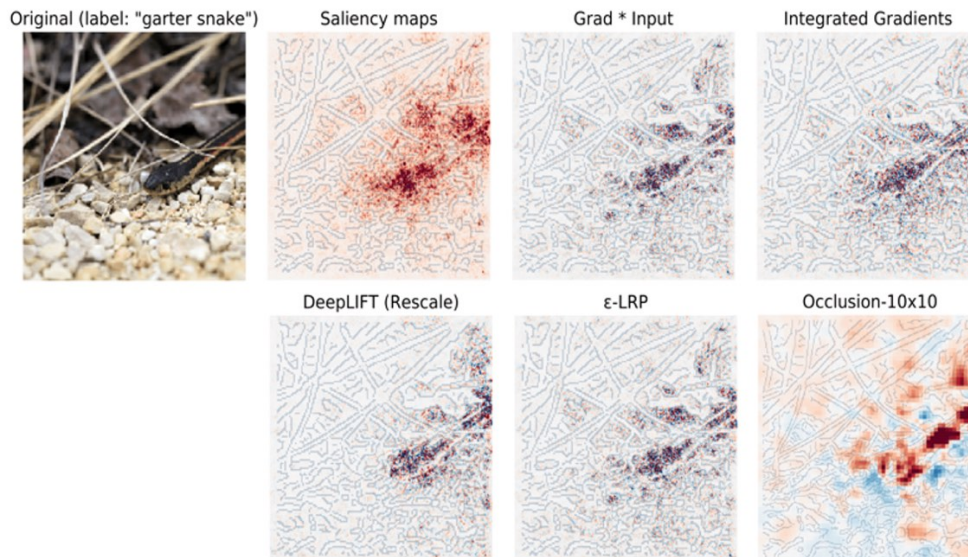


FIGURE 1.21 – Visualisation d'image par les techniques de l'explicabilité [ACÖG18].

Par ailleurs, plusieurs bibliothèques sont développées pour fournir les implémentations des techniques sur l'explicabilité des réseaux de neurones. Certaines parmi elles supportées sur le Framework TensorFlow sont énumérées dans la Table 1.1. Celle de tfexplain a été considérée dans nos expérimentations pour débloquer la boîte noire d'une IA agricole.

1.3.6 Métriques d'évaluation

Les métriques d'évaluation permettent de mesurer la qualité de modèle prédictif, juger de la confiance qu'on peut avoir par rapport au résultat fourni par le modèle. Cela donne aussi des éléments objectifs pour comparer plusieurs modèles. Dans le cadre classification supervisée, il existe divers moyens d'évaluer un classificateur. Les plus courantes seront définies ci-après. Ensuite nous présentons vers la fin de la section, la démarche de validation croisée qui utilise ces métriques lors de son évaluation.

Les métriques communes comme la précision, le rappel, le poids du score f1 et la justesse ou accuracy sont les plus communes pour mesurer la performance de la classification.

La **justesse (accuracy)**, calcule le rapport entre le nombre de classes correctement prédites et le nombre d'exemples total (Équation 1.7).

La **précision** calcule les échantillons positifs qui sont correctement prédits par rapport à tous les échantillons prédits dans une classe positive (Équation 1.5).

Librairie	Application	Méthodes disponibles	Site Web
tf-explain	Tensorflow 2.x	Activations Visualization Vanilla Gradients Gradients*Inputs Occlusion Sensitivity Grad CAM SmoothGrad Integrated Gradients	https://github.com/sicara/tf-explain
AIX360	Tensorflow, Pytorch, Scikit-learn	Contrastive Explanations Method LIME SHAP Etc.	https://github.com/Trusted-AI/AIX360
DeepExplain	Tensorflow 1.x	Saliency maps Gradient * Input Integrated Gradients DeepLIFT, in its first variant with Rescale rule (*) ϵ -LRP (*)	https://github.com/marcoancona/DeepExplain
DeepLIFT	Tensorflow 1.x/ Py-torch	Gradients Deconvolutional Networks Guided Backpropagation LRP Integrated Gradients	https://github.com/kundajelab/deeplift
iNNvestigate	TensorFlow 1.x	Saliency DeConvNet LRP Integrated Gradient Deep Taylor DeepLift Etc.	https://github.com/albermax/innvestigate

TABLE 1.1 – Bibliothèques pour l'explicabilité des réseaux de neurones

1.3. Concepts de l'apprentissage profond

Le **rappel** décrit la capacité du classificateur à trouver tous les exemples positifs (Équation 1.6).

$$Precision = \frac{TP}{TP + FP} \quad (1.5)$$

$$Rappel = \frac{TP}{TP + FN} \quad (1.6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.7)$$

$$Accuracy(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n 1(\hat{y}_i = y_i)$$

avec \hat{y}_i est la valeur prédite de la i -ème observation (échantillon) et y_i sa valeur réelle et n le nombre d'observations (d'échantillons).

Les métriques de rappel ou de précision sont des bonnes mesures du succès de la prédiction lorsque les classes ne sont pas équilibrées. La métrique $F\beta$ -score est une moyenne harmonique de la précision et du rappel. Avec un meilleur score à 1 et un pire score à 0, elle est calculée comme suit (Équation 1.8) :

$$F_\beta = (1 + \beta^2) \frac{Precision * Rappel}{\beta^2 Precision + Rappel} \quad (1.8)$$

Lorsque $\beta = 1$, la formule devient :

$$F_1 = 2 \frac{Precision * Rappel}{Precision + Rappel} \quad (1.9)$$

Ces métriques peuvent être "pondérées" pour chaque étiquette lorsqu'il faut tenir compte du déséqui- libre de données dans l'évaluation. Elles calculent pour chaque étiquette la moyenne pondérée par support (le nombre d'échantillons vrais pour chaque étiquette) :

$$Precision_{weighted} = w_i \left(\frac{TP_i}{TP_i + FP_i} \right)$$

$$Rappel_{weighted} = w_i \left(\frac{TP_i}{TP_i + FN_i} \right)$$

$$F1\ Score_{weighted} = w_i \left(\frac{Precision_i * Recall_i}{Precision_i + Recall_i} \right)$$

$$\text{avec } w_i = \frac{k_i}{\sum_i k_i}$$

où k_i indique les étiquettes uniques à compter pour la proportion totale de toutes les étiquettes et TP, TN, FP, FN sont les vrais positifs (true positive), les vrais négatifs (true negative), les faux positifs (false positive) et les faux négatifs (false negative).

Une autre métrique d'évaluation efficace est celle de la **moyenne de précision (AP)** [PVG⁺11]. Cette métrique nous aidera à comparer les performances de différentes architectures en fournissant la valeur de prédiction pour chaque étiquette. La moyenne de précision est un résumé d'une

courbe précision-rappel en faisant varier un seuil de décision et elle est calculée par la formule suivante (Équation 1.10) :

$$AP = \sum_n (R_n - R_{n-1}) P_n, n > 0 \quad (1.10)$$

Où P_n et R_n sont la précision et le rappel au n-ième seuil.

Une autre mesure courante pour comparer les modèles de prédiction est celle de la **courbe ROC (receiver operating characteristic curve)**. C'est un graphique montrant les performances d'un modèle de classification. Elle résume le compromis le taux de vrais positifs (synonyme de rappel) et le taux de faux positifs pour un modèle prédictif à l'aide de différents seuils de probabilité. Elle est créée en traçant le taux de faux positifs sur l'axe des x par rapport au taux de vrais positifs sur l'axe des y . Cette courbe sert également à comparer différents classificateurs. Elle est généralement associée à une mesure AUC (area under the curve) qui calcule l'aire sous la courbe ROC. Plus une courbe a des valeurs élevées, plus l'aire sous la courbe est grande, moins le classificateur fait d'erreurs.

La technique de la **validation croisée** sert à évaluer la performance d'un modèle prédictif. Les résultats d'un modèle prédictif vont dépendre d'un choix aléatoire particulier pour les ensembles d'entraînement et de validation. Il peut arriver qu'un modèle soit trop spécifique aux observations qui ont servi à l'entraîner. Cette situation est dite « sur-apprentissage ». Pour éviter qu'un modèle ne tombe pas dans ce cas, la démarche consiste à entraîner le modèle avec cette technique de validation croisée. Elle va découper le jeu de données en k parties (ou plis) et à tour de rôle, chacune des k -parties est utilisée pour valider le modèle, le reste (les $k-1$ autres parties) est utilisé pour son entraînement (Figure 1.22). La mesure de performance est la moyenne des valeurs calculées obtenues à chaque tour. Cette technique est très coûteuse en calcul, mais ne gaspille pas trop de données par rapport à une fixation aléatoire de la taille des échantillons.

Dans le cas d'un problème de classification, nous créons généralement les k parties, de sorte qu'elles contiennent à peu près les mêmes proportions d'observation pour chaque classe. Il s'agit d'un type de validation croisée dite stratifiée dont le but est d'éviter un déséquilibre de données important dans la distribution des données.

1.3.7 Framework et ensemble de données

De nos jours, plusieurs frameworks, bibliothèques et jeux ont été développés pour accélérer les recherches et les cas d'applications en apprentissage profond. Grâce à ces outils, le processus d'entraînement est devenu facile.

La Table 1.2 énumère les frameworks et les bibliothèques les plus connus en apprentissage profond. Le Framework TensorFlow est considéré comme plus efficace et plus facile à utiliser à travers nos expériences ou sur la base de notation de GitHub. Il bénéficie d'une large communauté de recherche, de développeurs et sert de base à beaucoup d'autres bibliothèques complémentaires. Il a la capacité de fonctionner sur plusieurs plateformes et s'utilise avec divers langages de programmation comme Python, C++ ou JavaScript. Il a permis le déploiement des modèles d'apprentissage sur les objets embarqués en le convertissant en une version plus légère.

Par ailleurs, des ensembles de données de références sont exploités pour différentes tâches d'apprentissages (classification ou régression) supervisées ou non. Ils contiennent des données

1.3. Concepts de l'apprentissage profond

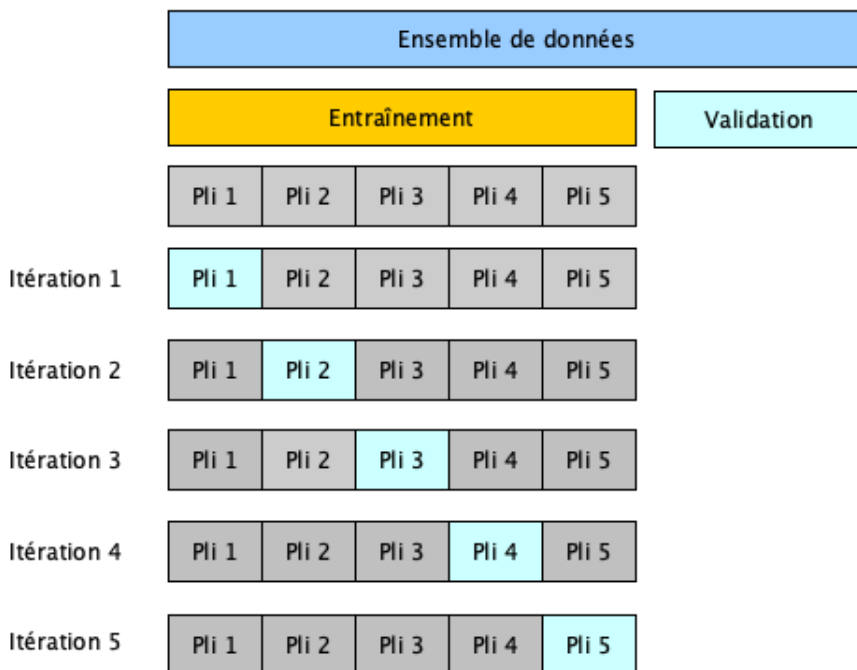


FIGURE 1.22 – Validation croisée à 5 plis

Framework	Licence	Langage de base	Site Web
TensorFlow	Apache 2.0	C++, Python	https://www.tensorflow.org/
Keras	MIT	Python	https://keras.io
Caffe	BSD	C++	https://caffe.berkeleyvision.org
Torch	BSD	C, Lua	http://torch.ch
DL4J	Apache 2.0	Java	https://deeplearning4j.org
MXNet	Apache 2.0	C++	https://mxnet.apache.org/
MatConvNet	Oxford	MATLAB	https://www.vlfeat.org/matconvnet

TABLE 1.2 – Liste des frameworks les plus courants

dans plusieurs formats (texte, images, signal, etc.). Quelques-uns d'entre eux (représentés par la Table 1.3) sont exploités dans les expérimentations lors de l'entraînement des réseaux de neurones.

Ensemble de données	Nombre de classes	Application	Site Web
MNIST	10	Classification des chiffres	http://yann.lecun.com/exdb/mnist
ImageNet	1000	Classification d'images, localisation, détection d'objets, etc.	https://www.image-net.org
CIFAR10/100	10/100	Classification d'images	https://www.cs.toronto.edu/~kriz/cifar.html
PascalVOC	20	Classification d'images, localisation, détection, segmentation d'objets.	http://host.robots.ox.ac.uk/pascal/VOC
NUS Wide	81	Classification d'images, localisation, détection d'objets, etc.	https://lms.comp.nus.edu.sg/wp-content/uploads/2019/research/nuswide/NUS-WIDE.html
Youtube-8M	3862	Classification de vidéo	https://research.google.com/youtube8m
Microsoft COCO	80	Détection d'objets, segmentation sémantique	https://cocodataset.org
Amazon rainforest	17	Classification d'images multi-labels	https://www.kaggle.com/c/planet-understanding-the-amazon-from-space
IP102	102	Classification d'images	https://github.com/xpwu95/IP102

TABLE 1.3 – Jeux de données de référence

1.4 Agriculture de précision

Apparue, il y a des milliers d'années, l'agriculture consiste à se servir d'un milieu comme support de production alimentaire. Dans son développement, les exploitants ajustaient les parcelles sur la base d'idées intuitives : utilisation des outils artisanaux, passage des engrais, et

1.4. Agriculture de précision

protection des cultures. Cela était possible parce que les parcelles étaient de petites tailles et il y avait des manques de moyens techniques et économiques. La donne a changé par l'agrandissement des exploitations, partagées en plusieurs unités d'hectares. Dans ces conditions, l'approche conventionnelle de l'agriculture manque de lucidité notamment avec les limites de sous-utilisation des intrants et le risque environnemental.

Les opportunités offertes par les nouvelles technologies du numérique ont donc fortement fait progresser l'agriculture vers une industrialisation. Cette mécanisation a fait passer l'agriculture à une nouvelle ère, l'Agriculture 4.0 où l'agriculture intensive s'impose désormais en contribuant à l'efficacité des procédés de production agricole. Grâce aux progrès fournis par les technologies, l'agriculture a subi d'énormes changements visant à la rendre plus intelligente et plus précise. Les technologies de positionnement (comme le global positioning system, ou GPS) et l'informatique embarquée permettent d'obtenir des informations géographiques et de réaliser des interventions modulaires sur les exploitations. Cette réflexion est à l'origine du concept agriculture de précision, concept qui se traduit par « la bonne intervention au bon endroit et au bon moment [Zim08, ZLB97] ».

En pratique, l'agriculture de précision est fondée sur l'intégration des technologies et de la communication dans la gestion des parcelles (Figures 1.23 et 1.25). Elle vise à moduler les pratiques culturales en fonction de la variabilité (sol, couverture végétale, etc.) intraparcellaire afin de piloter correctement le processus de production agricole et optimiser avec précision les interventions culturales.

À l'aube d'une croissance de la population mondiale, le défi actuel à relever est d'améliorer la qualité et la quantité des produits agroalimentaires tout en respectant la santé de l'homme et de son environnement. L'agriculture de précision répond donc aux besoins des agriculteurs comme étant un système agricole intégré basé sur l'information et la production. Son objectif est d'accroître l'efficacité, la productivité et la rentabilité de la production à long terme, sur un site spécifique et sur l'ensemble de l'exploitation, tout en minimisant les impacts involontaires sur la faune et l'environnement [TN07]. Selon [Gan05], l'agriculture de précision se définit comme un ensemble de technologies qui ont contribué à propulser l'agriculture dans le monde informatisé de l'information et qui est conçu pour aider les agriculteurs à obtenir un plus grand contrôle sur la gestion des opérations agricoles. En effet, l'agriculture de précision est une mise à niveau des stratégies conventionnelles de prise de décision spécifique concernant la production des cultures.

Nous assistons dans ses dernières décennies à l'existence d'une volonté des entreprises agro-industrielles et des coopératives agricoles, d'explorer les potentialités du numérique pour contribuer au développement des modèles d'amélioration de la santé des exploitations agricoles. Il s'agit des technologies des grandes données, de l'Internet des objets, de l'informatique en nuage (cloud) et de l'intelligence artificielle. Pour illustrer cette floraison d'outils pour l'agriculture de précision, nous pouvons citer les exemples suivants à titre illustratif :

AIRINOV offre une application avec tous les services agronomiques dont le but est de guider l'agriculteur dans ses premiers pas vers la modulation de la dose d'azote avec la géolocalisation des tracteurs dans les parcelles.

Le projet international dirigé par l'Université de Twente aux Pays-Bas et financé par l'Union européenne « **Spurring a Transformation for Agriculture through Remote Sense (STARS)** » avait pour vocation de favoriser la transformation de l'agriculture par la télédétection grâce aux images satellitaires à haute définition et des drones. Cela est rendu possible par l'acquisition des images multi-spectrales des parcelles à surveiller à l'aide d'objets connectés.

La société **Geosys** offre des solutions s'appuyant sur les résultats de recherches en Agronomie, des Systèmes d'Information et du Web, et de données géographiques (imagerie satellite, données climatiques. . .). Celles-ci fournissent des systèmes des systèmes d'aide à la décision pour les professionnels du monde agricole permettant par exemple de corriger les problèmes liés à la variabilité intraparcellaire.

Farmstar a conçu un outil d'agriculture de précision et accompagne des milliers d'agriculteurs soucieux de combiner rendement, qualité de récolte et respect environnemental. Cet outil est basé sur expertise agronomique associée à la télédétection par satellite permettant par exemple de fournir des services (indicateur de biomasse, chlorophylle, etc.) en production d'informations agricoles.

SkyWatch propose une nouvelle technologie satellitaire pour collecter des images panchromatiques haute résolution et multi-spectrales pour diverses applications dans le domaine agricole (par exemple le suivi de la santé des cultures) et de la surveillance environnementale.

Agrimeteo - NewFarm Agriconsult est un réseau collaboratif d'experts d'agronomes, de scientifiques, d'informaticiens et de fabricants d'objets connectés fournissant des solutions aux agriculteurs afin d'optimiser leurs décisions agronomiques.

L'agriculture de précision comporte plusieurs enjeux de nature agronomique, technologique et économique afin de répondre précisément aux préoccupations des agro-industries. L'enjeu technologique se caractérise par l'usage des outils embarqués (capteurs de biomasse ou de chlorophylle), des systèmes aéroportés (drones ou satellites), et des systèmes de cartographie. L'enjeu agronomique vise à améliorer le ratio intrant/rendement, la sélection des variétés de cultures adaptées aux contextes phytosanitaires. L'enjeu environnemental se caractérise par la limitation de l'érosion du sol et les pertes d'azote nitrique par lessivage. Il s'agit ainsi de favoriser la fertilisation optimale tout en préservant la santé humaine et l'environnement. L'enjeu économique contribue à garantir la rentabilité des activités de l'exploitation en diminuant le coût de la production, avec la limitation de l'utilisation excessive des intrants tout en garantissant la qualité et le rendement des cultures.

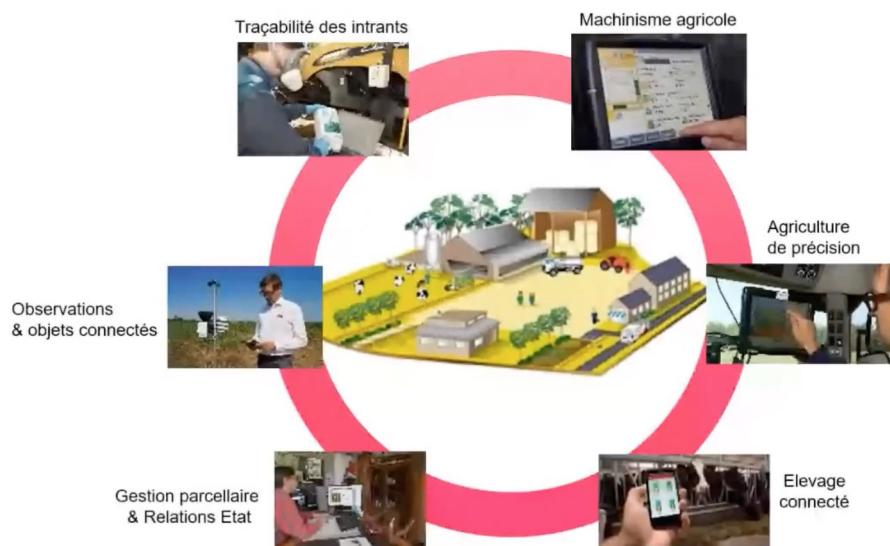


FIGURE 1.23 – Usage de la donnée sur les exploitations agricoles (Source : JEMS Group)

L'agriculture de précision génère de nombreuses informations issues des données collectées

1.4. Agriculture de précision

sur les parcelles par voie terrestre, par les objets connectés ou par voie aérienne par les outils satellitaires (Figure 1.23). Il en découle donc le réel besoin de développer des outils et des méthodes pour stocker, analyser, interpréter, visualiser et diffuser les données collectées afin d'en tirer des valeurs ajoutées (Figure 1.25). Ces applications agricoles représentent les moyens de production de connaissances utiles au service des systèmes d'aide à la décision [WGV17] pouvant accompagner les exploitants agricoles et les décideurs publics ou privés.

À l'ère moderne de la révolution industrielle, l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond sont devenus très populaires dans les recherches scientifiques avec de nombreuses applications en traitement de langue naturelle, classification d'images, diagnostic de maladies, fouille de texte, etc. Aujourd'hui, l'apprentissage profond connaît de nombreux cas d'exemples appliqués à la gestion des cultures et des sols. Pour ces raisons, l'apprentissage profond apparaît comme une réponse pertinente qui dépasse les modèles agronomiques classiques pour mieux aider les agricultures dans le positionnement dans la chaîne de valeur agricole (Figure 1.24).

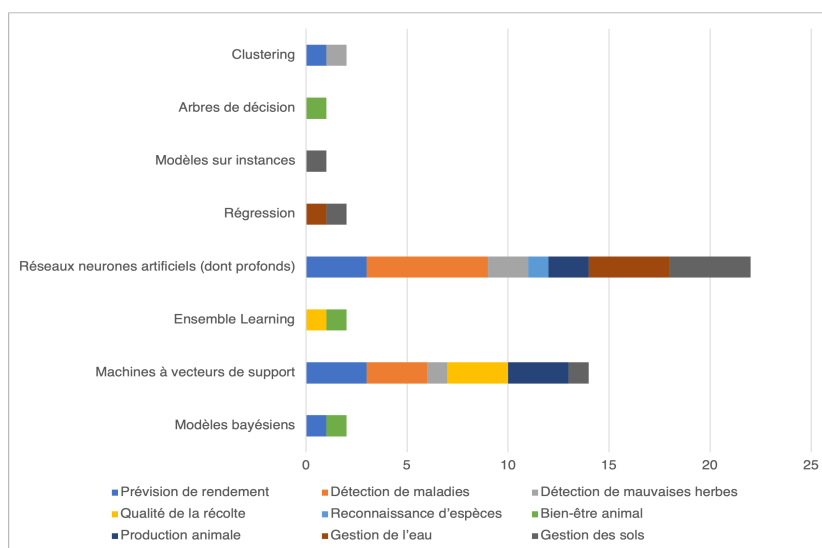


FIGURE 1.24 – Apprentissage automatique pour des usages agricoles d'après l'étude couvrant une période de 2004 à 2018 [LBM⁺18]

La transformation numérique de l'agriculture soutenue par l'apprentissage profond constitue une nouvelle carte de progrès scientifiques et techniques. Elle commence à changer la vision des agriculteurs en apportant de nombreux avantages pour gérer les changements fréquents des conditions extérieures. Les solutions développées ont permis de compléter les technologies (agronomique et informatique) déjà mises en œuvre pour accroître la production et moderniser certaines étapes de l'agriculture de précision.

L'apprentissage profond a apporté une contribution importante grâce à des dispositifs automatiques et intelligents et à des modèles de prédiction qui utilisent des algorithmes d'apprentissage pour améliorer diverses pratiques agricoles (Figure 1.26). Ces transformations et usage des concepts de l'apprentissage machine ont conduit à la mise en place de plusieurs applications ou méthodes telles que la gestion de l'eau et du sol, la détection des maladies des cultures, les mauvaises herbes, le comptage des fruits et la prédiction du rendement.

Les potentialités de l'apprentissage profond pour la surveillance des cultures ont été observées dans plusieurs études (comme montré dans le chapitre 2). Les chapitres 4 et 5 men-

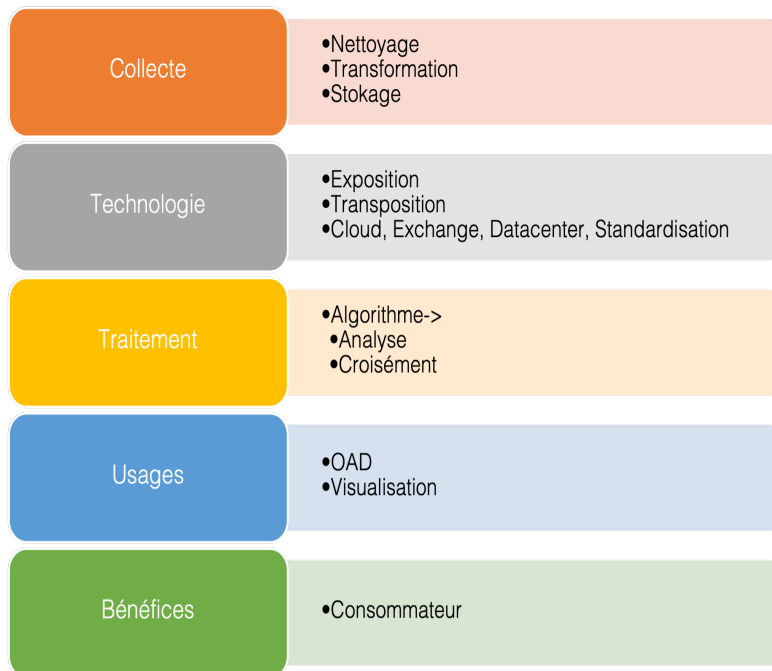


FIGURE 1.25 – Traitement des données agricoles

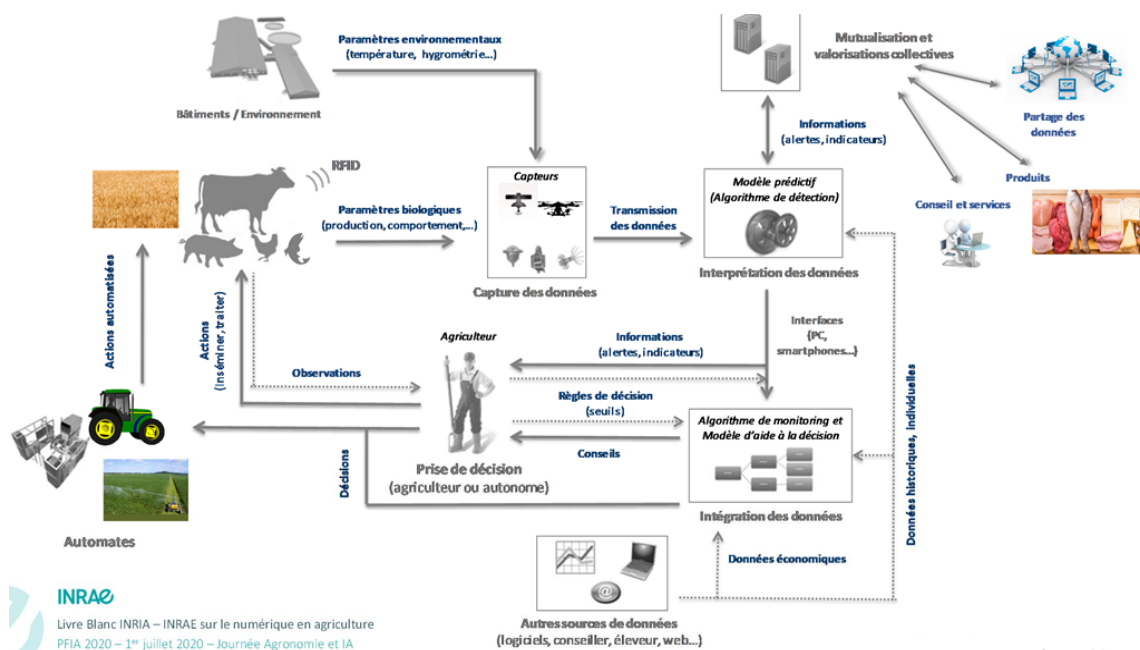


FIGURE 1.26 – Usages du numérique dans l'Agriculture

tionnent des cas d'exemple de détection des maladies ou des insectes nuisibles. À cela s'ajoute, la sélection des cultures, le rendement, des détections des mauvaises herbes qui jouent un rôle clé dans la détermination du rendement. Toutes ces nouvelles pratiques nécessitent une assistance technique des agriculteurs pour une agriculture durable. Ainsi l'apprentissage profond est devenu une force motrice de l'agriculture pour le changement.

En résumé, les applications de l'intelligence artificielle ou de l'apprentissage profond offrent de

1.5. Apprentissage profond et limitation des données

nombreuses possibilités (opportunités) pour améliorer les diverses étapes de l'agriculture afin de faire face aux différentes difficultés et atteindre les objectifs de l'agriculture de précision [Gup19, ZLL⁺18] :

1. Le traitement de l'information agricole : Le suivi de l'état des plantes et des animaux est vital pour la production agricole.
2. Les modèles font de la détection des maladies un processus réalisable, ce qui améliore le potentiel de production de cultures saines.
3. Le contrôle optimal des systèmes de production agricole : Les stratégies de contrôle des systèmes de production agricole reposent souvent sur l'expérience de l'agriculteur ou sur les connaissances d'un expert, ce qui ne tient pas compte de l'état physiologique des plantes (animaux) ou de la demande en temps réel.
4. Les pratiques de gestion des cultures ont atteint de nouveaux sommets, et il est devenu assez pratique pour les agriculteurs de gérer les cultures avec un minimum d'efforts.
5. Les équipements intelligents des machines agricoles : la production agricole implique de nombreux types de tâches.
6. Les données agricoles sont analysées efficacement par les modèles de l'apprentissage profond afin de faire des prédictions précises.
7. La gestion du système économique agricole : Le rendement agricole en lui-même n'est pas suffisant pour l'agriculture. De nombreux autres facteurs doivent être pris en compte, tels que les prix et la qualité des produits agricoles. Il est très important de prévoir les prix des produits agricoles.
8. L'IA a considérablement révolutionné le système de prévision météorologique dans l'agriculture, ce qui joue un rôle essentiel dans l'agriculture.

Le secteur agricole a été remodelé par les techniques de l'apprentissage profond. Plusieurs pratiques de l'agriculture ont été impactées par l'évolution des technologies. Cependant, l'humain doit rester la clé de ce changement pour contrôler les risques émis par les diverses intégrations technologiques et l'interprétation des informations.

1.5 Apprentissage profond et limitation des données

Nous avons présenté dans cette section, l'apprentissage par transfert entre un domaine source et un domaine cible, et l'augmentation des données qui sont des techniques pouvant améliorer l'entraînement des réseaux de neurones convolutifs.

1.5.1 Augmentation des données

L'augmentation des données est une approche qui enrichit et génère davantage de nouvelles données pour les données existantes. Elle consiste à créer (par rotation, retournement, variation de couleur, bruit, ...) des répliques transformées de chaque instance dans un ensemble de données d'apprentissage (Figure 1.27)². Elle est nécessaire pour augmenter les données disponibles et accroître la taille des données fournies aux classificateurs afin de compenser le coût de

2. <https://medium.com/analytics-vidhya/data-augmentation-in-deep-learning-3d7a539f7a28>

la collecte de données supplémentaires. Elle est appliquée dans le processus d'entraînement du modèle d'apprentissage profond par le biais de transformations aléatoires telles que le retournement, le zoom, le décalage et la remise à l'échelle. Pour une valeur numérique x donnée, les transformations suivantes sont effectuées :

- zoom : le redimensionnement de chaque image sur l'intervalle $[1 - x, 1 + x]$, x est une valeur flottante supérieure à 0.
- rotation : chaque image est tournée sur l'intervalle $[0, x]$.
- flip : l'image originale est transformée par une inversion miroir sur un axe horizontal.
- rescale : la multiplication des données par une valeur spécifique fournie après l'application des autres transformations.



FIGURE 1.27 – Augmentation de données

1.5.2 Apprentissage par transfert

L'apprentissage profond requiert une grande quantité de données d'apprentissage pour entraîner les réseaux de neurones. Ces données sont exploitées par des réseaux d'architectures complexes qui disposent d'un nombre important de paramètres d'apprentissage. Entraîner ces réseaux à partir de zéro est très coûteux en termes de ressources matérielles (vitesse de calculs et stockage). En même temps, il y a un manque d'ensembles de données de référence pour certains domaines comme l'agriculture. Les chercheurs développent leur propre collecte de données. Des projets comme PlantVillage contribuent ainsi à réduire la difficulté. En conséquence, il est difficile de construire des données d'entraînement étiquetées suffisantes et correctes pour tous les problèmes de l'agriculture. Ces faits limitent la croissance du marché des applications agricoles [CKKT19]. Un moyen permettant de réduire l'effort à ce stade consiste à utiliser un apprentissage par transfert [PY10], qui offre une solution efficace pour une prédiction précise avec moins d'échantillons à utiliser lors de l'entraînement du réseau.

L'apprentissage par transfert est la capacité d'un système à reconnaître et à appliquer les connaissances et les compétences acquises dans des tâches précédentes à de nouvelles tâches (dans

de nouveaux domaines). Autrement dit, il est possible de transférer des connaissances sur les données d'un domaine source au domaine cible (Figure 1.28).

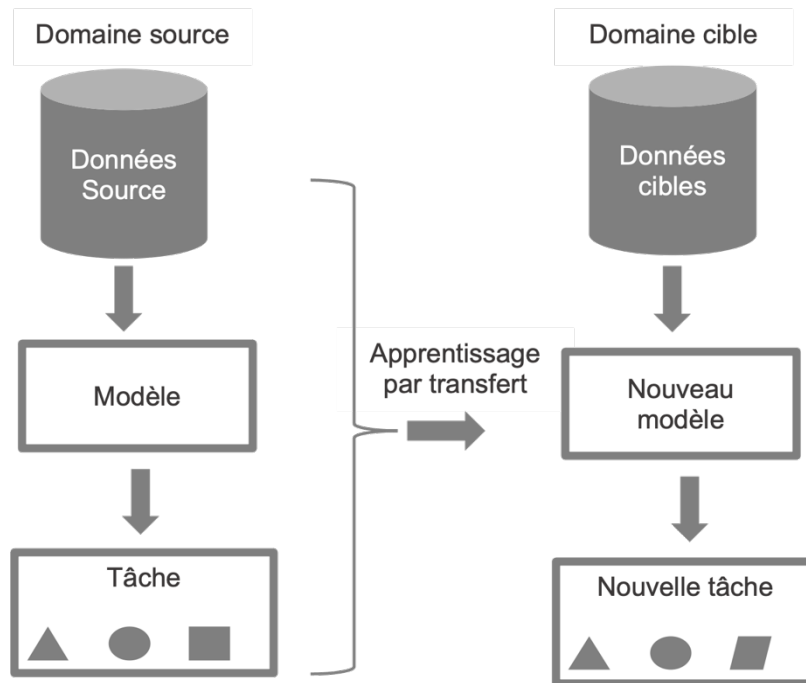


FIGURE 1.28 – Principe de l'apprentissage par transfert

Pour formaliser l'apprentissage par transfert en suivant [PY10], un domaine $D = \{X, P^X\}$ se compose d'un espace de caractéristiques X et une distribution de probabilité marginal P^X , où chaque observation $x \in X$. Une tâche T est définie sur un espace d'étiquettes Y , et d'une fonction prédictive $f(\cdot)$ des observations non vues, telles que $T = \{Y, f(\cdot); f(x) = P(y|x)\}$. Soit $D_s = \{x_{s_i}, y_{s_i}; x_{s_i} \in X_s \text{ et } y_{s_i} \in Y_s \text{ pour } i = 1..n_s\}$, les données étiquetées d'un domaine source et de $D_t = \{x_{t_i}, y_{t_i}; x_{t_i} \in X_t \text{ et } y_{t_i} \in Y_t \text{ pour } i = 1..n_t\}$, données étiquetées d'un domaine cible. L'apprentissage par transfert consiste à améliorer l'apprentissage de la fonction prédictive $f_t(\cdot)$ pour le domaine cible en exploitant les connaissances D_s et D_t où $D_s \neq D_t$, $T_s \neq T_t$.

L'approche de l'apprentissage par transfert est particulièrement utile dans le contexte des réseaux de neurones convolutifs profonds pour la classification et la réduction du bruit à l'aide de plusieurs architectures profondes entraînées [AAMB18]. Avec l'apprentissage par transfert, les réseaux de neurones profonds peuvent apprendre des relations très compliquées (complexes) qui peuvent conduire à un sur-apprentissage (overfitting) par faute de données. Le sur-apprentissage concerne les réseaux de neurones ayant plus de paramètres d'apprentissage que nécessaire par rapport à la taille de données. Cependant, nous pouvons nous baser sur certaines méthodes (techniques) telles que la régularisation et le dropout pour prévenir ce problème [SHK⁺14].

Le concept de l'apprentissage par transfert est couramment utilisé en apprentissage profond lors de la classification d'images. Dans de nombreuses applications, il se base sur la similarité entre les espaces de caractéristiques ou des étiquettes apprises d'un domaine source et cible avec principalement deux options possibles :

La première option utilise un réseau déjà entraîné que nous pouvons réutiliser ou appliquer sur un domaine source. Dans ce cas, nous retiendrons deux cas d'usage : **le réglage ou l'ajus-**

			Jeu de données si-		Jeu de données	
			milaires		différentes	
Faible	jeu	de	feature	extractor	feature	extactor
données			(première couches)		(dernières couches)	
Large	jeu	de	fine-tuning		A partir de zéro	
données						

TABLE 1.4 – Appliquer l'apprentissage par transfert

tement fin (fine-tuning) et l'extraction de caractéristiques (feature extraction).

Appliquer un réglage fin ou l'extraction de caractéristiques dépend nécessairement de la taille de l'ensemble de données cible et de sa similarité avec l'ensemble de données sources (voir la Table 1.4)³. L'opération de transfert utilise un réseau pré-entraîné sur un ensemble de données source en procédant par :

- i. Une extraction de caractéristiques (feature extraction) : elle consiste à utiliser les caractéristiques d'un réseau pré-entraîné pour représenter les données (images) de nouveaux ensembles de données. Ces caractéristiques sont utilisées pour entraîner un nouveau classificateur. L'extraction de caractéristiques a du sens si nous avons un grand nombre de variables impliquées dans des données complexes (dans l'apprentissage).
- ii. Réglage de fin (fine-tuning) : l'ajustement fin du réseau pré-entraîné est pertinent lorsque le jeu de données cible est très grand. Il consiste à dégeler (unfreezing) quelques couches supérieures d'un modèle utilisé pour l'extraction de caractéristiques ou non. De plus, les paramètres de toutes les couches (sauf la dernière) sont initialisés comme ceux du réseau pré-entraîné.

L'architecture de base d'un CNN est résumée par la Figure 3.1. Techniquement, en cas d'approche d'apprentissage par transfert, la première partie (**module de convolution**), les couches de convolution dites couches basses sont partagées et utilisées pour extraire les caractéristiques des données d'entrée. Dans la seconde partie (**module de classification**) utilisée pour l'ajustement fin, les couches entièrement connectées ou couches hautes sont systématiquement remplacées par une ou plusieurs nouvelles couches pour la classification des données associées au problème cible.

La deuxième option est basée sur l'adaptation du domaine où deux approches sont envisageables : le **transport optimal** [CFHR17, CFT14, RCFT18] et l'**information mutuelle** [ARS18] pour guider la sélection de la meilleure source de données. Il s'agit de minimiser l'erreur ou la divergence entre les deux domaines, conduisant ainsi à un transfert dit négatif.

Le problème de l'adaptation de domaine a refait surface avec les travaux de [CFT14] qui évaluent la distance entre domaines. Les techniques proposées par le transport optimal et l'information mutuelle permettent de réduire cette distance entre les domaines source et cible. La théorie du transport optimal a été formulée par le géomètre français Gaspard Monge au 18^e siècle. Il revient dans le travail de Kantorovitch [Kan58] et revisité dans [Vil09]. La théorie du transport optimal définit une Distance de Wasserstein (ou distance de transport optimal) pour

3. <http://cs231n.github.io/transfer-learning>

mesurer la distance entre des distributions de probabilité sur un espace métrique spécifié. Le transport optimal s'est avéré utile pour l'adaptation au domaine dans le cadre de l'apprentissage non supervisé [CFHR17, CFT14] qui consiste à trouver de bonne source de données. Dans ce sens, [BEF19] ont pu réaliser par une approche modulaire composée de trois étapes : (i) le calcul des distances entre ensembles de données par paires (ii) la transformation des distances en scores, et (iii) la garantie de la variété des ensembles d'entraînement.

Il semble intuitivement raisonnable que des procédures appropriées permettent de sélectionner la source la plus proche à partir des caractéristiques extraites par le réseau de neurones pour un apprentissage par transfert optimal (transférabilité maximale). Ces procédures supposent que les caractéristiques extraites d'un domaine source doivent donner des informations supplémentaires et pertinentes qui ne sont pas précédemment prises en compte par les données d'entraînement du domaine cible [ARS18]. La transférabilité maximale peut être calculée en utilisant des sources ouvertes sélectionnées sur les critères basés sur la théorie de l'information mutuelle. Dans les deux cas de figure, le réseau de neurones est testé sur une ou plusieurs sources de données afin de sélectionner celle qui est la plus efficace.

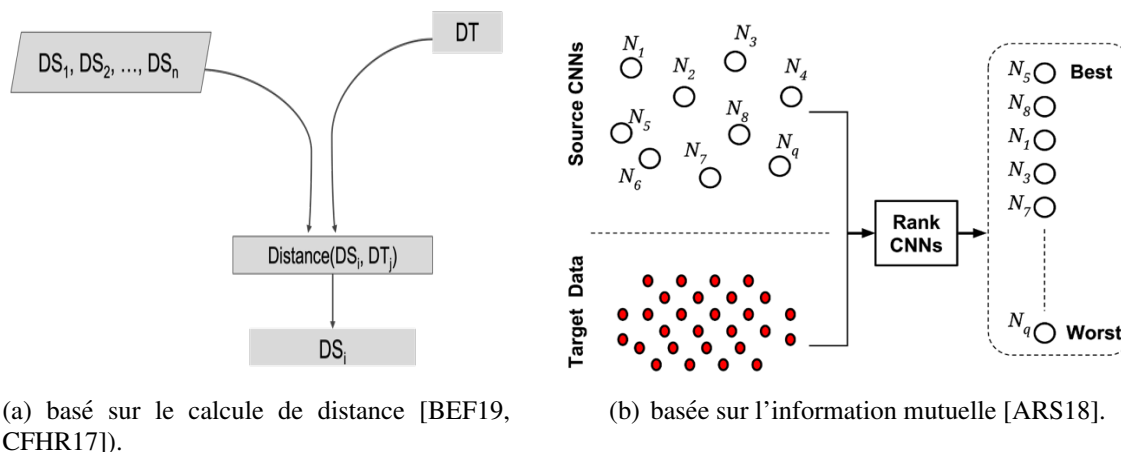


FIGURE 1.29 – Sélections des domaines sources.

1.6 Les bases de données utilisées

Dans cette section, nous allons présenter les ensembles de données mis en place pour atteindre les objectifs de cette thèse. L'apprentissage profond nécessite une grande quantité de données pendant entraînement d'un modèle prédictif. Pour conduire nos travaux, nous utilisons des données de référence publique disponibles ou téléchargées sur Internet. Chaque ensemble de données est divisé en trois : l'ensemble de données d'entraînement, de validation et de test. Les ensembles de données PascalVOC, NusWide ou Amazon Forest sont des données à caractère multi-labels. Ils sont exploités dans le cadre expérimental du chapitre 3 pour une classification multi-labels. L'ensemble de données mildiou est utilisé dans le chapitre 4 pour la détection de la maladie du mildiou sur le mil tandis que l'ensemble de données IP102 pour la détection et la localisation des insectes ravageurs est employé dans le chapitre 5.

1.6.1 Maladie du mildiou du mil

Les images du mil ont été manuellement téléchargées sur Internet et recadrées pour construire le jeu de données d'entraînement. Au total, nous avons 126 images dans notre jeu de données. Il est composé de différentes formes de maladies de mildiou et d'images saines (Figure 1.30). Sur l'ensemble de données d'entraînement, 70 images portent sur les maladies et 29 sont saines. L'ensemble de données tests comprend 18 images malades et 9 sans maladie.

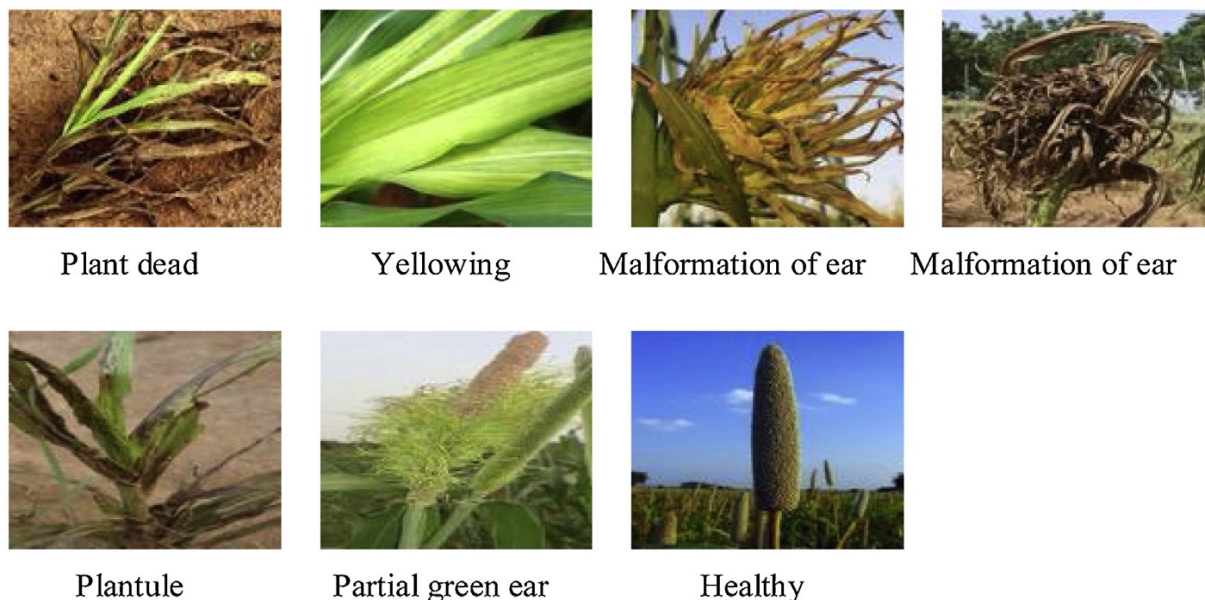


FIGURE 1.30 – Exemple d'images de la maladie du mildiou

1.6.2 IP102

Cet ensemble de données de référence IP102 est utilisé dans les expérimentations du chapitre 5. C'est un ensemble de données à grande échelle pour la reconnaissance des insectes ravageurs [WZL⁺19]. L'ensemble de données IP102 contient plus de 75 000 images pour 102 classes ou catégories de ravageurs avec 737 images en moyenne par classe. L'ensemble des données est subdivisé comme suit : 45 095 images dans l'ensemble d'entraînement (train set), 7 508 images dans l'ensemble de validation (validation set) et 22 619 images dans l'ensemble de test (test set). La Table 1.5 fournit un extrait de l'ensemble de données, à savoir le nombre d'images, l'identifiant des classes à prédire et le nom de la super-classe.

Le jeu de données IP102 est considéré comme l'un des plus grands jeux de données sur les insectes nuisibles, lancé en 2019. Il comporte plusieurs défis pour évaluer les performances de classification : (i) la couleur de fond de l'objet est similaire à la couleur de l'objet ; (ii) les images couvrent tout le cycle du ravageur ; (iii) certains ravageurs se ressemblent. La Figure 1.31 montre quelques exemples d'images.

1.6.3 PASCAL VOC

Le Pascal VOC [EVGW⁺07, EVGW⁺12] est un jeu de données très populaire pour construire et évaluer des algorithmes concernant la classification d'images, la détection d'objets et la seg-

Identifiant de la classe	Super classe	Nombre d'images
14	Rice	8417
9	Wheat	2491
10	Mango	4482
13	Corn	14015

TABLE 1.5 – Quantité d'échantillons de types de parasites provenant de l'ensemble de données IP102



FIGURE 1.31 – Exemples d'images tirées de l'ensemble de données IP102

mentation. Le jeu de données d'images contient 20 étiquettes ou classes d'objets (telles que personne, chien, avion, bouteille, etc.). Ce jeu de données contient 9 963 images réparties entre les jeux de données d'entraînement/validation (trainval) et de test (Figure 1.32). Les jeux de données d'entraînement et de test comprennent respectivement 5 011 et 4 952 échantillons. Le Pascal VOC propose un autre référentiel, par exemple le Pascal VOC 2012, largement utilisé dans les expérimentations d'apprentissage profond. Le jeu de données VOC 2012 contient 22 531 images pour 20 catégories. Le jeu de données est divisé en sous-ensemble d'entraînement, de validation et de test avec 11 540 images dans l'entraînement et la validation.

1.6.4 AMAZON RAINFOREST

L'ensemble de données Amazon rainforest [Pla17] contient des photos satellites de la forêt tropicale amazonienne. Ce jeu de données a servi dans le concours de sciences de données sur le site de Kaggle. Ce jeu de données est composé de 40 479 images pour les jeux de données d'entraînement et de validation et de 40 669 images pour le jeu d'évaluation. Amazon rainforest est un repère standard de vision par ordinateur qui comprend des échantillons à étiquettes multiples

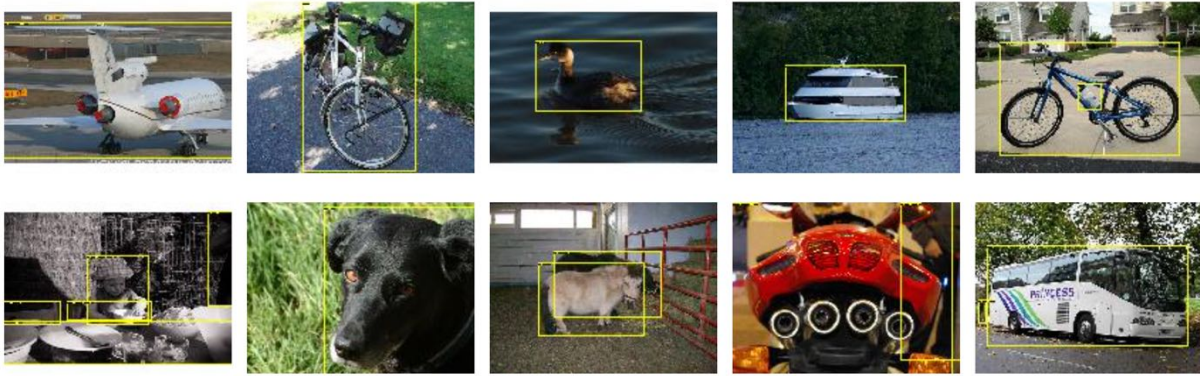


FIGURE 1.32 – Exemples d'images issues PascalVOC

étiquetés à travers dix-sept (17) classes d'étiquettes. On peut citer par exemple, l'agriculture, la rivière, les routes et ainsi de suite (Figure 1.33).



FIGURE 1.33 – Exemples d'images issues d'Amazon rainforest

1.6.5 NUS-WIDE

Le jeu de données NusWide [CTH⁺09] contient 269 648 images et 5 018 tags provenant de Flickr. Ces images sont annotées manuellement avec 81 concepts ou classes. Les images sont très diverses et comprennent des concepts tels que des événements (e.g la natation), des scènes/lieux (e.g des aéroports), des personnes (e.g la police), des objets (e.g des animaux) (Figure 1.34). En plus, les images sont très différentes les unes des autres. Dans le cadre expérimental, nous considérons ce jeu de données contenant 5000 images pour l'entraînement et 1000 images pour les tests.

1.7 Conclusion

Ce chapitre a présenté un aperçu de l'apprentissage profond et exposé les concepts fondamentaux des réseaux de neurones artificiels en passant par les réseaux de neurones convolutifs. Nous avons ensuite présenté quelques techniques permettant de remédier au phénomène de sur-apprentissage à savoir l'augmentation des données ou et l'apprentissage par transfert. Comment

1.7. Conclusion

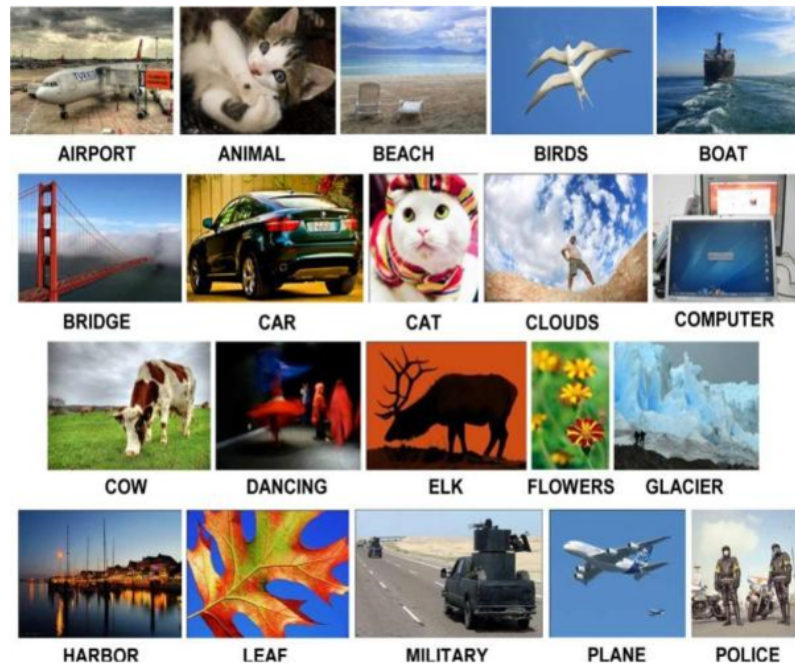


FIGURE 1.34 – Exemples d’images issues NUSWide

faire confiance ou justifier la recommandation d’un réseau de neurones profonds ? Les travaux présentés sur les méthodes de l’explicabilité permettent d’obtenir une explication du processus ayant conduit à cette recommandation.

A partir de la présentation de l’agriculture de précision, nous remarquons que, les modèles émergents de réseaux de neurones convolutifs (CNN) ont intégré dans la chaîne de valeur agricole. Nous avons trouvé un intérêt d’étudier ces diverses applications pour la surveillance de la santé des cultures. De ce fait, le prochain chapitre présente notre première contribution qui définit la place occupée par l’apprentissage profond dans le domaine de l’agriculture. Il est recensé à travers cette étude, un ensemble de travaux de recherches sur les deux thématiques afin de dégager les apports et manques qui déboucheront sur nos perspectives de recherches.

Chapitre 2

Apprentissage profond à la rencontre de l'agriculture de précision : une analyse bibliographique

2.1 Introduction

L'agriculture est une pratique qui consiste à cultiver des produits alimentaires. Cette pratique nécessite une assistance technique pour éviter la surproduction et l'insuffisance alimentaire. De nos jours, le secteur agricole est en pleine transformation numérique afin d'augmenter sa productivité. Les défis liés à son développement sont nombreux (comme la détection des maladies, la prédiction du rendement des cultures ou la pulvérisation de pointe).

L'introduction de l'intelligence artificielle et de l'apprentissage automatique a eu un fort impact dans ce domaine en le faisant passer à une nouvelle ère digitale. Il s'agit de l'agriculture numérique ou de précision dans laquelle on assiste à une forte utilisation de l'informatique intelligente, de la robotique, des drones ou des capteurs. Les opportunités offertes par ces technologies numériques ont favorisé l'industrialisation, la surveillance des processus de production et déclenché l'Agriculture 4.0. Elles consistent à améliorer la vie des agriculteurs en fournissant de précieux conseils d'optimisation des processus d'exploitation agricole s'appuyant sur les données et l'automatisation des tâches.

L'observation visuelle des plantes par les experts était généralement effectuée par le biais d'un diagnostic lorsque cela est nécessaire. Néanmoins, cette méthode peu rentable prend habituellement beaucoup de temps. Pour résoudre ces problèmes, il est nécessaire d'utiliser des techniques avancées et intelligentes. A cet effet, les algorithmes classiques (machine à vecteur de support, forêt aléatoire, k plus proches voisins) de l'apprentissage automatique ont été appliqués dans de nombreuses études. [AHRM14, ASAV13, YGYN14, HRDR21]. Cependant, les récentes performances de l'apprentissage profond, un sous-ensemble de l'apprentissage automatique, ont été efficaces pour plusieurs domaines. Il permet à la machine d'effectuer des traitements complexes sur des données volumineuses, en fournissant des prédictions avec des résultats prometteurs. En conséquence, le domaine agricole s'est orienté vers des solutions basées sur l'apprentissage profond. Dans ce sens, des entreprises agro-industrielles et des coopératives agricoles explorent ces potentialités pour contribuer au développement des modèles de surveillance des exploitations agricoles. Les progrès accomplis jusqu'à présent sont encourageants. De nombreux aspects ont été traités notamment la reconnaissance de mauvaises herbes [LBMS18, PKA19, RvHB⁺20],

l'identification des maladies des plantes [CKKT19, KHC20, GZY⁺20], la détection des insectes [DLJ⁺19, CHL20b, ASB20], la récolte des fruits [MC17, BBSF19, GXF]. Toutefois, cette agriculture de précision soulève aussi de nombreux problèmes et défis relatifs à la transition écologique du système agricole, tels que la réduction de la consommation d'énergie, la minimisation des intrants et la place de l'humain dans la réalisation des systèmes intelligents. Ainsi, dans ce chapitre, nous présentons les résultats d'une étude bibliométrique de 440 récentes recherches liées à l'implication de l'apprentissage profond en agriculture. Nous évaluons les contributions des chercheurs de ces sous-domaines de l'intelligence artificielle qui sont principalement impliqués dans le développement d'une agriculture intelligente. Ce chapitre est organisé de la manière suivante : la section 2.2 présente la technique ainsi que les logiciels bibliométriques, la section 2.3 décrit les analyses bibliographiques effectuées sur les données collectées, tandis que la section 2.4 concerne la discussion avec des observations sur les défis identifiés, et enfin la section 2.5 souligne les informations conclusives ainsi que les perspectives.

2.2 Techniques Bibliométriques

L'analyse bibliographique ou la bibliométrie est l'analyse quantitative de la production scientifique et l'analyse des réseaux de cette production [KIP⁺08, DB09, CX16, AC17]. A partir d'une requête liée à un sujet ou un domaine donné, elle permet d'identifier les recherches, les auteurs ainsi que leurs relations par rapport aux différentes citations et publications. Elle comporte principalement les étapes : (a) la collecte de travaux connexes ou données bibliographiques (b) l'examen et l'analyse détaillée de ces travaux et (c) une visualisation graphique ou tabulaire des résultats collectés (voir la Figure 2.1).

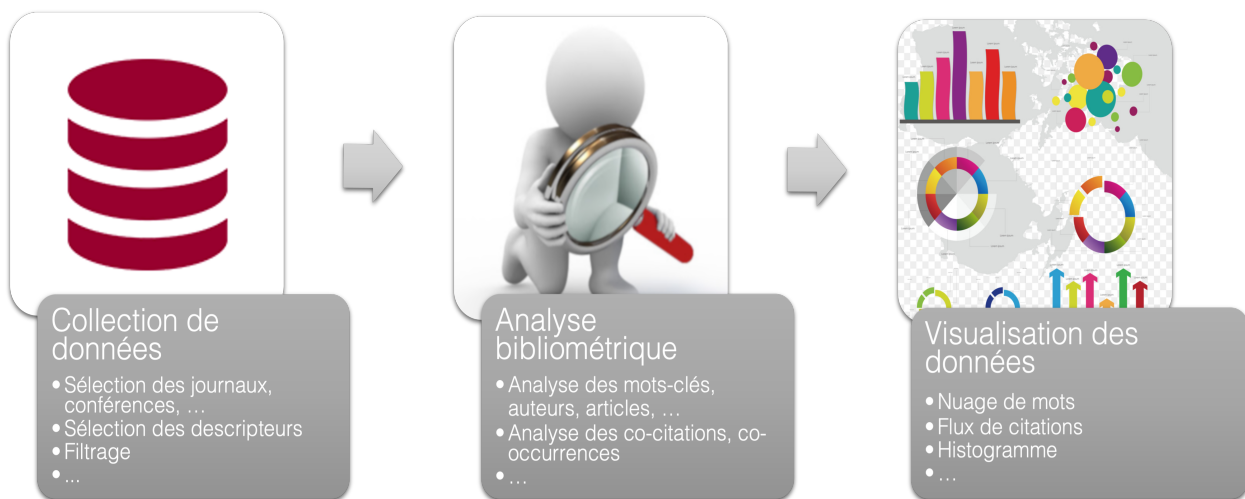


FIGURE 2.1 – Processus bibliométrique

La première étape est une recherche par mots-clés de documents de conférences ou d'articles de journaux, de chapitres de livres et autres moyens de communication afin d'obtenir des données bibliographiques. Cette étape a été effectuée dans les bases de données scientifiques IEEE Xplore digital library et Association for Computing Machinery (ACM) digital library, ainsi que dans les services d'indexation scientifique comme Web of Science (WoS), Elsevier

2.3. Analyse bibliométrique : Vue d'ensemble

Scopus ou Google Scholar. WoS semble être très utilisé grâce à sa pluridisciplinarité. En effet, les bases de données peuvent être très généralistes comme Google Scholar ou très spécialisées comme IEEE Xplorer. Par ailleurs, Scopus, développé par Elsevier ne répond clairement pas aux critères d'accès libre. Par conséquent, dans ce travail, nous nous concentrons sur WoS qui offre également des fonctionnalités de recherche avancée utiles pour sélectionner des ensembles significatifs d'articles pouvant être considérés comme une base intéressante pour construire notre analyse bibliométrique. Il convient de souligner que les données recueillies ont été traitées afin d'éliminer les informations parasites qui auraient entravé les étapes ultérieures de notre analyse.

Il existe un certain nombre d'outils pour analyser les résultats des requêtes de recherche bibliographique. Nous pouvons citer Bibliometrix [AC17] ou VOSviewer [vW10] qui sont d'une utilisation simple avec des interfaces graphiques intuitives.

Bibliometrix est un package de R permettant d'effectuer des recherches quantitatives en scientométrie (par exemple science de la mesure et analyse de la science) et en bibliométrie. Il permet d'importer des données bibliographiques de plusieurs sources comme Scopus ou WoS. En même temps, il propose un système d'évaluation basé sur la co-citation ainsi que la mesure de la collaboration scientifique.

VOSviewer est un logiciel libre pour la construction et la visualisation de réseaux bibliométriques liés aux citations ou aux co-citations, au couplage bibliographique (bibliographic coupling) et la co-occurrence des mots. Le logiciel peut être utilisé pour construire et visualiser des réseaux extraits d'un corpus de littérature scientifique.

Nous avons utilisé dans ce travail Bibliometrix pour analyser les résultats des requêtes de recherche depuis Web of Science.

2.3 Analyse bibliométrique : Vue d'ensemble

L'objectif de cette étude est de cartographier la structure scientifique de la recherche de l'apprentissage automatique dans le secteur de l'agriculture par l'analyse bibliométrique, la visualisation et l'analyse en réseau. Pour ce faire, comme mots-clés de recherche, nous avons utilisé la requête suivante le 3 avril 2021 sur le site web Web of Science ¹ :

("deep learning") AND ("agriculture*" OR "farming")

La question a intentionnellement été formulée d'une manière très générale afin d'élargir le champ de recherches observées. Ainsi, nous avons identifié 440 articles publiés de 2016 à 2021 grâce à la requête ci-dessus. Les données bibliographiques ont été téléchargées sous forme de fichier au format BibTeX et analysées à l'aide de l'outil Bibliometrix.

2.3.1 Sources et auteurs les plus pertinents

La Figure 2.2 montre le nombre de publications depuis 2016 et leur répartition dans le temps des 440 articles scientifiques portant sur l'apprentissage profond en agriculture en utilisant la

1. <https://www.webofscience.com>

base de données WoS. Nous observons que le nombre de publications a augmenté de manière significative à 2020.

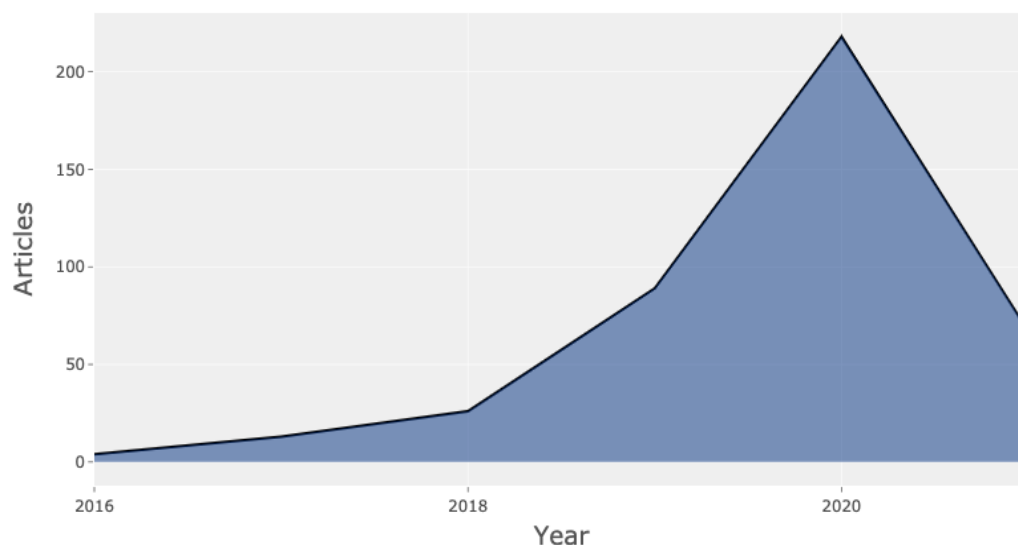


FIGURE 2.2 – Courbe des productions scientifiques annuelles

Par ailleurs, les journaux spécialisés des études agronomiques ou de l'imagerie satellitaire sont les plus actifs en termes de publications d'articles scientifiques (Figure 2.3). Par exemple, «Computers and Electronics In Agriculture» d'Elsevier couvrant le développement et l'application des systèmes informatiques, des logiciels, de l'électronique et des systèmes de surveillance pour résoudre les problèmes de l'agriculture est la source ou le journal le plus utilisé pour les publications avec 66 publications. Il est suivi des journaux en télédétection comme «Remote Sensing» avec 45 publications au total. Parmi les sources les plus citées, nous avons «IEEE Access» de l'Institute of Electrical and Electronics Engineers avec 34 publications au total.

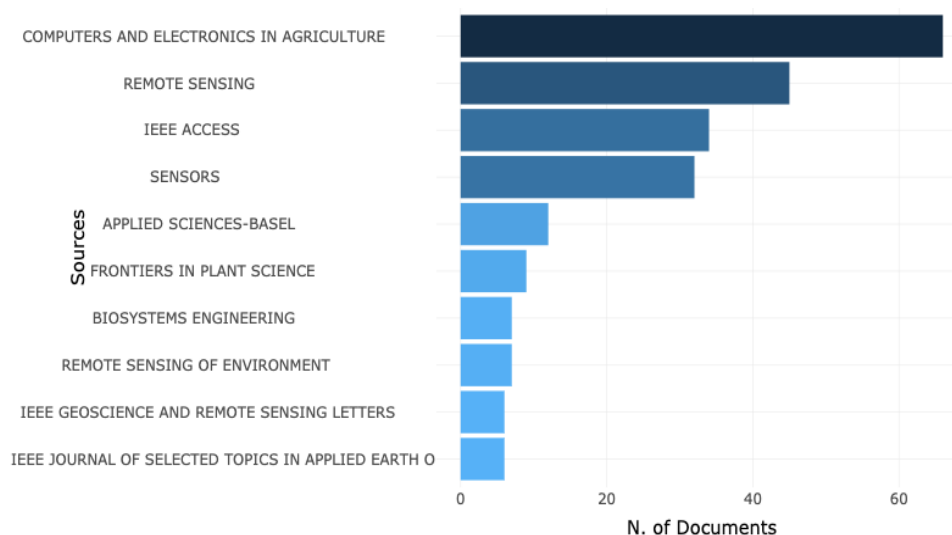


FIGURE 2.3 – Les 10 principales sources les plus pertinentes

La Table 2.1 présente l'impact des cinq (5) premiers auteurs (top-5 ranking) (respectivement des pays) les plus productifs en fonction de l'indice h (h-index), du nombre total de citations

2.3. Analyse bibliométrique : Vue d'ensemble

(Total de Citation ou TC) et du nombre de publications (Number de Publication ou NP). L'indice h (ou indice de Hirsch ou h-index en anglais) est une métrique du niveau de l'auteur permettant de quantifier la productivité scientifique et l'impact d'un scientifique en fonction du niveau de citation c de ses publications p [Hir05]. Le nombre total de citations après $n + 1$ années est alors de :

$$h = \frac{c}{(1 + \frac{c}{p})}n$$

Rang	Auteurs (h-Index)	Auteurs TC	Auteurs NP	Pays TC	Pays NP
1	LIU X 5	KUSSUL N 398	LIU X 8	CHINA 745	CHINA 109
2	CHEN C 4	LAVRENIUK M 398	WANG J 6	USA 690	USA 53
3	LUO J 4	SKAKUN S 398	LI J 6	UKRAINE 398	INDIA 42
4	ZHAO Y 4	SHELESTOV A 391	ZHANG J 6	KOREA 257	KOREA 24
5	PARTEL V 3	FUENTES A 176	ZHANG Z 6	FRANCE 182	AUSTRALIA 18

TABLE 2.1 – Top-5 des publications des auteurs par rapport à H Index, TC (Total de Citation) et NP (Nombre de Publication)

A la lecture de la Table 2.1, LIU X. et CHEN C. sont les auteurs les plus influents en termes de h-Index. KUSSUL N., LAVRENIUK M. et SKAKUN S. obtiennent un TC de 398 chacun. LIU X. (respectivement CHEN C., LUO J, ZHAO Y.) atteint une valeur de 5 (respectivement 4) en termes de h-index. En termes du nombre de publications, LIU X possède 8 documents. Les autres auteurs ont écrit chacun 6 contributions.

Par ailleurs, les pays comme la Chine et les États-Unis d'Amérique arrivent en tête du classement grâce à leurs TC et NP. La Chine obtient un total de 745 citations provenant de 109 publications (contributions). Ces résultats s'expliquent par le fait que la Chine et les États-Unis ont connu une transformation radicale des systèmes agricoles en ayant pour conséquence une forte production agroalimentaire². Pour mieux conduire les exploitations, ces pays ont mis en avant l'intégration et la mise en œuvre d'outils numériques dans les industries agricoles. La digitalisation des moyens de production et de surveillance vise à augmenter considérablement les économies numériques et agricoles des pays concernés. Par exemple, d'ici 2025, l'économie numérique agricole de la Chine devrait représenter 15% de la valeur ajoutée de l'agriculture chinoise, 70% des zones rurales étant alors couvertes par Internet [Xin20].

2. <http://www.fao.org/faostat>

2.3.2 Affiliation la plus pertinente

La Figure 2.4 montre les affiliations les plus pertinentes qui ont effectué le plus de publications. Cette figure confirme les résultats observés dans la Table 2.1 (colonne Pays). Les universités américaines « IOWA State University » et « University of Florida », arrivent en tête du classement avec un maximum de publications de 18 articles pour chacun d’eux. La Chine avec « Beijing Technology And Business University », et « Zhejiang University » se positionne en troisième position avec 16 et 15 publications. Au niveau européen, « Université de Montpellier » de la France a été classée quatrième grâce à ses 15 contributions.

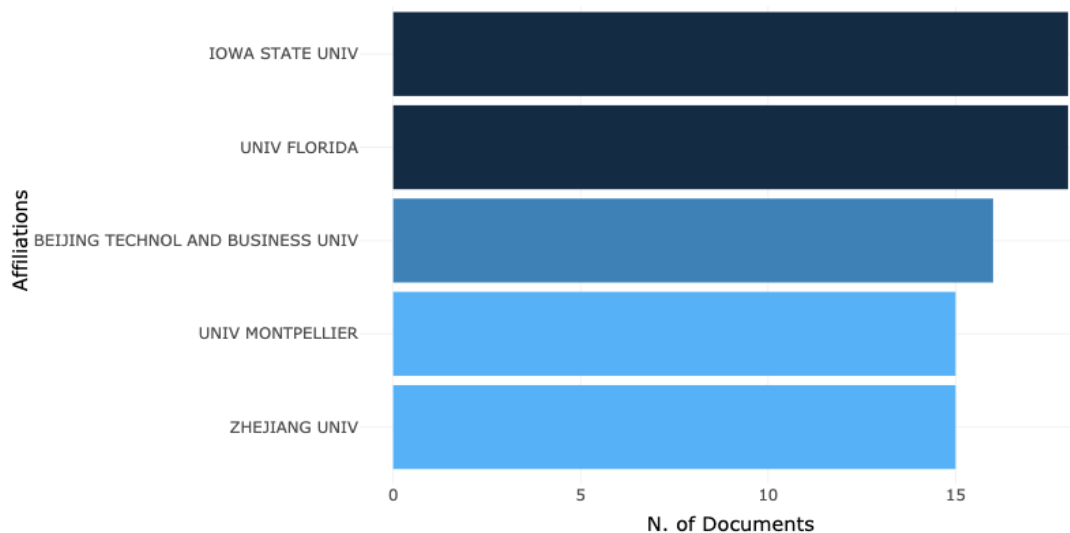


FIGURE 2.4 – Affiliations les plus pertinentes

2.3.3 Top manuscrits par citations

La Table 2.2 résume le top-5 des publications (manuscrits ou documents scientifiques) en termes de nombre total de citations (TC) ou de moyenne de citation par an (Average Citations per Year). La publication de Kussul et collègues nommée « Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data » publiée en 2017 a obtenu le nombre maximum de citations avec 391 citations et un nombre de citations par an de 78,2.

Kussul et al comparent aux réseaux de neurones convolutifs, les modèles d’apprentissage de perceptrons multi-couches et la forêt aléatoire pour la classification des cultures (comme le blé, le maïs, le soja, etc.) pour la région de Kyiv en Ukraine [KLSS17]. Des données multi-temporelles sont acquises par les satellites Landsat-8 et Sentinel-1A pour le besoin de cette classification. Des expérimentations ont montré que les CNN atteignent une meilleure performance avec une moyenne de précision de 88%.

Fuentes et ses collègues ont proposé une approche basée sur l’apprentissage profond pour détecter les maladies et les ravageurs des plantes chez les cultures de tomates [FYKP17]. Ils ont combiné des méta-architectures comme Convolutional Neural Network (Faster R-CNN), Region-based Fully Convolutional Network (R-FCN), and Single Shot Multibox Detector (SSD) avec d’autres architectures basées sur des extracteurs de caractéristiques telles que VGGNet du Visual Geometry Group (VGG) de l’université d’Oxford ou Residual Neural Network (ResNet) de Microsoft Research. Pour accroître la performance de leur approche, les auteurs ont

2.3. Analyse bibliométrique : Vue d'ensemble

TC Rang	Publication (Auteur, Année, Source de la publication, Référence)	Total Citations
1	KUSSUL N, 2017, IEEE GEOSCI REMOTE SENS LETT [KLSS17]	391
2	FUENTES A, 2017, SENSORS [FYKP17]	176
3	ZHONG L, 2019, REMOTE SENS ENVIRON [ZHZ19]	124
4	RAHNEMOONFAR M, 2017, SENSORS [MC17]	122
5	ZHU J, 2018, IEEE INTERNET THINGS J [ZSJS18]	92

Rang TC par an	Publication (Auteur, Année, Source de la publication, Référence)	TC par an
1	KUSSUL N, 2017, IEEE GEOSCI REMOTE SENS LETT [KLSS17]	78,2
2	ZHONG L, 2019, REMOTE SENS ENVIRON [ZHZ19]	41,333
3	FUENTES A, 2017, SENSORS [FYKP17]	35,2
4	RAHNEMOONFAR M, 2017, SENSORS [MC17]	24,4
5	ZHU J, 2018, IEEE INTERNET THINGS J [ZSJS18]	23

TABLE 2.2 – Documents les plus cités

utilisé la technique de l'augmentation de données durant l'expérimentation. Les résultats comparatifs ont prouvé la performance de leurs combinaisons architecturales sur les architectures standards avec une moyenne AP (Average Precision) de 83% pour Faster R-CNN avec VGG-16 ou 82.53% pour SSD avec ResNet-50.

Zhong et al ont proposé la classification des cultures basée sur l'apprentissage profond des séries temporelles télédéteectées [ZHZ19]. Dans cette étude, les modèles d'apprentissage profond basés sur la mémoire à long terme (LSTM) et sur les couches de convolutions unidimensionnelles (Conv1D) sont comparés aux modèles classiques comme XGBoost, Forêts aléatoires et Machine à Vecteur de Support. L'expérimentation a montré que les couches convolutives Conv1D sont efficaces et efficientes pour la représentation des séries temporelles dans les tâches de classification multi-temporelles avec une précision de 85.54% par rapport aux autres modèles cités ci-dessus. En somme, 13 catégories (comme le riz, le maïs, la tomate, le vignoble, etc.) de cultures d'été du comté de Yolo, en Californie, aux Etats-Unis, ont été sélectionnées pour la tâche classification.

Dans [MC17], les auteurs ont élaboré une méthodologie basée sur les réseaux de neurones convolutifs pour l'estimation du rendement des cultures. Leur objectif consiste à réduire le coût d'étiquetage (marquage) des données d'entraînement pour le problème du dénombrement des objets en créant un ensemble de jeux de données secondaires pour l'entraînement. À l'aide d'une version modifiée de l'architecture couplant les modèles GoogleNet ou Inception et Re-

sNet, leur algorithme compte le nombre exact de fruits ou des fleurs. Les résultats expérimentaux entraînés sur des données simulées ont donné une précision moyenne de 91% sur un ensemble d'images réelles.

Aujourd'hui, les technologies des objets connectés (IoT) existants doivent ajouter des éléments intelligents et passer de la « perception » à la « cognition » en combinant l'IoT avec des méthodes cognitives [Fou20]. Dans l'IoT cognitif, la technologie de mise en réseau d'auto-organisation peut utiliser la collaboration de groupes entre les nœuds pour accomplir la mission commune, qui reflète le raisonnement, la distribution et la robustesse de l'IoT [SDMC20]. L'IoT cognitif permet aux organisations d'apprendre à partir des données provenant d'appareils connectés, de capteurs, de machines et d'autres sources, et infuse l'intelligence dans les opérations industrielles et les expériences des acteurs sur les produits et les machines. En effet, ils constituent un levier de développement pour l'agriculture du futur avec l'augmentation de la capacité d'apprendre, de penser et de comprendre des problèmes dans des situations complexes. Dans cette digitalisation du système agricole, les utilisateurs ont un fort besoin d'automatisation de la collecte et l'analyse des données sur les sites d'exploitation et d'amélioration des défaillances du système productif (désherbage, récolte, traitement, etc.). Zhu et al. mettent en œuvre un nouveau mécanisme d'ordonnancement de transmission des informations entre les objets connectés avec une technique d'apprentissage profond par renforcement (Deep Q-Learning) [ZSJS18]. La méthode Q-Learning permet d'apprendre une politique, qui indique quelle action effectuer dans chaque état du système. Ceci facilite la comparaison des récompenses probables sur la prise des actions disponibles à optimiser sans avoir la connaissance initiale des états de l'environnement. L'apprentissage de l'algorithme Q-Learning est un processus d'optimisation graduelle jusqu'à l'atteinte d'une convergence par la sélection des actions optimales. De ce fait, le Deep Q-Learning joue un rôle important dans la transmission et la réception des paquets de données dans les réseaux sans fil pour maximiser le débit du système lors des échanges entre les nœuds de ce système.

2.3.4 Mots-clés les plus pertinents

La figure 2.5 montre un graphique des mots-clés d'auteurs (most popular author keywords) les plus pertinents par leur popularité dans les données collectées avec une visualisation appelée nuage de mots (Word Cloud). La taille d'un élément de graphique est proportionnelle au nombre de documents où le mot-clé apparaît. Ces mots-clés correspondent aux principaux sujets d'intérêts de recherche en apprentissage profond dans l'agriculture. Les mots-clés obtenus sont entre autres : agriculture de précision (smart agriculture), traitement d'images (image processing), segmentation ou prédiction de culture (segmentation ou prediction crop) ou reconnaissance des images (image recognition), détection des maladies ou mauvaises herbes (diseases ou weed detection), apprentissage par transfert (transfer learning), réseaux de neurones convolutifs (convolutional neural network), vision par ordinateur (computer vision), extraction des caractéristiques (feature extraction) et classification.

Grâce à ces mots-clés, nous remarquons que l'agriculture est confrontée à de nombreuses difficultés comme l'identification des maladies et la reconnaissance des adventices (mauvaises herbes) qui sont couramment traitées. Une résolution traditionnelle de ces difficultés peut favoriser chez les agriculteurs une utilisation abusive des pesticides pouvant dégrader la qualité de la production avec des effets ou conséquences néfastes sur la santé humaine et l'environne-

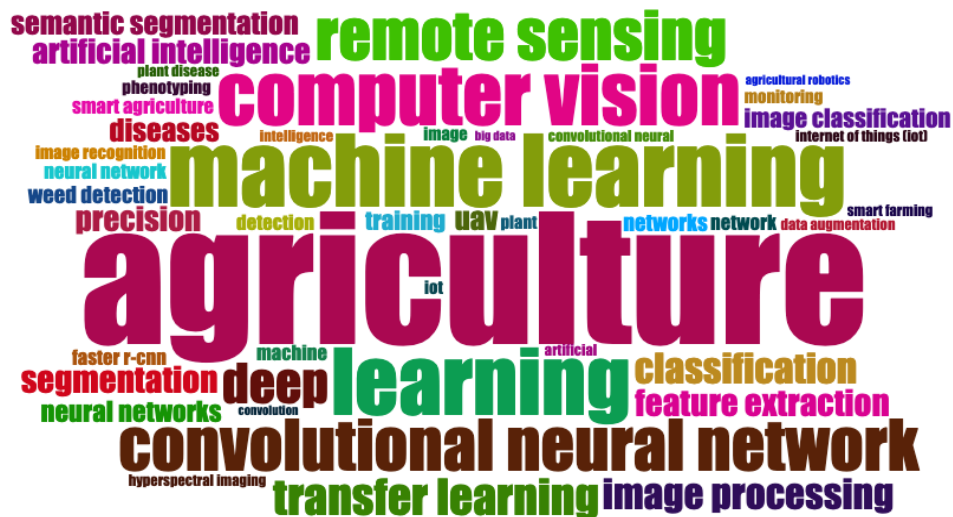


FIGURE 2.5 – Mots-clés les plus pertinents

ment. Pour lever ces difficultés, une piste de solution prometteuse consiste à la digitalisation de l'agriculture traditionnelle offrant des opportunités d'automatisation du traitement des données collectées. Les méthodes avancées d'automatisation des données de l'agriculture consistent à appliquer les techniques d'apprentissage automatique comme les réseaux de neurones convolutifs qui ont montré des résultats encourageants, en particulier dans le domaine de la vision par ordinateur. Le principe de fonctionnement de ces techniques d'apprentissage profond facilite l'extraction des caractéristiques, l'apprentissage par transfert ou la télédétection dans le contrôle et la gestion des exploitations agricoles. Ceci encourage l'identification précoce des problèmes dans les cultures avec des leviers de limitation de l'utilisation des intrants (produits phytosanitaires, traitements chimiques, etc.), et d'amélioration des rendements agricoles. Des travaux s'y rapportant à des auteurs recueillis sur WoS sont présentés dans la section suivante.

2.3.5 Relations mots-clés, pays et sources avec un diagramme de flux

Les pays comme la Chine, les États-Unis, l'Inde ou le Brésil ont une communauté de recherche très active sur ces thématiques relevant de l'agriculture de précision. Sur la Figure 2.6, on a une visualisation des principaux éléments de trois champs (par exemple, les journaux, les mots-clés, les pays), et leurs liaisons à travers un diagramme de Sankey (type de diagramme de flux). Dans cette représentation de flux avec le diagramme de Sankey, la taille d'un élément indique son importance qui est proportionnelle aux nombres de nœuds qu'il possède. L'analyse montre les sources (journaux le plus souvent) dans lesquels les auteurs ont le plus publié et les sujets de recherche les plus abordés identifiés à travers les mots-clés dans les pays concernés.

2.3.6 Répartition des travaux de recherche en agriculture de précision

En agriculture, la reconnaissance et l'identification traditionnelle (l'inspection visuelle par des experts ou l'examen biologique) des problèmes et leurs risques dans les exploitations demandent beaucoup de temps et deviennent quasi-impossibles lorsque les exploitations sont grandes. Les solutions de l'intelligence artificielle utilisent la vision par ordinateur dont le principal but est de permettre à une machine d'analyser, de traiter et de comprendre une ou

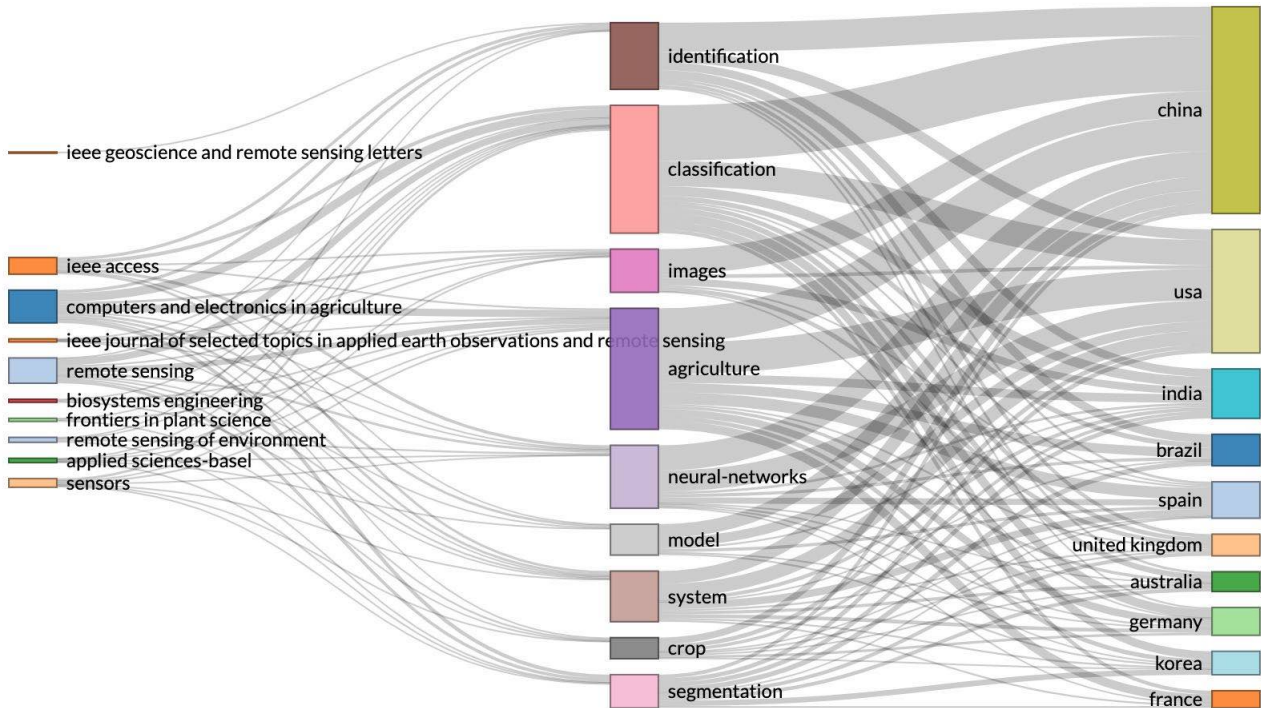


FIGURE 2.6 – Visualisation du flux de citations par le diagramme de Sankey

plusieurs images prises à partir d'un système d'acquisition de données. Cette vision par ordinateur est basée sur la localisation ou la reconnaissance automatique des objets et améliore considérablement la qualité des activités agricoles. Les Tableaux 2.3 à 2.5 fournissent une classification des certains travaux de recherche d'apprentissage automatique appliquée au secteur agricole en fonction de leur approche, du type de problèmes et des résultats obtenus.

Article	Approche	Type de problèmes	Résultat
[PKA19]	CNN basé sur le détecteur d'objets YOLOV3	Développement et évaluation d'un pulvérisateur intelligent afin de faire la distinction entre les mauvaises herbes et les plantes de tomates sur des données réelles et artificielles	Une des expérimentations fournit une précision de 90% et un rappel global de 89% avec des plantes artificielles, et respectivement de 59% et 44% avec des plantes réelles.

TABLE 2.3: Présentation des travaux de recherche sur la détection de mauvaises herbes

2.3. Analyse bibliométrique : Vue d'ensemble

Article	Approche	Type de problèmes	Résultat
[RvHB ⁺ 20]	CNN basé sur le détecteur d'objets YOLOv3 + augmentation de données (data augmentation)	Système de détection de mauvaises herbes au niveau de la plante, intégré à un robot de pulvérisation autonome de pomme de terre	Rappel (Recall) : 57% ; Précision (Precision) : 84%
[BHC18]	CNN (ResNet18) + extraction de caractéristiques (feature extraction) et affinement du modèle (fine-tuning)	Apprentissage profond non supervisé avec étiquetage des données pour la détection de mauvaises herbes dans les cultures sur des images de drones	Champs d'épinards : Areas Under the Curve (AUC) : 94,34% (étiquetage non supervisé des données) et 95,70% (étiquetage supervisé des données). Champ de haricots : AUC : 88,73% (étiquetage non supervisé des données) et 94,84% (étiquetage supervisé des données)
[LBMS18]	Combinaison des réseaux CNN (Fully Convolutional DenseNet) + AutoEncodeur	Système de classification de mauvaises herbes sur les betteraves avec un réseau entièrement convolutifs et des informations séquentielles	F1 score : 92.4% sur dataset n°1 et 86.6% sur dataset n°2.

TABLE 2.3: Présentation des travaux de recherche sur la détection de mauvaises herbes

Article	Approche	Type de problèmes	Résultat
[EMA ⁺ 20]	CNN+ transfer learning (réseau affiné et pré-entraîné). Combinaison des réseaux de neurones convolutifs comme Xception, Inception-Resnet, VGGNet, MobileNet et DenseNet pour l'extraction de caractéristiques avec d'autres classificateurs d'apprentissage automatique	Identification des cultures et mauvaises herbes à l'aide de l'apprentissage par transfert	F1-score : 99,29% (Fine-tuned DenseNet + Support Vector Machine). Par l'extraction des caractéristiques (Feature extraction) avec DenseNet + Support Vector Machine a un score F1 de 98 %.
[TIT ⁺ 21](Tufail et al., 2021)	ResNet18 pré-entraîné sur ImageNet, Fined-tuned MobileNet v2	Détection de mauvaises herbes sur des cultures et de tabac pour un pulvérisateur à rampe monté sur un tracteur.	Accuracy : ResNet18 100% MobileNet v2 : 81 %

TABLE 2.3: Présentation des travaux de recherche sur la détection de mauvaises herbes

Article	Approche	Type de problèmes	Résultat
[KHC20]	Segmentation d'images avec l'architecture SegNet formée par un Encoder-Decoder VGG16 + LeNet5 + augmentation des données (data augmentation)	Détection des maladies de la vigne à l'aide d'images multispectrales de drones par l'apprentissage profond	F1 :92.81%, Précision (Precision) : 92. 78%, Rappel (Recall) : 92.91%, Justesse (Accuray) : 95.02%

TABLE 2.4: Présentation des travaux de recherche sur les maladies de culture

2.3. Analyse bibliométrique : Vue d'ensemble

Article	Approche	Type de problèmes	Résultat
[GZY ⁺ 20]	Définition d'un modèle CNN avec une architecture spécifique (Proposed CNN model)	Identification des maladies (pourriture noire, plaque bactérienne et maladies de la rouille) basée sur l'apprentissage profond dans l'agriculture de précision	Accuracy : 83.57%
[VCS20]	Proposed CNN (Capsule Networks ou CapNet)	CapNet pour la classification des maladies de la pomme de terre.	Accuracy : 91.83%
[AKS ⁺ 19]	CNN (AlexNet, VGG19, Inception v3, DenseNet201, ResNet152) + augmentation de données avec les réseaux antagonistes génératifs (Generative Adversarial Network (GANs))	Résolution des limites des approches de détection des maladies des plantes basée sur l'apprentissage profond	Mean Average Precision : 93.67%
[CKKT19]	CNN (VGG16) + apprentissage par transfert + augmentation des données (Transfer learning + data augmentation)	Détection de la maladie de mildiou dans le mil à base d'apprentissage par transfert	Accuracy : 95% F1 : 91.75% Precision : 90.50%, Recall : 94.50%
[PSA ⁺ 19]	CNN (ResNet50)	Classification des maladies des plantes incorporant des informations contextuelles (comme les espèces végétales, et les conditions météorologiques)	Accuracy : 98%

TABLE 2.4: Présentation des travaux de recherche sur les maladies de culture

Article	Approche	Type de problèmes	Résultat
[NJS ⁺ 19]	3D-CNN	Identification des maladies des plantes de soja sur une image hyperspectrale à l'aide de l'apprentissage profond avec une approche d'explicabilité et de visualisation du modèle basé sur une carte de saillance (saliency-map)	Accuracy : 95.73% ; Recall : 92% ; Precision : 82% ; F1-Score : 87%
[SCJJ19]	CNN	Classification des feuilles de mangues infectées par l'antracnose	Accuracy : 97.13%
[FYKP17]	Faster RCNN, R-FCN, SSD combiné + VGGNet, ResNet	Détection des maladies et les ravageurs des plantes (diseases and pests) chez les cultures de tomates	Moyenne des précisions moyennes (mAP pour mean Average Precision) de 83%

TABLE 2.4: Présentation des travaux de recherche sur les maladies de culture

Article	Approche	Type de problèmes	Résultat
[CHL20b]	Réseaux de neurones convolutifs de type modèle de YOLO pour la détection des objets en temps réel (Convolutional Network YOLO model)	Amélioration de la détection des parasites par l'extraction des caractéristiques (entier et partiel) basée sur l'apprentissage par transfert	Moyenne des précisions moyennes (mAP pour mean Average Precision) : 90% ; Accuracy : 90% ; Recall : 85%
[ASB20]	CNN (ResNet34) + apprentissage par transfert	Classification des parasites dans les cultures de coton en utilisant les réseaux de neurones résiduels profonds (deep residual networks)	Accuracy : 98%

TABLE 2.5: Présentation des travaux de recherche sur la reconnaissance des parasites

2.3. Analyse bibliométrique : Vue d'ensemble

Article	Approche	Type de problèmes	Résultat
[WWX ⁺ 20]	Deeppest : Réseaux de neurones convolutifs + informations contextuelles (comme les informations géographiques, la température ambiante)	Détection des parasites à partir des images des cultures du riz ou du blé	Moyenne des précisions moyennes (mAP pour mean Average Precision) : 74.3% ;
[LWZ ⁺ 19]	CNN + augmentation de données (data augmentation)	Localisation et reconnaissance des parasites dans les exploitations de blé (wheat)	Moyenne des précisions moyennes (mAP pour mean Average Precision) : 83.23% ;
[DLJ ⁺ 19]	CNN (AlexNet)	Reconnaissance des parasites basée sur un réseau de neurones convolutifs avec un apprentissage par transfert	Accuracy : 93.86%
[TR19]	CNN + apprentissage de transfert (transfer learning) + augmentation des données (data augmentation)	Classification des insectes ravageurs des cultures basée sur un réseau neurones convolutifs et un apprentissage par transfert	Accuracy : 97.74%
[CHL ⁺ 20a]	CNN (YOLOv3) + augmentation de données (data augmentation)	Système intelligent d'agriculture de précision basé sur des capteurs d'environnement et IoT pour la détection des parasites	Accuracy : 90% ; mAP : 92%

TABLE 2.5: Présentation des travaux de recherche sur la reconnaissance des parasites

Les avancées techniques de l'apprentissage profond à base des réseaux de neurones convolutifs (CNN) ont été largement appliquées avec succès dans la classification des cultures, l'identification de mauvaises herbes, la détection des maladies, le dénombrement des fruits, même à des stades précoces de l'apparition des infections. Cependant, les modèles CNN ne tiennent pas compte des relations spatiales entre les caractéristiques d'une image et ils sont sensibles aux variations dans les images d'une classe à prédire, d'où la nécessité d'avoir une grande quantité de données d'apprentissage. Dans certaines études, ce problème est résolu par la technique de

l'augmentation de données qui s'effectue par la rotation, l'agrandissement d'une zone ou autres méthodes de transformation de l'image comme par l'utilisation des GANs.

Par ailleurs, les CNN comme les modèles utilisant les architectures AlexNet, VGGNet, GoogleNet, ReseNet ou DenseNet qui ont déjà montré leur performance lors des concours « ImageNet Large Scale Visual Recognition Competition (ILSVRC) » ont fourni des résultats satisfaisants lors des évaluations expérimentales. Ces résultats sont parfois améliorés par les techniques d'apprentissage par transfert, la combinaison de plusieurs réseaux de convolutions ou le pré-traitement (comme la segmentation) des données d'apprentissage. Les mesures (métriques) d'évaluation, considérées comme Justesse (Accuracy), F-mesure (F-score), Précision (Precision), Rappel (Recall), Moyenne des « moyennes de précisions » (mAP pour mean Average Precision) sont les plus utilisées. Elles indiquent chacune le pourcentage des instances mal classifiées par rapport aux instances correctement classifiées. Cependant, les mesures F1-score ou mAP semblent être des bons compromis lorsqu'il y a lieu de comparer plusieurs classificateurs sur des ensembles de données distincts (différents).

Les ensembles de données d'apprentissage généralement constitués d'images en couleur au format Rouge Vert Bleu (RVB ou Red Green and Blue ou RGB) et spectrales sont multi-sources (diverses et variés). Pour certains ensembles, elles sont acquises sur le terrain (dans des conditions réelles) par des drones, des appareils photos ou fournies par des systèmes satellitaires ; pour beaucoup d'autres, il s'agit des données de références disponibles en libre accès sur Internet (PlantVillage³, AgriPest [WLX⁺21] ou NBAIR⁴).

2.3.7 Pays et publications des auteurs en correspondance

La Figure 2.7 donne une représentation graphique des pays et des publications des auteurs en correspondance. Avec la Table 2.6, elles fournissent leur nombre de publications et l'analyse du nombre de publications avec les auteurs d'un seul pays (SCP), dans plusieurs pays (MCP) et du ratio de publications dans plusieurs pays. La Chine et les États-Unis sont les pays les plus importants avec un total de 109 et 53 publications.

La Chine est en tête de la liste avec 77 publications avec les auteurs d'un seul pays (SCP) et 32 des publications de plusieurs pays (MCP), avec un ratio MCP de 0,29. Les États-Unis ont 40 publications avec les auteurs d'un seul pays (SCP) et 13 des publications avec les auteurs de plusieurs pays (MCP), avec un ratio MCP de 0,24.

Country	Articles	SCP	MCP	MCP Ratio
CHINA	109	77	32	0,2936
USA	53	40	13	0,2453
INDIA	42	38	4	0,0952
KOREA	24	19	5	0,2083
AUSTRALIA	18	11	7	0,3889

TABLE 2.6 – Publication des pays pour les auteurs en correspondance.

3. <https://github.com/spMohanty/PlantVillage-Dataset>

4. <http://www.nbair.res.in/insectpests/pestsearch.php> (National Bureau of Agricultural Insect Resources)

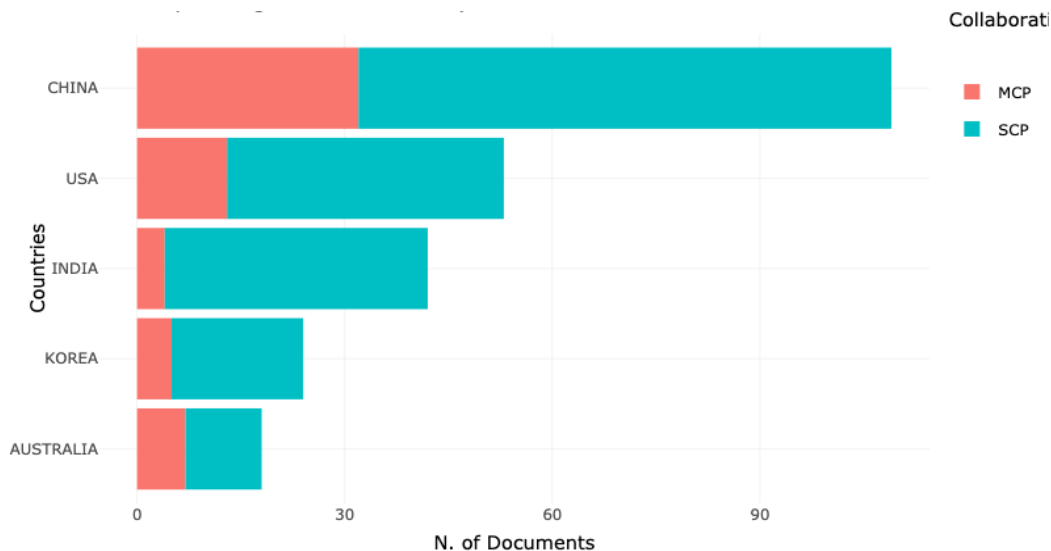


FIGURE 2.7 – Publication des pays pour les auteurs en correspondance

2.4 Discussion

Les résultats expérimentaux des travaux étudiés présentent certaines limites liées à la visualisation des facteurs influents dans la prédiction et à l'utilisation de multiples classificateurs (classificateurs) :

En premier lieu, si divers classificateurs sont inclus dans les expériences, alors il devient nécessaire d'effectuer des tests statistiques pour analyser la signification des résultats obtenus [BCDZ17]. Demšar et al. ont examiné plusieurs [Dem06] tests statistiques non paramétriques, sûrs, robustes et étudié leur adéquation dans le domaine de l'apprentissage automatique pour analyser les différences significatives dans les performances de classification. Les tests qui ont été recommandés sont le test des rangs signés de Wilcoxon pour la comparaison de deux classificateurs et le test de Friedman avec les tests post-hoc pour la comparaison de plusieurs classificateurs sur plusieurs ensembles de données. Les résultats de ce test peuvent être présentés à l'aide des diagrammes de différence critique (Critical Difference ou CD) pour plus de clarté.

En second lieu, lorsque nous voulons estimer la fiabilité d'un modèle prédictif basé sur l'échantillonnage, la technique de validation croisée (voir la section 1.3.6) sur les données d'apprentissage est utile. Autrement, les résultats du processus d'apprentissage vont dépendre d'un choix aléatoire particulier pour les ensembles d'entraînement et de validation. Cette technique effectue une suite de découpage distincte des données en un ensemble d'entraînement et de validation.

En troisième lieu, l'exactitude de la prédiction d'un modèle est inversement proportionnelle à son explicabilité, i.e les modèles d'apprentissage profond ont un fonctionnement opaque pour les humains. Ce concept d'explicabilité ou d'intelligence artificielle explicable (eXplainable AI ou XAI) a été soutenu par l'Agence du Département de la Défense des États-Unis (DARPA) donnant lieu à une réorientation des problématiques liées à l'apprentissage automatique [GA19]. On note l'absence d'éléments relatifs à cette nouvelle préoccupation d'une intelligence artificielle explicable dans la plupart des travaux étudiés. L'objectif de l'explicabilité est de fournir aux utilisateurs, une garantie, une justification, une confiance dans l'utilisation des modèles d'apprentissage et de proposer des recommandations pour l'amélioration de leurs

performances. Par exemple, nous pouvons visualiser l'ensemble de caractéristiques (comme les pixels d'une image) qui ont contribué à la décision du modèle ou fournir des éléments de compréhension sur son fonctionnement [RFFT17, SVZ14, ZF13].

Il y a un grand besoin d'apporter des éclairages sur l'usage des méthodes l'apprentissage automatique dans les domaines sensibles touchant à la santé ou à l'alimentation de l'humain. C'est dans cette optique que se situent les potentialités de l'apport de l'explicabilité [BBM⁺15] dans le domaine agricole. Il en découle des opportunités de faire participer les exploitants [CSHB18, RSG16, SVZ14] agricoles comme acteurs dans le processus d'apprentissage automatique à partir des données de terrain pour une pratique agricole plus intelligente, plus humaine et plus écologique. La participation de ces acteurs dans le processus d'apprentissage peut se faire à trois niveaux :

- L'acquisition et la préparation des données d'apprentissage qui seront traitées ultérieurement,
- L'assistance à l'analyse des résultats partiels pour guider la suite du processus de l'apprentissage de manière progressive,
- L'évaluation du processus d'apprentissage par l'expression du niveau de satisfaction avec des interactions.

Ainsi nous présentons une optimisation interactive du processus d'apprentissage automatique des données agricoles par la Figure 2.8. Elle permet de concaténer toutes les informations et les résultats obtenus durant le processus d'apprentissage avec la participation des utilisateurs finaux (les acteurs agricoles).

Le déploiement d'un système agricole de précision intégrant une intelligence artificielle explicable est possible dans des situations réelles. En effet, il existe actuellement quelques exemples pratiques d'applications agricoles utilisant l'apprentissage automatique tout en essayant de répondre à cette nouvelle exigence d'explicabilité. Par exemple, des cartes de visualisations ont été utilisées pour l'identification des niveaux de stress et des types d'infections au niveau des plantes [GBS⁺18, NJS⁺19].

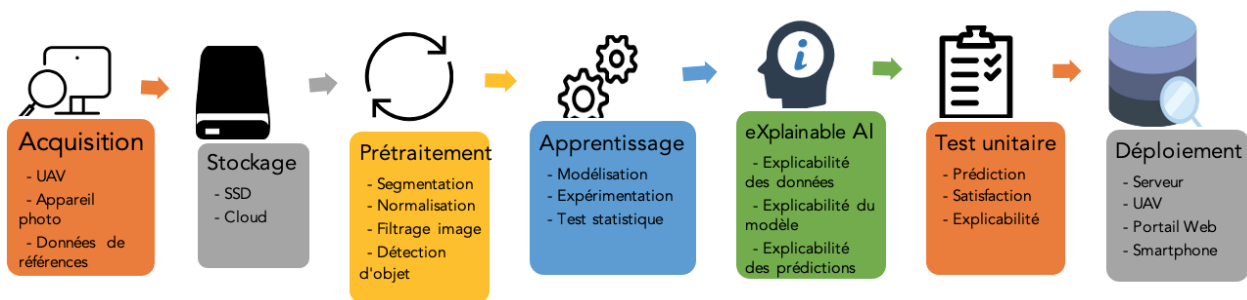


FIGURE 2.8 – Développement d'un système agricole intelligent (SSD : Solid-State Drive ; UAV : Unmanned Aerial Vehicle)

2.5 Conclusion

L'agriculture de précision est une science récente qui se retrouve certainement encore dans sa phase de jeunesse. L'intelligence artificielle peut avoir une contribution déterminante dans l'analyse des données agricoles, en l'occurrence en utilisant la vision par ordinateur permettant

à une machine d'analyser et de comprendre automatiquement le monde visuel. Ceci offre l'opportunité de développer des systèmes intelligents qui apparaissent comme l'un des moyens pour répondre aux défis de performance économique, environnementale et sociale en agriculture. La revue de littérature effectuée montre une floraison de travaux décrivant le développement au cours des dernières années de divers systèmes utilisant l'apprentissage automatique profond dans la résolution de divers problèmes de l'agriculture de précision. Les problèmes traités concernent principalement les pertes de rendements occasionnées par les facteurs affectant la bonne croissance des cultures agricoles et la fertilité des sols. Sur ce domaine de recherche scientifique, les analyses bibliométriques montrent que les pays les plus actifs sont la Chine, les États-Unis, l'Inde et le Brésil. Les thématiques abordées recouvrent entre autres les mots-clés « identification des maladies », « extraction des caractéristiques », « apprentissage par transfert », « détection des maladies », « détection de mauvaises herbes », « réseaux de neurones à convolution », « traitement d'images ». Les réseaux de neurones à convolution (CNN) sont l'une des architectures les plus couramment utilisées du fait de ses applications performantes dans les méthodes d'apprentissage profond.

Toutefois, l'usage des outils intelligents en agriculture de précision soulève de nombreux défis dont la résolution est utile pour renforcer ce domaine et faciliter son appropriation par les acteurs du domaine :

- La faiblesse des données d'apprentissage en termes de quantité et de qualité constitue un important frein pour plus de précisions dans les décisions suggérées par les systèmes intelligents (les systèmes d'aide à la décision ou les outils d'analyses spatiales) dans le domaine agricole en pleine évolution numérique avec de nouvelles contraintes de respect de normes environnementales.
- La profondeur des couches dans les réseaux de neurones profonds, le nombre élevé des paramètres et la complexité des algorithmes d'apprentissage de données ne simplifient pas les temps de calcul. Ceci peut engendrer des procédures additionnelles (e.g. contraintes supplémentaires pour les systèmes embarqués) en vue de leur déploiement de manière appropriée sur les objets connectés en agriculture.
- La compréhension des mécanismes de raisonnement sous-jacents aux systèmes d'apprentissage automatique profond doit être améliorée. Il s'avère donc utile de fournir des pistes d'explications pour une meilleure compréhension des prédictions fournies.

Une tentative de contournement du problème de la faiblesse des données peut augmenter la précision des architectures issues de l'apprentissage profond. Cette problématique est discutée dans le prochain chapitre qui met l'accent sur la réutilisation ou le partage des connaissances provenant des tâches précédemment apprises pour une nouvelle tâche.

Chapitre 3

Réseaux de neurones convolutifs pour la classification multi-labels des images

3.1 Introduction

L'apprentissage profond est très utilisé pour les tâches de classification d'images binaire et multi-classes [ASM17, SS19, SSC17]. Dans ces nombreux cas, la segmentation d'objet [GOO⁺18, LKAL16], la détection d'objets [CH16, GKCA17], la télédétection [ALYS17, FK18] et autres applications [SGAS17, Was18], l'apprentissage profond obtient de meilleurs résultats. La classification multi-labels est une classe importante du problème d'apprentissage automatique qui trouve ses applications dans l'annotation d'images [BLSB04, WYZZ09], la catégorisation de textes, l'annotation de fonctions génétiques [LYG⁺10, YLCL18]. Elle consiste à apprendre simultanément plusieurs étiquettes pour une donnée quelconque. Lorsque l'on résout les problèmes de MLC, on s'attend généralement à ce que l'algorithme tienne compte la corrélation entre des étiquettes des données pour améliorer les performances de prédiction [HYG⁺19, HLW⁺17, PPZM18]. L'extraction de la corrélation cachée est habituellement une tâche difficile lorsqu'on traite des images cas avec plusieurs objets en avant ou arrière-plan. D'un côté, le fait d'ignorer les relations entre les étiquettes risque d'aboutir a des prédictions incohérentes, et d'un autre côté, le fait de prendre en compte ces relations risque de propager l'erreur de prédiction d'une étiquette a tous les autres qui lui sont reliées [LMR17]. L'approche que nous proposons permet l'apprentissage des relations entre les étiquettes au travers d'une description vectorielle binaire de chaque étiquette et d'un apprentissage des multiples caractéristiques de chaque image.

Généralement, on retrouve deux grandes catégories pour le problème de la classification multi-labels [TK07] : (i) les méthodes de transformation du problème et (ii) les méthodes d'adaptation de l'algorithme. Avec la méthode de transformation de problème, un ensemble de données à étiquettes multiples est converti en un ensemble de données à étiquette unique. Certaines transformations de problèmes tiennent compte des corrélations entre les étiquettes pendant le processus de classification [WLW⁺18, WGZ18] et d'autres ignorent cette dépendance [MGCV18]. Un exemple de transformation connu est celui de la pertinence binaire (Binary relevance). C'est une méthode très répandue de transformation de problèmes qui consiste à apprendre un certain nombre de classificateurs binaires, soit un classificateur pour chaque étiquette distinctive dans l'ensemble fini des étiquettes. Les méthodes d'adaptation de l'algorithme proprement dites portent l'idée d'une extension pour des algorithmes d'apprentissage spécifiques

(comme le modèle génératif probabiliste, les champs aléatoires conditionnels, l'approche de déconvolution) afin de manipuler directement des données à étiquettes multiples pour faire des prédictions multi-labels. Par exemple, l'algorithme des champs aléatoires conditionnels cherche à exploiter la corrélation entre les étiquettes pour une meilleure classification.

Les méthodes de classification d'images courantes avec les réseaux de neurones profonds sont celles d'une classification binaire et multi-classes. Ces méthodes ont besoin d'une adaptation pour relier les probabilités d'appartenance d'une observation à des catégories d'images lors d'une classification multi-labels. Dans ce contexte, nous proposons une nouvelle approche de classification multi-labels en construisant un réseau de neurones à multiples branches. Nous essayons de répondre aux problèmes et questions suivantes : tout d'abord, comment transformer un problème de classification multi-labels en un problème de classification binaire? Deuxièmement, comment résoudre chaque problème afin de trouver les classes d'étiquettes pertinentes? La solution à ce problème utilise la technique de diviser pour mieux régner dont chaque problème considéré comme une tâche à résoudre, est subdivisé en des sous-tâches. Nous combinons ces différentes sous-tâches pour construire l'espace final des étiquettes. Partant de là, la tâche de classification consiste en (i) la construction d'un modèle multi-branches (ii) le traitement par chaque branche d'une partie du problème et (iii) la combinaison des résultats de chaque branche pour construire l'espace d'étiquette final qui est résultat de la prédiction.

Ce travail est organisé comme suit : la section 3.2 présente quelques travaux connexes, la section 3.3 explique le principe fondamental de l'apprentissage multitâches, la section 3.4 décrit les architectures du réseau de neurones à multiples branches, les sections 3.5 et 3.6 évaluent des performances du réseau de neurones à multiples branches dans la classification multi-labels, la section 3.7 discute des résultats obtenus et la section 3.8 conclut ce chapitre.

3.2 Travaux connexes

Dans cette section, nous passons en revue quelques travaux connexes sur les problèmes de classification multi-labels et le réseau de neurones multi-branches .

Yidong et al. utilisent le réseau de neurones à multiples branches pour la détection automatique du glaucome, une maladie oculaire qui endommage le nerf optique. Le modèle qu'ils proposent est composé de trois branches [CLX18] : la première branche considère l'entrée entière de l'image et extrait les caractéristiques dans un CNN. La deuxième branche fait usage de l'architecture fast region-based CNN (systèmes de détection d'objets utilisant un CNN) pour obtenir la région du disque optique. La troisième branche est constituée de couches complètement connectées pour déterminer les éléments de connaissance du domaine contenant des images et des éléments (données) non-images.

Bianco et al. proposent un réseau de neurones à multiples branches pour la catégorisation des peintures. Les différentes branches classifient les artistes et leurs styles [BMS17]. Leur modèle est constitué de cinq modules : trois branches pour collecter les indications des éléments de texture de bas niveau et, elles utilisent la disposition brute de la peinture. Les sorties des trois branches sont liées pour représenter simultanément les caractéristiques brutes et fines de la peinture. Enfin, un module de classification exploite les blocs résiduels afin de traiter le problème des gradients évanescents.

Zhuang et ses collaborateurs définissent un apprentissage par transfert pour la classification des attributs du visage tels que l'attraction, le sexe, la race, etc. Ils ont construit un réseau à structure

complexe qui se compose de trois sous-réseaux [ZYC⁺18] : le premier sous-réseau est le réseau de détection des visages (Face detection Network ou FNet), le deuxième est appelé réseau d'apprentissage multi-labels (Multi-label learning Network ou MNet) et le dernier sous-réseau est le réseau d'apprentissage par transfert (Transfer learning Network ou TNet). Le FNet qui se base sur Faster RCNN est réentraîné pour la détection des visages. Le MNet est également ajusté par le FNet et sert à prédire divers attributs du visage avec des données étiquetées. TNet est basé sur MNet pour la prédiction à partir des ensembles de données ayant des attributs du visage non étiquetés. Les résultats expérimentaux sur des données de visage montrent l'efficacité de la méthode proposée par rapport aux méthodes standards en entraînant plusieurs attributs simultanément. Le MNet proposé exploite la corrélation entre 40 attributs pour construire le groupement d'attributs.

Zu et al. définissent un réseau de neurones convolutifs appelé Multi-Label Convolutional Network ou MLCNN pour prédire les attributs des piétons [ZLLL17] tels que le sexe, les vêtements, la jupe, etc. À partir d'un jeu de données appelé PETA : (i) l'image entière est divisée en plusieurs segments de taille égale et intégrée dans le MLCNN ; (ii) chaque partie est filtrée indépendamment et entièrement connectée pour construire une représentation des caractéristiques. Dans cette méthode, le réseau de neurones multi-labels développé est constitué d'une combinaison de plusieurs fonctions de coût de classification d'attributs binaires. L'ensemble de données utilisé comprend 19 000 images de piétons, où chaque image est annotée avec un ou plusieurs attributs considérés. Les auteurs utilisent la fonction softmax pour chaque nœud afin de prédire chaque attribut. Ils calculent une fonction de coût qui regroupe tous les classificateurs d'attributs. Leurs expériences ont montré que la méthode proposée est considérablement plus performante que la méthode SVM proposée dans d'autres articles cités par les auteurs. Ils obtiennent un meilleur score de classification de 44 attributs sur 45 pour diverses métriques telles que la justesse de classification, les taux de rappel et les régions sous les courbes ROC (Areas Under Curves ou AUC).

Wosiak et al. étudient la classification multi-labels pour l'identification de la comorbidité, chez les patients affectés par plusieurs maladies [WGZ18]. Les auteurs proposent deux méthodes de classification, à savoir (i) la méthode Independent Label (IL) qui prédit tous les labels sans tenir compte des données de corrélation et (ii) la méthode Label Chain (LC) qui fait usage des dépendances tout au long du processus de classification. Leurs résultats expérimentaux sont évalués par plusieurs métriques (justesse, distance de hamming, précision, rappel, f1-score). Les résultats obtenus sont également consolidés par inférence statistique, dans trois ensembles de données médicales sur les enfants. Les classificateurs suivants ont été évalués pour les méthodes IL et LC : k-Nearest Neighbors, Naive Bayes, Support Vector Machine. Ils ont conclu que la méthode LC fournit de meilleurs résultats que la méthode IL pour toutes les mesures d'évaluation et d'inférence statistique.

En médecine, Maxwell [MLY⁺17] a exploité l'apprentissage profond pour identifier les patients atteints de plusieurs maladies en même temps. Il s'agit des maladies chroniques (telles que le diabète, l'hypertension artérielle, la stéatose hépatique, etc.) examinées sur 110 300 examens médicaux. Leur analyse a montré que l'apprentissage profond peut être utilisé pour apprendre les contributions de chaque caractéristique qui ont un impact sur le risque de maladie chronique d'un patient. Dans leur processus de classification, les auteurs ont mis en œuvre un réseau de neurones profonds basé sur la transformation des problèmes et l'adaptation des algorithmes. Les résultats sont comparés aux méthodes traditionnelles (RANdom k-labELsets ou RAKEL),

Multi-label k-Nearest Neighbors ou MLkNN à l'aide des métriques de classification : justesse, précision, rappel et F-score. Ils ont obtenu un score de précision de 92,07% pour le DNN contre 51,03% pour MLKNN ou 85,67% pour RAKEL.

D'autres cas d'applications de la classification multi-labels existent dans la littérature : la participation des membres à des communautés d'innovation en ligne [DCDR18], la classification des genres de films [WB17], la recherche visuelle à grande échelle [XFLH17] et bien d'autres [GV15, MGCV18].

Ce travail vise à améliorer les réseaux de neurones profonds pour une classification multi-labels des images. Cette préoccupation présente des similitudes avec certains des travaux brièvement décrits ci-dessus. Premièrement, nous partageons avec la catégorie de « transformation du problème », la nécessité d'aborder les difficultés liées à la transformation des données multi-labels. Deuxièmement, notre problème nécessite également « l'adaptation des algorithmes » apprentissage profond pour la tâche de classification. Par conséquent, nous avons conjointement engagé ces deux principes à l'aide des réseaux de neurones convolutifs pour construire une approche générique appelée, réseau de neurones multi-branches. Ce réseau offre deux cas d'usages de classification multi-labels : (i) réseau de neurones à sorties multiples incluant plusieurs branches pour prédire simultanément les étiquettes (labels) des données d'entrée et (ii) réseau de neurones à caractéristiques multiples qui renferment plusieurs branches de réseaux pré-entraînés pour extraire les caractéristiques des données d'entrée.

La réalisation d'une prédiction avec un réseau de neurones à multiples branches est un concept sous-jacent de l'apprentissage multitâche qui se construit par la formation de plusieurs tâches conjointes. En conséquence, nous présentons dans la section suivant le cadre de l'apprentissage multitâche.

3.3 Apprentissage multitâches

L'apprentissage multitâches [Car97] a connu une croissance rapide ces dernières années. Rich Caruana avait présenté dans ses travaux en 1997 comment tirer le meilleur parti de l'apprentissage multitâches en le combinant avec des méthodes classiques. Il a démontré que l'apprentissage multitâches améliore les performances des algorithmes d'apprentissage tels que les réseaux de neurones artificiels, le k plus proches voisins, la régression ou les arbres de décisions. Il a ensuite conclu qu'il y aura de nombreuses occasions de l'appliquer à des problèmes du monde réel. Les travaux de la dernière décennie de [CZ14, HDN19, LWL⁺19, PAW⁺17, TBSN12, YL18] sur la MTL montrent effectivement son efficacité pour résoudre des problèmes de classification ou de régression.

L'avantage du MTL est également avancé par rapport à l'apprentissage à tâche unique (STL). On constate que les modèles MTL sont efficaces grâce à l'utilisation d'une représentation partagée des connaissances entre les différentes sous-tâches [NGMN19]. Ce qui est appris par une tâche peut aider les autres tâches à mieux apprendre et à fournir un apprentissage plus robuste. Ainsi, la combinaison de toutes les tâches au sein d'un modèle unique peut améliorer la performance du modèle.

En raison des performances des réseaux neuronaux formés individuellement pour une seule tâche, les chercheurs ont porté leur attention sur la formation de réseaux capables de résoudre plusieurs tâches en même temps. Ainsi, l'apprentissage multitâche connaît une large applica-

bilité dans les problèmes du monde réel, mais elle nécessite la mise en œuvre de méthodes spécifiquement adaptées. Les sous-tâches ou tâches auxiliaires liées au problème traité doivent être considérées comme un tout. Leur exécution en silos (sans partage d'informations) peut provoquer un transfert de connaissance négatif entre les tâches. Le transfert négatif se traduit par la détérioration des performances du modèle d'apprentissage profond pendant l'apprentissage [ARS18]. Nous pouvons corriger ce problème à l'aide d'une technique de régularisation qui consiste à introduire des informations supplémentaires telles qu'une pénalité (par exemple, la régularisation L1 ou L2, la régularisation du dropout, la pondération des pertes) pour la prévention du sur- et du sous-ajustement [YXA18].

Un important cadre MTL sur l'apprentissage profond a été décrit dans [Rud17]. Le MTL est utilisé dans des situations où nous aimerions obtenir des prédictions de plusieurs tâches en même temps. Il s'agit d'un apprentissage conjoint avec des tâches auxiliaires, qui peut améliorer l'efficacité de la tâche principale. Durant la phase d'apprentissage, les informations et les connaissances sont partagées entre les tâches auxiliaires pour permettre une meilleure efficacité de généralisation de la tâche principale. Il est considéré comme une forme d'apprentissage par transfert inductif qui contient deux méthodes génériques dans l'apprentissage profond : le partage dur et souple des paramètres des couches cachées.

Le partage dur des paramètres : il soutient que des couches cachées sont partagées entre toutes les tâches et que certaines couches de sortie spécifiques aux tâches sont conservées. Afin de réduire le risque de sur-apprentissage le modèle est entraîné avec une représentation qui englobe toutes les tâches auxiliaires.

Le partage souple des paramètres : chaque tâche est définie indépendamment avec son propre modèle et ses propres paramètres dans un premier temps. Ensuite, des techniques de régularisation sont utilisées pour ajuster la distance entre les paramètres du modèle en fonction de leur similarité.

Dans ce travail, nous présentons les algorithmes de classification multi-labels en concevant les deux méthodes présentées, c'est-à-dire le partage dur et le partage souple des paramètres. Ces deux méthodes sont incorporées dans des réseaux de neurones convolutifs afin de comparer leurs performances de classification. Le partage dur a été utilisé sans en modifier les principes fondamentaux, tandis que dans le partage souple, nous avons procédé à un ajustement qui consiste à combiner des tâches auxiliaires au sein d'une couche cachée.

3.4 Méthodologie

3.4.1 Vue d'ensemble

L'apprentissage profond requiert une grande quantité de données pour entraîner un réseau de neurones profonds. La collecte de données multi-labels à grande échelle est une tâche complexe et fastidieuse. Il est donc possible de transposer les idées véhiculées par le principe "Diviser et Conquérir" dans le MTL pour la classification de données multi-labels. Ainsi, la résolution d'une tâche nécessite sa décomposition pour obtenir la résolution de plusieurs tâches auxiliaires dont leur exécution est imbriquée, faite séquentiellement ou parallèlement. Ces informations nous ont permis de construire deux architectures de réseaux de neurones profonds basées sur le partage dur (réseau de neurones à sorties multiples) et souple (réseau de neurones

à caractéristiques multiples) des caractéristiques de couches cachées. Le résultat attendu vise à améliorer la performance d'un modèle de classification multi-labels.

La première architecture appelée « réseau de neurones à sorties multiples » est formée d'un réseau pré-entraîné partagé entre plusieurs classificateurs binaires i.e autant d'étiquettes (labels) que de classificateurs binaires avec la taille de la sortie à 1. En second lieu, la deuxième architecture appelée « réseau de neurones à caractéristiques multiples », formée à partir de plusieurs réseaux pré-entraînés pour un classificateur dont la taille en sortie est égale au nombre d'étiquettes à prédire. Notre approche combine le MTL et transformation de données. Elle offre une forme d'apprentissage par transfert à travers le déploiement d'un ou plusieurs extracteurs de caractéristiques. Ainsi, nous pouvons nous servir des connaissances acquises par un réseau de neurones pré-entraînés lors des tâches de classification (Figure 3.1) :

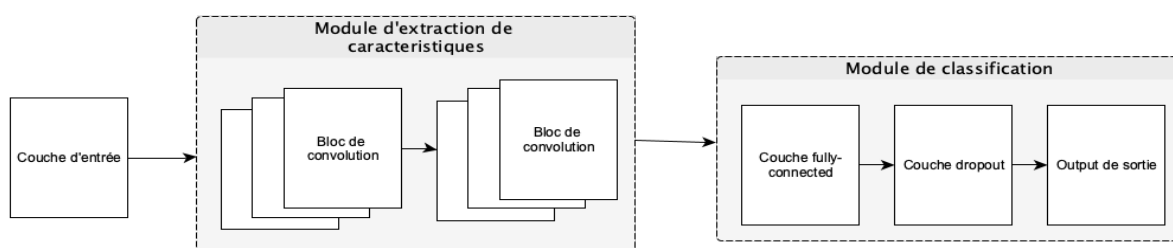


FIGURE 3.1 – Architecture de base d'un réseau de neurones convolutifs

Cette figure 3.1 montre séparément l'entrée du réseau, le module d'**extraction de caractéristiques** d'une entrée et le **module de classification** :

- Au niveau de la couche d'entrée, le type de données est une image multi-labels (une image appartenant à une ou plusieurs classes/catégories).
- Le module d'extraction de caractéristiques est constitué de couches d'entrée, de convolution, de mise en commun (dropout layer), entièrement connectées (fully-connected layer) et de sorties (output layer). Chaque image reçue en entrée sera filtrée, réduite et corrigée plusieurs fois par ce module. Son but est d'extraire les éléments caractéristiques de l'image (contours, coins, etc.).
- En général, le module de classification comprend les couches d'abandon (facultatives) et les couches entièrement connectées qui déterminent le lien entre la position des caractéristiques dans l'image. Ce type de couche reçoit un vecteur en entrée et produit un nouveau vecteur qui sera utilisé pour calculer la sortie du réseau de neurones. Pendant l'apprentissage par transfert, les couches du module de classification seront remplacées par une ou plusieurs nouvelles couches pour la classification des données du domaine cible.

3.4.2 Réseau de neurones profonds multi-branches pour les classifications multi-labels

Cette section explique en détail les architectures proposées. La tâche principale à résoudre est un problème de classification multi-labels. Les tâches auxiliaires (ou sous-tâches) définies constituent les branches (sous-tâches ou sous-réseaux) dans chacune des architectures. La combinaison de l'apprentissage par transfert, de l'apprentissage multitâches et de la transformation

des données multi-labels nous amène à proposer deux systèmes d'apprentissage alternatifs :

(i) un système appelé un « **réseau de neurones à sorties multiples (Neural Networks with multi-outputs)** » dont l'architecture est constituée d'un ensemble de classificateurs binaires disposés parallèlement.

(ii) un deuxième système appelé « **réseau de neurones à caractéristiques multiples (Neural Networks with multi-features)** » dans lequel plusieurs sous-réseaux de neurones convolutifs pré-entraînés parallèles sont utilisés par un classificateur multi-labels.

3.4.3 Réseau de neurones à sorties multiples

Tout type de réseaux de neurones peut être utilisé pour implémenter un réseau de neurones à sorties multiples. À cet égard, un principe invariable est le suivant : les couches inférieures du réseau sont partagées et les couches supérieures sont spécifiques aux tâches auxiliaires (Figure 3.2).

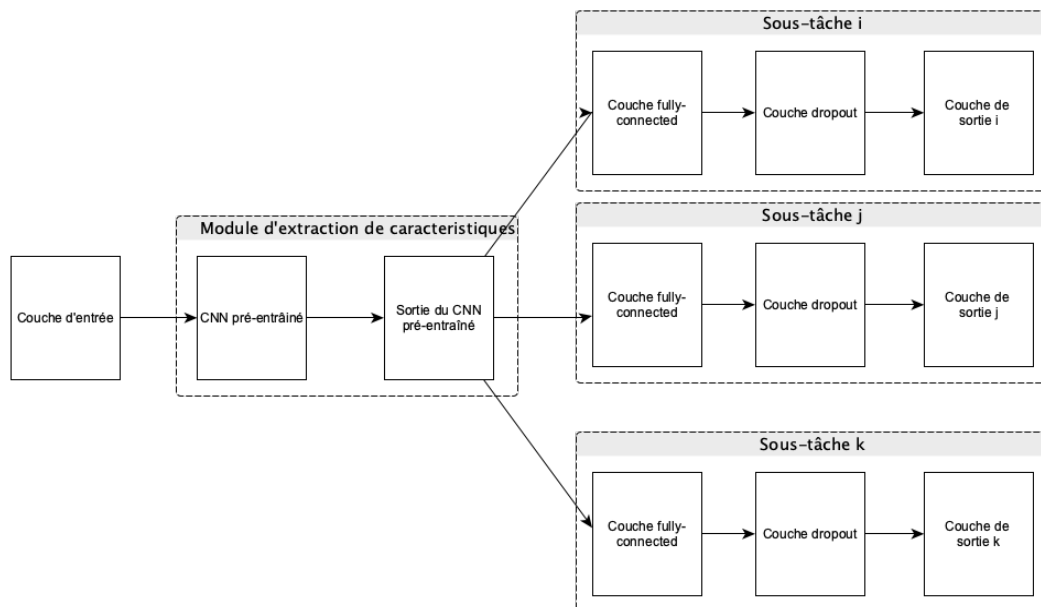


FIGURE 3.2 – Réseau de neurones à sorties multiples.

Les couches (cachées) partagées entre les sous-réseaux sont acquises pour le processus d'extraction de caractéristiques à partir d'un réseau pré-entraîné. Les tâches auxiliaires définies (décrivant des tâches spécifiques) sont formées comme suit :

- Si L est la taille de l'ensemble d'étiquettes, $L \geq 2$, nous aurons L sous-réseaux (c'est-à-dire L branches) constitués de plusieurs couches de neurones associés à L tâches auxiliaires.
- Chaque sous-réseau contribue à prédire une sortie j souhaitée liée à l'étiquette \widehat{Y}_j , $j = 1 \dots L$.

Lors de la conception du modèle de classification, les couches partagées sont étendues (comme dans l'Équation 3.1) par des couches supplémentaires, les couches entièrement connectées,

les couches d'abandon ou dropout et une couche d'activation dans toutes les branches (pour toutes les tâches auxiliaires ou spécifiques). La fonction d'activation ajoutée permet le passage ou non de l'information si le seuil de stimulation est atteint. Concrètement, son rôle est de décider si la réponse du neurone doit être activée ou non. Dans cette situation, les fonctions d'activation choisies sont employées pour produire la sortie de chaque branche.

$$join_{layer} = fuse(X, I_j) \quad j = 1 \dots L \quad (3.1)$$

où X désigne la sortie des couches partagées et I_j la couche d'entrée d'une branche j . La couche I_j est complétée par d'autres couches complètement connectées ou d'abandon (dropout).

Une fonction d'activation après la dernière couche complètement connectée est créée. Elle reçoit une valeur d'entrée, la transforme et retourne une valeur de probabilité qui sera utilisée pour associer un échantillon à une étiquette. Par conséquent, les étiquettes de vérité terrain sont déduites par cette fonction de seuillage (Équation 3.3). Deux options sont possibles pour établir ces étiquettes depuis l'entrée : (i) un seuil global est donné pour toutes les étiquettes ou (ii) un seuil spécifique est donné pour chaque étiquette.

La détermination du seuil global s'effectue en trois étapes (Équation 3.4) :

Étape 1 : Supposons que θ soit un ensemble avec une progression arithmétique ayant une séquence de nombres. Par conséquent θ_i est tel que la différence entre les termes consécutifs, donc θ_i est constant (représentée par Équation 3.2).

Par exemple, l'ensemble θ a des valeurs ajustées par unité de 10% (de 0% à 100%) avec $\theta_0 = 0$, $\theta_1 = 10\%$ et $\theta_n = 100\%$. Les opérations d'incrémentations peuvent être effectuées par une autre valeur r (différente de 10%), ainsi θ_i serait incrémenté de r , puis $\theta_{i+1} + r$ se voit attribuer la valeur de :

$$\begin{aligned} \theta_0 &= 0 \\ \theta_{n+1} &= \theta_n + r \end{aligned} \quad (3.2)$$

où les valeurs θ sont comprises entre 0 et 1 ; r est la différence commune (ou raison), dont les valeurs sont comprises entre 0% et 100%.

Étape 2 : Nous calculons successivement la valeur métrique choisie Mv pour tous les θ_i

Étape 3 : Nous recherchons la valeur θ_m procurant la valeur maximale de Mv

$$Mv = \begin{cases} 0, & x < \theta_m \\ 1, & x \geq \theta_m \end{cases} \quad (3.3)$$

où $x \in Mv$

$$argmax_{\theta} Mv = \{\theta_m \in \theta \mid \forall i, Mv(\theta_i) \leq Mv(\theta_m)\} \quad (3.4)$$

où θ_m est le seuil indiqué.

3.4.4 Réseau de neurones à caractéristiques multiples

Tout type de réseaux de neurones profonds peut être utilisé pour implémenter le réseau de neurones à caractéristiques multiples. Dans ce travail, nous avons considéré des réseaux de neurones convolutifs avec un principe cohérent qui est le suivant : les couches inférieures sont différenciées par l'utilisation de réseaux pré-entraînés distincts (sous-réseaux pré-entraînés), et les couches supérieures sont partagées au moyen d'un classificateur commun (Figure 3.3). Chaque sous-réseau pré-entraîné peut être ré-entraîné sur une ou plusieurs données des domaines sources avec différents modèles ou différentes approches d'apprentissage par transfert. Les sous-réseaux pré-entraînés sont étendus par l'ajout d'une ou plusieurs couches cachées et leurs sorties sont concaténées pour former une nouvelle couche nommée $join_{layer}$ (comme indiqué dans l'Équation 3.5). Cependant, il est possible d'utiliser d'autres opérateurs de jointure tels que l'addition, la soustraction, la multiplication, la moyenne, le maximum ou le minimum comme alternatives à la concaténation des couches. Dans ce dernier cas, sa sortie est complétée par une ou plusieurs couches supplémentaires entièrement connectées, une couche d'abandon et une dernière couche prenant en charge la fonction d'activation pour calculer les sorties prédites liées à l'étiquette $\hat{Y}_j, j = 1 \dots L$. Les étiquettes de vérité sont déduites par une fonction de seuillage (comme indiqué dans les Équations 3.2 - 3.4) déterminant l'appartenance d'un échantillon à une classe. À la fin de cette étape, le classificateur est entièrement spécifié dans le réseau de neurones avec des caractéristiques multiples. Le résultat est une architecture de réseau de neurones profonds conçue spécifiquement pour réaliser des prédictions très significatives dans la classification multi-labels.

$$join_{layer} = concatenate(x_j, j = 1..k) \quad (3.5)$$

où x_j représente les sorties des sous-réseaux pré-entraînés et k la taille du sous-réseau pré-entraîné utilisé.

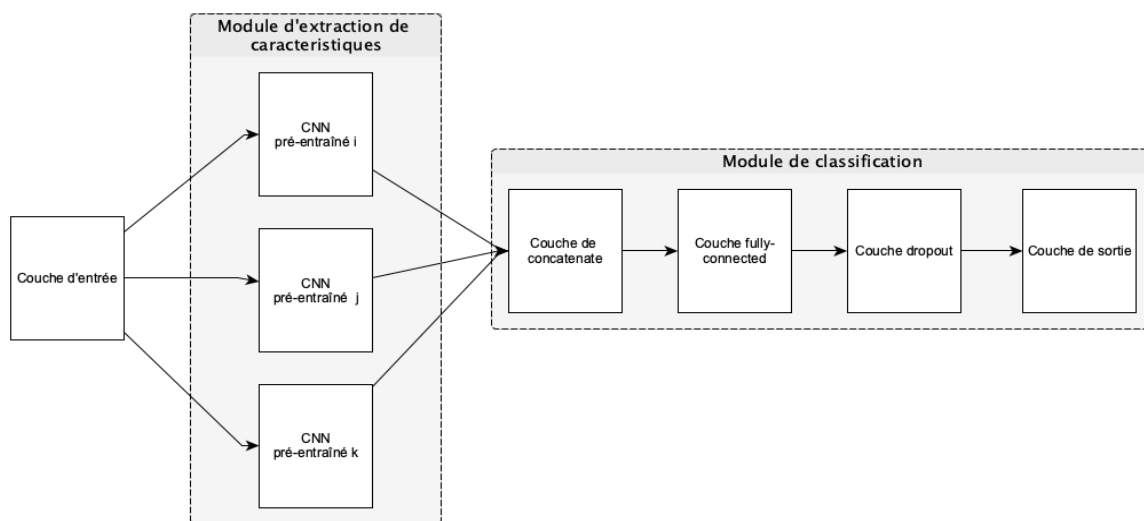


FIGURE 3.3 – Réseau de neurones à caractéristiques multiples.

3.4.5 Algorithme de réseau neurones profonds multi-branches

Dans cette sous-section, nous présentons les algorithmes de classification multi-labels et le processus d'apprentissage profond. La tâche de classification réalisée par le **réseau de neurones à caractéristiques multiples** est décrite par la Figure 3.4. Dans un premier temps, nous utilisons la méthode de transformation binarisée multi-labels dans le prétraitement des données. Cette transformation attribue à chaque échantillon un vecteur binaire $[y_1^j, y_2^j, y_i^j \dots y_L^j]$, $j = 1 \dots L, y \in Y$ indiquant la présence ou l'absence d'un label. Si $y_i^j = 1$ (respectivement $y_i^j = 0$) cela indique la présence du label i (respectivement l'absence du label i) dans l'échantillon x^j . La taille du vecteur binaire est égale au nombre d'étiquettes L .

Ensuite, des réseaux pré-entraînés sont mis en avant pour extraire les caractéristiques des données d'entrée. Dans l'étape suivante, l'opération de concaténation de ces modèles est effectuée en ajoutant une ou plusieurs couches entièrement connectées, des couches d'abandons et une fonction d'activation. Le modèle profond obtenu est utilisé pour les ensembles de données d'entraînement et de validation.

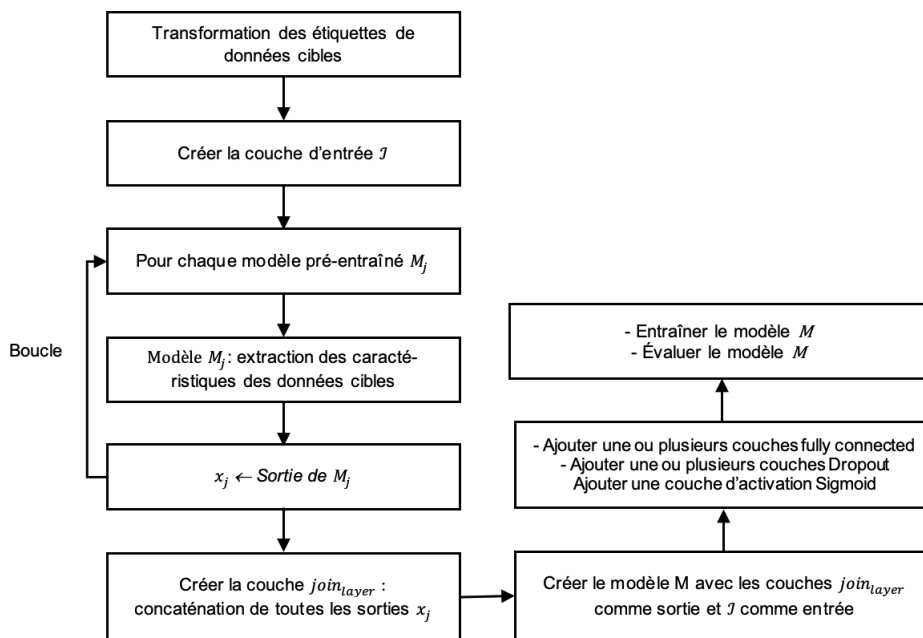


FIGURE 3.4 – Procédure de classification d'un réseau avec des caractéristiques multiples

La tâche de classification introduite avec le **réseau de neurones à sorties multiples** est décrite par la Figure 3.5. La méthode de transformation binaire multi-labels est employée pour le prétraitement des données dans un premier temps. Cette transformation attribue à chaque échantillon un vecteur binaire $[y_1^j, y_2^j, y_i^j, \dots, y_L^j]$, $j = 1 \dots L, y \in Y$ indiquant la présence ou l'absence d'une étiquette. Si $y_i^j = 1$ (respectivement $y_i^j = 0$) cela dénote la présence du label i (respectivement la non-présence du label i) dans l'échantillon x^j . Le nombre d'étiquettes L est égal à la taille du vecteur binaire.

Deuxièmement, les couches partagées (cachées) sont réalisées à partir d'un réseau pré-entraîné sur les domaines sources. Ensuite, nous étendons ce modèle pré-entraîné par une ou plusieurs couches complémentaires de la tâche spécifique. Les sorties des couches complémentaires indiquent la probabilité d'appartenir ou non à une étiquette dans l'étape suivante. Cependant, si

vous avez L classes, vous aurez L tâches spécifiques soit L sorties sont générées. Le modèle pré-entraîné et les couches additionnelles sont fusionnés pour former un réseau à sorties multiples. La définition du réseau peut être assez complexe lorsque nous avons un nombre élevé d'étiquettes. Une technique de simplification peut consister à établir des groupes d'étiquettes afin de réduire le nombre de tâches.

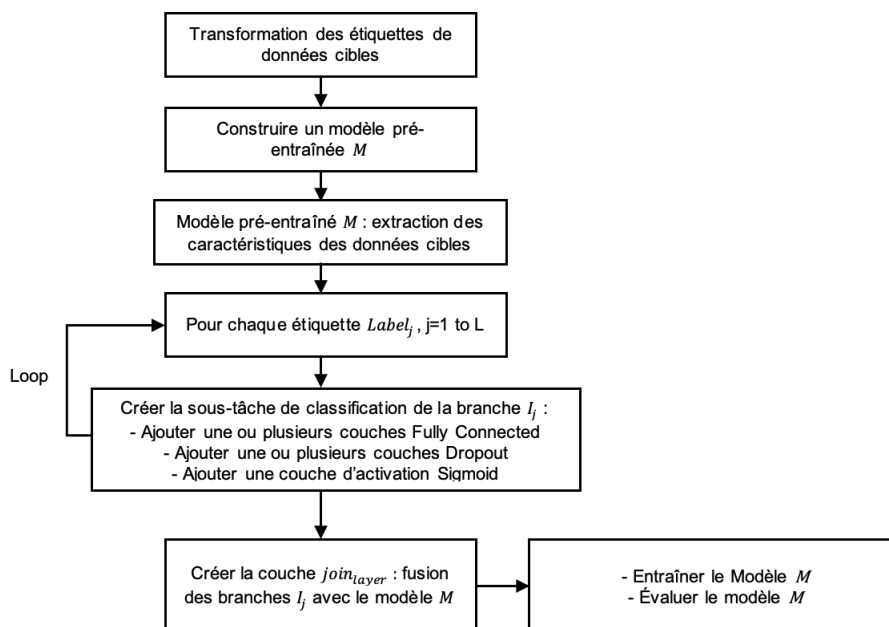


FIGURE 3.5 – Procédure de classification d'un réseau de neurones à sorties multiples

3.5 Expérimentons

3.5.1 Jeux de données

L'architecture proposée est entraînée, validée et évaluée sur les jeux de données Pascal Visual Object Classes (VOC) 2007 & 2012 [EVGW⁺07], NusWide [CTH⁺09] et Amazon rainforest [Pla17]. Tous les jeux de données sont décrits dans la section 1.6

3.5.2 Paramètres d'expérimentation

Nous validons les architectures proposées sur les réseaux pré-entraînés suivants : VGG16 [SZ14], ResNet [HZRS15] et InceptionV3 [SVI⁺15] qui sont définis dans la section 1.3.4. Ces architectures sont toutes initialisées sur le jeu de données ImageNet et elles ont été gagnantes au moins une fois dans la compétition Large Scale Visual Recognition Challenge (ILSVRC). Les résultats de sortie sont prédits par la fonction d'activation sigmoïde sur la dernière couche. L'arrêt du processus d'entraînement est contrôlé par la fonction d'arrêt précoce (early stopping), qui est principalement définie par les éléments suivants : (i) la valeur maximale ou minimale du paramètre à contrôler (ii) la valeur du nombre d'améliorations requises et (iii) la patience qui est la valeur conditionnant l'arrêt avec le nombre d'époques sans amélioration sur la "valeur du paramètre à contrôler". Les valeurs limites sont définies pour chaque paramètre

de la fonction d'arrêt précoce afin de contrôler l'arrêt du processus d'entraînement, utile pour éviter les problèmes de sur- et de sous-apprentissage. Cependant, le déséquilibre des données peut augmenter de manière significative le risque de problèmes de sur-apprentissage. Dans de tels cas, il est recommandable d'utiliser l'échantillonnage stratifié pour que les fréquences relatives des classes soient approximativement préservées dans chaque pli de formation durant l'entraînement.

Plusieurs autres éléments sont nécessaires pour optimiser l'apprentissage d'un modèle et le rendre efficace pour les tâches de classification ultérieures. Pour accomplir cette tâche d'optimisation, nous avons utilisé l'optimiseur appelé Adam¹. Il est également nécessaire de fournir d'autres moyens pour les tâches d'optimisation qui concernent les principaux hyper-paramètres, à savoir le taux d'apprentissage, les époques et la taille du lot. La Table 3.1 présente les valeurs des hyper-paramètres utilisés dans l'apprentissage des deux architectures proposées.

Hyper-paramètres	Value
Optimiseur	Adam
Taux d'apprentissage	1e-4
Époque	30
Taille du lot	32

TABLE 3.1 – Liste des hyper-paramètres

Le **taux d'apprentissage (Learning rate)** est un paramètre d'ajustement dans un algorithme d'optimisation qui détermine la taille du pas à chaque itération tout en se déplaçant vers un minimum de la fonction de perte. Pour mener à bien la phase d'entraînement, nous allons réaliser plusieurs **époques (Epoch)**, c'est-à-dire plusieurs passages complets de l'ensemble des données d'entraînement. Pour chaque donnée, nous allons réaliser une prédiction et calculer l'erreur pour ensuite procéder à la mise à jour des poids du réseau de neurones. Les époques servent à déterminer si la minimisation de la fonction d'erreur globale calculée sur l'ensemble des erreurs de prédictions est sur la bonne voie. La **taille du lot (Batch size)** fait référence au nombre d'exemples de données utilisés dans une itération. Un ensemble de données d'entraînement peut être divisé en un ou plusieurs lots.

3.5.3 Métriques d'évaluations

Les métriques standards comme la précision, le rappel, f1-score et la justesse sont adoptées pour mesurer la performance du classificateur. Les définitions de ces métriques sont fournies dans la section 1.3.6

3.5.4 Architecture CNN

La structure du réseau de neurones à sorties multiples est illustrée ci-dessous :

- La couche 1 est constituée des données d'entrée de 256x256 pixels.
- La couche 2 consiste en un réseau inception-v3 pré-entraîné dans lequel nous avons ajouté deux couches entièrement connectées de 4096 neurones. La couche d'abandon

1. Algorithme d'estimation du mouvement adaptatif (Adaptive Movement Estimation en anglais)

est insérée entre les couches entièrement connectées et possède une valeur de 0,3 pour l'expérimentation.

- La dernière couche est une couche entièrement connectée à une dimension (nœuds) et la fonction sigmoïde est ajoutée pour obtenir la probabilité que les étiquettes aient une valeur comprise entre 0 et 1.

La structure du réseau à caractéristiques multiples est décrite explicitement comme suit :

- La couche 1 est l'entrée du réseau, et elle possède 256x256 pixels.
- La couche 2 (respectivement la couche 3) forme la première (respectivement la deuxième) branche pour extraire les caractéristiques des données à l'aide du réseau pré-entraîné VGG16 (respectivement Inception-v3).
- La couche 4 est obtenue en concaténant les sorties des couches 2 et 3.
- Les trois dernières couches sont entièrement connectées avec 4096 neurones. Les étiquettes sont prédites avec la fonction d'activation sigmoïde qui donne le score des étiquettes ayant une valeur entre 0 et 1.
- Entre la dernière couche et les couches entièrement connectées, la couche d'abandon (dropout) est insérée avec une valeur de 0,3 pour l'expérimentation

3.6 Résultats

3.6.1 Evaluation de la performance prédictive

Nous opposons les modèles de classification basés sur **ResNet50**, **Inception-v3**, **VGG16**, **réseau de neurones à sorties multiples basé sur Inception-v3**, **réseau à caractéristiques multiples formé par VGG16 et Inception-v3** en termes de performance prédictive de classification pour montrer leurs efficacités. Tous les modèles sont implémentés dans le framework Keras [C⁺15]. Ils sont entraînés et testés avec 1 GPU Nvidia, 50G de mémoire fournis par le centre de Calcul CALMIP² en Midi-Pyrénées.

Les Tables 3.2 à 3.5 montrent les résultats expérimentaux des classificateurs pour les quatre (4) métriques d'évaluation choisies et quatre (4) ensembles de données. Les résultats collectés de toutes les observations sont comparés et interprétés à l'aide des tests statistiques. Les scores en gras indiquent les meilleurs résultats sur les tableaux de résultats.

3.6.2 Comparaison statistique des classificateurs

Pour discuter de la comparaison des performances des classificateurs sur plusieurs ensembles de données, des tests statistiques non-paramétriques sont nécessaires. L'auteur [Dem06] examine plusieurs tests statistiques tout en étudiant leur adéquation en termes de ce qu'ils mesurent réellement et leur sûreté sur des hypothèses émises sur les données :

- H_0 : hypothèse nulle, définie qu'il n'y a pas de différences entre la performance des classificateurs.

2. <https://www.calmip.univ-toulouse.fr>

Jeux de données	resnet50	inception-v3	vgg16	réseau neurones sorties multiples	réseau de neurones à caractéristiques multiples
Amazon rainforest	87.05	62.27	84.27	55.26	83.41
Voc2007	83.21	85.01	44.20	85.18	92.63
Voc2012	83.21	86.19	59.35	90.58	92.87
Nuswide	70.90	72.51	3.12	80.57	88.17

TABLE 3.2 – Comparaison des performances entre les différents classificateurs avec la mesure « précision ».

Jeux de données	resnet50	inception-v3	vgg16	réseau neurones sorties multiples	réseau de neurones à caractéristiques multiples
Amazon rainforest	81.72	70.51	83.10	80.32	83.19
Voc2007	68.27	66.06	19.89	90.29	92.38
Voc2012	68.27	63.33	28.39	91.06	92.31
Nuswide	59.10	53.03	0.02	90.3	91.67

TABLE 3.3 – Comparaison des performances entre les différents classificateurs avec la mesure « rappel ».

— $H1$: rejet de l’hypothèse nulle selon laquelle il existe au moins une différence statistiquement significative entre les classificateurs.

Ainsi, pour comparer plus de deux classificateurs sur plusieurs ensembles de données, le test de Friedman est employé pour analyser les différences significatives dans les performances de classificateurs. Ce test est défini comme suit (Équation 3.6) :

$$\chi_F^2 = \frac{12N}{k(k+1)} \left(\sum_j R_j^2 - \frac{k(k+1)^2}{4} \right) \quad (3.6)$$

où N est le nombre d’observations, k le nombre de classificateurs évalués et R_j est le rang moyen du j -th des k classificateurs définis comme ce qui suit (Équation 3.7) :

$$R_j = \frac{1}{N} \sum_j r_j^i \quad (3.7)$$

avec r_j^i , le rang du j -th des k classificateurs.

Pour toutes les observations réalisées, le test a fourni une valeur dénommée p -value de 0.004 (soit p -value < 0.05), alors l’hypothèse nulle $H0$ est rejetée au niveau de signification.

3.6. Résultats

Jeux de données	resnet50	inception-v3	vgg16	réseau de neurones à sorties multiples	réseau de neurones à caractéristiques multiples
Amazon rainforest	75.65	61.96	75.43	64.21	83.41
Voc2007	74.74	73.93	23.71	88.60	88.81
Voc2012	74.74	71.71	30.68	88.59	88.64
Nuswide	63.22	59.48	0.04	88.60	87.71

TABLE 3.4 – Comparaison des performances entre les différents classificateurs avec la mesure « f1-score ».

Jeux de données	resnet50	inception-v3	vgg16	réseau de neurones à sorties multiples	réseau de neurones à caractéristiques multiples
Amazon rainforest	75.97	20.98	91.61	68.14	90.03
Voc2007	59.67	59.26	21.17	68.40	73.64
Voc2012	59.67	57.54	24.73	69.25	78.57
Nuswide	46.64	43.56	8.37	64.62	69.25

TABLE 3.5 – Comparaison des performances entre les différents classificateurs avec la mesure « moyenne de précision ».

En conséquence le test post-hoc de **Nemenyi (Nemenyi post-hoc test)** qui compare tous les classificateurs entre eux, est appliqué avec le niveau de signification de 5% (le plus souvent défini). Ce niveau de signification est un seuil qui détermine si le résultat d’une étude peut être considéré comme statistiquement significatif après la réalisation des tests statistiques prévus. Ce test nous montre dans quelle mesure il existe une différence statistique significative entre les classificateurs ou non. Pour cela, un seuil de **différence critique (critical difference)** (Table 3.6) est calculé pour établir le niveau de signification de performance entre les classificateurs. La différence de critique (CD) est calculée comme suit par l’Équation 3.8 :

$$CD = q_{\alpha} \sqrt{\frac{k(k+1)}{6N}} \quad (3.8)$$

avec q_{α} la valeur critique du test post-hoc de Nemenyi.

Avec cinq classificateurs, la valeur de CD est de 1.53. Si la différence des rangs entre deux classificateurs est plus grande que cette valeur, la différence de performance est significative (Table 3.6).

À partir de la Table 3.6, les résultats montrent que les différences de rang entre « réseau de neurones à caractéristiques multiples » et « ResNet50 » (soit une valeur de 1,69) ; entre « réseau

Méthodes	Average rank	Critical Distance (CD)
ResNet50	3,0	
Inception-v3	4,0	
VGG16	4,25	1,53
Réseau de neurones à sorties multiples formé avec Inception-v3	2,4375	
Réseau de neurones à caractéristiques multiples formé avec VGG16 et Inception-v3	1,3125	

TABLE 3.6 – Rang de classement de tous les classificateurs.

de neurones à sorties multiples » et « VGG16 » (soit 1,81) sont supérieures à la valeur 1,53 de CD. En conséquence, ces résultats ont des différences statistiquement significatives. Les différences entre les rangs moyens comme « ResNet50 » et « Inception-v3 » ou « réseau de neurones à caractéristiques multiples » et « réseau de neurones à sorties multiples » qui font respectivement 1.0 et 1.12, tous plus petits que CD, ne sont pas statistiquement significatives. En conséquence, ces architectures ont donc des performances équivalentes.

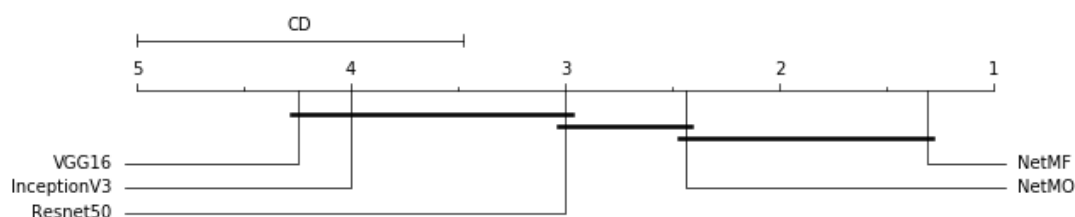


FIGURE 3.6 – Diagramme critique pour Nemenyi test. NetMF : réseau de neurones à caractéristiques multiples et NetMO : réseau de neurones à sorties multiples.

La Figure 3.6 présente le diagramme critique (Critical diagrams) qui donne une représentation visuelle de ce calcul de distance. Les chiffres sur l'axe horizontal indiquent la valeur moyenne du rang de chaque classificateur. La différence de performance n'est pas significative pour les classificateurs qui sont reliés par un trait noir. En conséquence, il n'y a aucune différence significative au sein des groupes suivants :

- NetMF, NetMO
- NetMO, Resnet50
- Resnet50, VGG16
- Resnet50, InceptionV3
- Inception-v3, VGG16

3.7 Discussions

3.7.1 Analyse technique

Généralement, dans un réseau de neurones profond, les premières couches apprennent les caractéristiques génériques des données d'apprentissage, tandis que les dernières couches apprennent leurs caractéristiques spécifiques. Plus précisément, dans notre étude, les observations suivantes ont été notées :

- Dans le partage souple (soft sharing), l'extraction des caractéristiques des données est augmentée (multipliée) par la présence de deux réseaux pré-entraînés.
- Cependant, dans le partage dur (hard sharing), ces caractéristiques sont obtenues par un seul réseau pré-entraîné.

De ce fait, nous observons une multiplication des possibilités de convergence vers de bonnes performances dans le partage souple par rapport au partage dur.

À première vue, le partage dur qui se présente sous forme de couches partagées entre plusieurs classificateurs est très spécialisé et devrait être efficace selon les travaux précédents décrits par [Rud17]. Cela se reflète dans chaque tâche de prédiction d'étiquettes qui est légèrement modélisée et paramétrée en fonction de la nature de la classification. Notre attente initiale était que le partage souple avec des réseaux à caractéristiques multiples produirait de meilleurs résultats. Ce réseau multiplie la probabilité de classification d'un label en combinant les résultats d'extraction des classificateurs. La combinaison de classificateurs augmente le nombre de caractéristiques pour représenter une image d'entrée, ce qui contribue à améliorer les performances du classificateur final.

3.7.2 Analyse statistique

Dans ce travail, nous avons comparé la performance de classifications de nos deux propositions (réseau à caractéristiques et à sorties multiples) aux méthodes de classification standards : ResNet50, Inception-v3 et VGG16. L'évaluation des performances de cinq méthodes de classification implique une comparaison détaillée de leurs caractéristiques. Les différents résultats expérimentaux limitent la possibilité de mieux comprendre les classificateurs avec leurs avantages et inconvénients. À cette fin, pour simplifier la comparaison, il est important d'envisager l'utilisation de mesures de performance appropriées et cohérentes :

- La comparaison des classificateurs sur la base d'indices de performance (Precision, Recall, F1-score et AP), le classement par rang (average ranking) et les tests statistiques sont également réalisés.
- Lors de l'analyse statistique, le test de Friedman suivi du test post-hoc de Nemenyi est la méthode recommandée pour comparer plusieurs classificateurs dans plusieurs jeux de données [Dem06].

D'une part, les tests statistiques sont nécessaires pour établir une certaine cohérence dans les expérimentations en indiquant l'importance des résultats issus de la comparaison des classificateurs. D'autre part, sur chaque jeu de données (Pascal VOC 2007, Pascal VOC 2012, Amazon Forest et NusWide), l'expérimentation a été menée dans des conditions de validation croisée à 10-plis sur chacun des jeux de données. Les classificateurs sont ensuite comparés statistiquement pour le test non-paramétrique de Friedman suivi du test post-hoc de Nemenyi pour toutes les observations. Le test de Nemenyi post-hoc est appliqué lorsque l'hypothèse nulle est

rejetée. La différence critique de ce test est utilisée comme seuil pour déterminer si les performances entre deux classificateurs sont significativement différentes. Le classement est établi en attribuant le rang le plus élevé au meilleur classificateur (Table 3.6).

3.7.3 Comparaison avec les algorithmes traditionnels et leurs explications

Les quatre points suivants vont décrire l'avantage de notre approche. Les classificateurs sont évalués sur quatre ensembles de données de référence. Les expériences statistiques montrent une amélioration notable de nos deux classificateurs par rapport aux divers classificateurs multi-labels simples.

1°) Le degré et le mode de contribution du traitement algorithmique à la classification

L'algorithme est générique au problème de la classification multi-labels. L'approche proposée n'est pas spécifique à un domaine et elle est utilisable dans de variétés de situations dites multi-tâches. Elle met en évidence une transformation de l'ensemble des étiquettes en un vecteur binaire pour une meilleure classification. Cependant, l'approche adoptée est essentiellement liée à la génération de plusieurs représentations potentielles (et leurs transformations) de la même information par différents réseaux. De ce fait, l'opération fondamentale effectuée sur les couches de sortie du module de convolution est celle de la concaténation de matrices (processus consistant à joindre une ou plusieurs matrices pour en faire une nouvelle). Toutefois, d'autres opérations telles que l'addition ou la soustraction de valeurs de neurones sont disponibles en option.

2°) Les données traitées et leurs sources

Pour chaque expérimentation d'apprentissage profond, la qualité et le volume des données d'entraînement et de test peuvent contribuer à améliorer la performance du modèle de classification. Certes, nous n'avons pas de préférence pour les données utilisées (Pascal VOC, Amazon rainforest, et NusWide) qui ne sont pas de même nature (section 1.6). Dans cette étude, c'est leur caractère multi-labels qui a été étudié. Néanmoins, les techniques d'apprentissage par transfert et de l'augmentation des volumes des données sources ont été d'une grande aide pour l'entraînement des modèles.

3°) Condition ou configuration expérimentale

Les paramètres d'apprentissage habituellement mentionnés sont l'optimiseur stochastique adaptatif (Adam), la descente de gradient stochastique, la propagation quadratique moyenne, etc. avec la valeur de leurs arguments comme le taux d'apprentissage, la taille du lot (nombre d'exemples d'apprentissage utilisés dans une itération); l'époque (nombre de passages du modèle dans le jeu de données). À cela s'ajoute la proportion d'échantillons de données d'apprentissage et de validation.

La conception de l'architecture du réseau joue également un rôle essentiel. Elle peut être difficile à maîtriser, d'où l'utilisation de standards comme VGGNet, GoogleNet ou Inception, ResNet, etc. De plus, un plan d'exécution peut être généré sur ces paramètres de traitement génériques en combinant les valeurs de leurs arguments. Cependant, il est très coûteux d'exécuter le même programme pour chaque optimiseur, pour chaque taux d'apprentissage, pour chaque taille de lot, etc.

3.8 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle approche de classification multi-labels. Les travaux ainsi décrits définissent un cadre adaptatif des réseaux de neurones profonds par le biais d'un apprentissage multi-tâches appelé « réseau de neurones profonds à plusieurs branches ». Ce cadre comprend un mécanisme de partage dur et souple des paramètres du réseau de neurones profonds. Les réseaux de neurones à plusieurs branches sont capables d'améliorer les performances des classificateurs multi-labels avec les notions de réseaux de neurones convolutifs à sorties multiples et caractéristiques multiples. Dans les deux cas considérés, les réseaux pré-entraînés ont servi à construire des tâches d'apprentissage. Les expériences ont été réalisées sur quatre jeux de données constitués d'images à étiquettes multiples. Ils ont montré que le partage souple des paramètres dans le cas d'un réseau à caractéristiques multiples fournit un meilleur résultat de la classification des images par rapport au partage dur des paramètres pour un réseau à sorties multiples. Les résultats obtenus nous permettent de conclure que les classificateurs conçus ont une performance comparable à celle des autres classificateurs standards en apprentissage profond. Pour une analyse significative des résultats, le test de Friedman suivi du test post-hoc de Nemenyi a été exploré avec un niveau de signification de 5%, pour la comparaison de plusieurs classificateurs. La visualisation des résultats de test est établie à l'aide de diagramme critique de différence.

L'architecture d'un réseau de neurones profond est un élément essentiel pour obtenir plus de performances désirées durant la phase d'entraînement. Par conséquent, l'application de différentes architectures ne conduit pas nécessairement à des résultats de même précision. Damodaran et ses collègues ont montré que les techniques de régularisation telles que l'initialisation aléatoire des poids, les critères d'abandon ou d'arrêt précoce doivent être évaluées avec soin et leurs valeurs comparées à des seuils requis [DFSC19]. Les hyper-paramètres que nous utilisons pendant l'expérimentation sont liés à la nature des données d'entraînement et ils vont influencer sur le temps du processus d'apprentissage. Certains hyper-paramètres peuvent être plus robustes que d'autres en fonction des applications considérées. Trouver le bon équilibre en ajustant leur valeur nécessite beaucoup d'expérimentations afin de se conformer aux exigences explicites dans une grande variété de domaines d'application. En conséquence, des modifications de ces paramètres sont possibles à l'avenir.

Des progrès considérables ont été accomplis dans la reconnaissance d'objets avec l'apprentissage profond, grâce à la disponibilité d'un ensemble de données annoté à grande échelle. Cependant, la taille limitée de l'ensemble de données devient un handicap pour former un réseau de neurones profonds. Par conséquent, le prochain chapitre étudie l'efficacité de l'apprentissage par transfert sur un tel ensemble de données pour la classification de la maladie du mildiou du mil.

Chapitre 4

Classification des maladies du mil à l'aide d'un réseau de neurones convolutifs profonds et de l'apprentissage par transfert

4.1 Introduction

Le mil est l'une des cultures vivrières les plus cultivées dans les régions d'Afrique et d'Asie. L'Inde, le Niger, la Chine, le Mali ont été les plus gros producteurs mondiaux en 2019¹. La présence des maladies au niveau de cette culture conduit à une perte au niveau de sa production. Les agriculteurs (avec les phytopathologistes) cherchent à lutter contre les maladies des plantes depuis des siècles. Ces maladies sont généralement causées par des virus, des bactéries ou des champignons qui apparaissent dans certaines conditions climatiques ou environnementales comme la dégradation des sols, les problèmes hydriques, le dérèglement climatique. À la base, elles sont sources de pertes économiques considérables pour les agriculteurs et menacent l'autosuffisance alimentaire de la population. Pour résoudre ces problèmes, il est nécessaire de détecter les maladies des plantes par des techniques avancées et intelligentes.

Pour relever ce défi, les chercheurs en relation avec les coopératives agricoles ont exploité les potentialités de l'intelligence artificielle (IA) dans le domaine de l'agriculture. Leurs rôles consistent à suivre l'état de santé des cultures pour prévenir la propagation des maladies. Une approche innovante de l'IA, l'apprentissage profond a montré un résultat prometteur, en particulier pour la vision par ordinateur en employant les réseaux neurones convolutifs (CNN). Ils offrent une nouvelle voie de détection basée sur la reconnaissance automatique des formes. À travers un CNN, il est possible d'analyser, d'interpréter et de visualiser une image afin d'obtenir des résultats significatifs (comme les motifs). Ainsi une analyse des images des cultures ou des parcelles à l'aide des CNN va permettre de générer un support de connaissances pour aider les agriculteurs à prendre des décisions grâce à l'exécution des systèmes d'aide à la décision. En conséquence, les agriculteurs peuvent choisir des méthodes appropriées de protections pertinentes des cultures et innover leur système de production.

Récemment, plusieurs applications de l'apprentissage profond ont trouvé des solutions à de

1. <http://fao.org/foostat>

nombreux problèmes de reconnaissance d'images et ont obtenu les meilleurs résultats dans de nombreux domaines de recherche, tels que les détections des maladies végétales [FYLP18, LYZ⁺17, MHS16] ou le diagnostic médical [CLX18, DCP⁺19, WRH17] et bien d'autres. D'une manière générale la motivation principale de ce travail consiste à identifier la maladie mildiou chez le mil. Cette maladie se caractérise par la transformation partielle ou totale de l'épi ou la feuille avec l'apparition des tâches brunâtre. Par la suite, les tâches vont s'étendre et provoquer la mort de la plante si elles n'ont pas été enrayerées rapidement. La recherche des voies et moyens pour diminuer ou remédier aux problèmes causés par le mildiou fait l'objet de cette étude. Pour ce faire, nous présentons le développement d'un modèle en utilisant une approche d'apprentissage profond, en particulier les réseaux de neurones convolutifs en utilisant l'image de la culture du mil. L'approche proposée dans ce travail applique la technique d'apprentissage par transfert pour compenser le manque de données d'entraînement et comprend trois étapes principales : (i) l'acquisition d'images (ii) la définition d'un réseau de neurones composé d'un "modèle pré-entraîné" pour une "extraction de caractéristiques" et (iii) la classification de la maladie du mildiou.

Le classificateur réalisé est basé sur le modèle VGG16 pré-entraîné ImageNet comme jeu de données sources. Nous avons un petit jeu de données de 126 images pour les catégories malades et saines pour identifier ou pas le mildiou. La tâche de classification à l'aide de VGG16 fournit les résultats suivants : (i) le classificateur des images non étiquetées est construit (ii) la méthode proposée a été utilisée pour l'extraction de caractéristiques afin d'identifier ou pas la maladie dans l'image d'entrée.

Ce chapitre est structuré comme suit : les travaux connexes sont présentés dans la section 4.2, la section 4.3 décrit la méthodologie adoptée. Dans la section 4.4, le contexte expérimental et les résultats de test sont présentés et la dernière section 4.5 conclut le travail et présente les recherches futures.

4.2 Travaux connexes

L'analyse d'images est un domaine de recherche important dans le secteur agricole et elle couvre plusieurs problèmes [ASM17, ASB20, AKS⁺19, BCSB20, CCZ⁺20, DKC⁺21, MHS16, SSC17] : reconnaissance d'images, classification d'images, détection d'anomalies, etc.

Cheng et al. ont développé un réseau de neurones convolutifs (CNN) combiné à des réseaux de neurones résiduels pour identifier les parasites qui menacent la croissance des cultures et le stockage des produits agricoles [CZC⁺17]. En entrée du réseau, une image du parasite active la première couche convolutive et le système de pooling, et permet ainsi de séparer le parasite du reste de l'image. Dix catégories de ravageurs ont été identifiées par le réseau de neurones.

Grinblat et al. proposent d'utiliser les CNN pour le problème de l'identification des plantes à partir de leur veine morphologique [GULG16]. Les plantes à classer sont trois (3) espèces de légumineuses différentes à savoir, le haricot blanc, le haricot rouge et le soja. Des modèles CNN de deux couches à six couches ont été construits. La justesse du modèle atteint une valeur moyenne de 96,9% avec un CNN à cinq couches contre trois algorithmes d'apprentissage automatique : machines à vecteurs de support (SVM), analyse discriminante pénalisée (PDA) et forêts aléatoires (RF).

Le riz est l'une des cultures vivrières les plus importantes au monde. Ses maladies ont un

effet dévastateur sur sa production et constituent également une grande menace pour la sécurité alimentaire. Lu et ses collègues [LYZ⁺17] ont identifié les maladies du riz à l'aide d'un CNN. L'étude a catégorisé dix (10) classes de maladies pour un total de 500 images de feuilles et de tige de riz. L'expérience a montré que le CNN donne un meilleur résultat comparativement aux modèles d'apprentissages automatiques conventionnels qui utilisent des techniques de bases de la reconnaissance des formes. Dans le cadre d'une stratégie de validation croisée à 10 plis, le modèle CNN proposé atteint une justesse de 95,48 %.

Shiqi Yu et al. proposent un modèle CNN capable de classer une image hyperspectrale, ce qui est une bonne pratique pour l'agriculture de précision ou la surveillance de l'environnement [SSC17]. Dans le cadre d'un apprentissage supervisé, ils définissent un CNN qui a appris à partir des images étiquetées (bâtiment, colline, pâturage, etc.). L'expérimentation fournit un résultat de 81,75 % de justesse par rapport aux méthodes traditionnelles d'apprentissage automatique comme les K plus proches voisins et la machine à vecteur de support.

Amara et al. réalisent une approche qui automatise le processus de classification des maladies de bananiers [ABA17] grâce à l'apprentissage profond. Les auteurs déploient l'architecture LeNet comme réseau de neurones convolutifs profonds pour classer les maladies sigatoka et speckle du bananier causées par des champignons déposés sur les feuilles. Sur deux ensembles de données d'images (en couleur, noir et blanc), ils obtiennent 98,61% de justesse avec les images en couleur et 94,44% pour les images en noir et blanc.

Les travaux de [RBM⁺17] évaluent la pertinence de l'apprentissage par transfert à partir d'un modèle de réseaux de neurones convolutif profonds sur les ensembles de données d'images de manioc. Leur modèle a identifié trois maladies et deux types de dommages causés par les ravageurs avec les résultats qui suivent : 98% pour la tache brune de la feuille, 96% pour les dommages causés par les acariens rouges, 95% pour les dommages causés par les acariens verts (GMD), 98% pour la maladie des stries brunes du manioc et 96% pour la maladie de la mosaïque du manioc (cassava mosaic disease ou CMD).

Rangarajan et al. [RPR18] ont utilisé deux modèles pré-entraînés, AlexNet et VGG16 pour classer des feuilles de tomates (six maladies et une classe saine) obtenues à partir du jeu de données PlantVillage [HS15]. La justesse de la classification a été de 99,24% pour VGG16 et de 96,51% pour AlexNet.

La détection des maladies des plantes à l'aide des réseaux de neurones convolutifs fournit une meilleure précision dans la tâche classification par rapport aux techniques traditionnelles de traitement d'images ou celles de l'apprentissage automatique conventionnel comme la machine à vecteurs de support.

4.3 Méthodologie proposée

La méthodologie proposée exploite l'apprentissage par transfert pour la détection ou l'identification de la maladie de mildiou chez le mil. Elle permet le développement d'un modèle de réseaux de neurones convolutifs basé sur VGG16. Nous disposons d'un petit jeu de données alors que l'apprentissage profond requiert beaucoup de données pour construire un classificateur efficace. En conséquence, l'apprentissage par transfert (section 1.5.2) combiné à l'augmentation de données (section 1.5.1) constituent un moyen qui peut augmenter l'efficacité du modèle d'apprentissage. L'augmentation de données va enrichir et étendre notre jeu de données en créant des variations d'une même image par des transformations mathématiques (rotation, translation, etc.). Ces transformations permettent d'éviter de plus le surapprentissage et d'aug-

menter la capacité de prédiction du réseau de neurones.

4.3.1 Vue d'ensemble du système

La Figure 4.1 donne une vue d'ensemble de notre système d'identification de la maladie mildiou. Il contient trois composants qui sont décrits comme suit :

(i) **Acquisition d'images** : nous avons créé notre propre ensemble de données étiquetées à partir d'images téléchargées depuis l'Internet ou capturées à partir de dispositifs numériques sur le terrain. Ces images constituent l'ensemble de données cibles dans le cadre de l'apprentissage par transfert. Elles se composent en deux catégories d'images : saines et malades et seront transmises au réseau de neurones pour l'entraînement.

(ii) **Entraînement du réseau de neurones convolutifs** : ce bloc comporte un modèle CNN pré-entraîné nécessaire à l'extraction de caractéristiques des images et la classification de nos images. Ce modèle pré-entraîné est un modèle de sauvegarde qui est déjà entraîné et testé un autre ensemble des données d'images. Il est ensuite modifié et entraîné sur la base de nos images.

(iii) **Classification des maladies** : à ce stade, nous avons notre classificateur déjà prêt. Cette dernière étape permet la prédiction d'une image d'entrée non vue par modèle. Elle consiste à déterminer la présence ou non de la maladie sur cette image. Le résultat de la classification (la sortie) représente les probabilités de la classe trouvée. Le détail du système implémenté est présenté dans la section suivante.

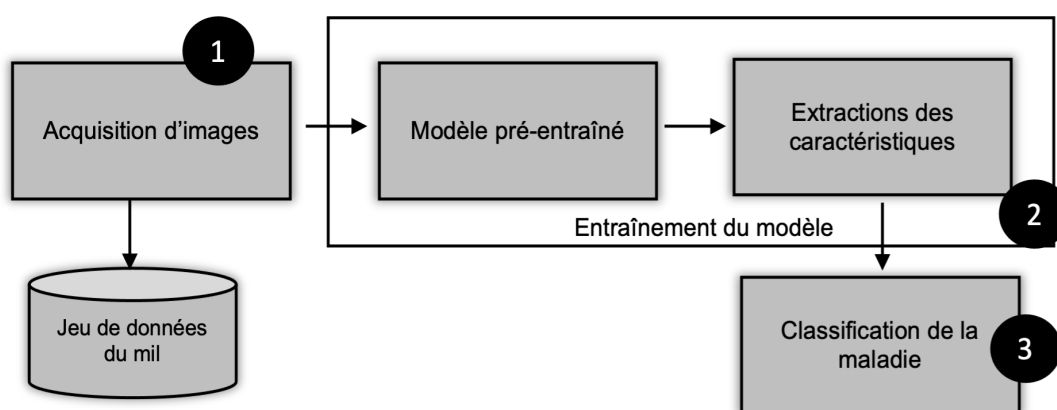


FIGURE 4.1 – Vue d'ensemble de la méthodologie proposée [CKKT19].

4.3.2 Approche basée sur l'extraction de caractéristiques

Notre système d'apprentissage et la classification de la maladie du mildiou mis en œuvre dans le réseau de neurones convolutif sont illustrés par la Figure 4.2. Le CNN mis en place utilise l'architecture de VGG16 dont le modèle a été entraîné sur le jeu de données ImageNet qui contient 1000 classes d'objets. Ce jeu de données contient des images plutôt génériques et très peu d'images sur les attaques fongiques du mil comme le mildiou. Un réseau de neurones pré-entraîné sur cette base serait a priori moins performant pour le problème d'identification de la maladie mildiou. Par conséquent, une adaptation du réseau de neurones est nécessaire pour

obtenir un résultat escompté. Nous proposons donc une architecture modifiée pour offrir une adaptation cohérente.

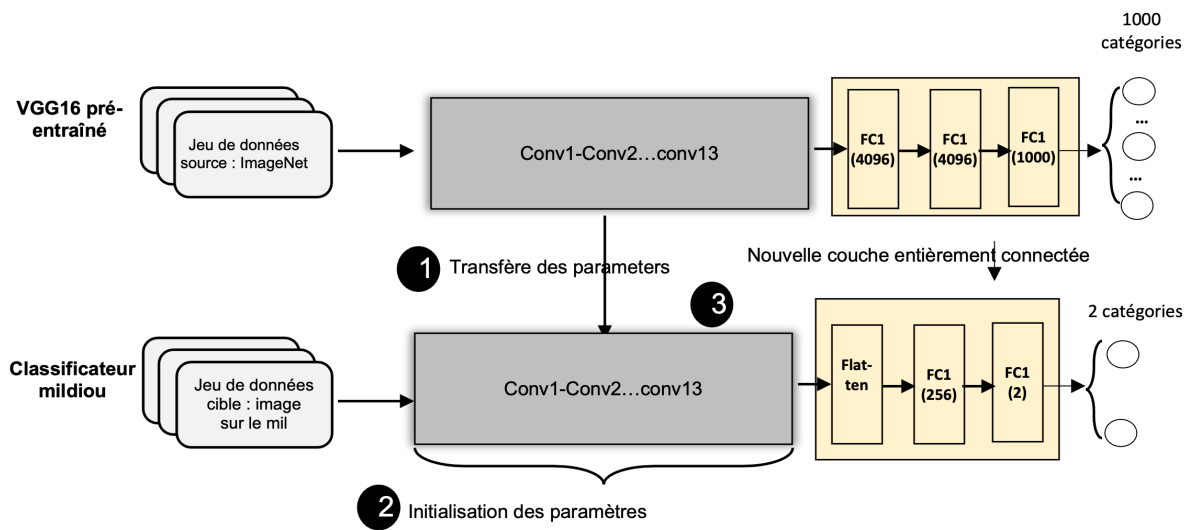


FIGURE 4.2 – Approche proposée pour l'extraction de caractéristiques [CKKT19].

Ce changement permet de modifier la classification des 1 000 classes d'ImageNet en une classification de deux (2) classes qui consiste à déterminer comme sorties la présence ou l'absence de mildiou. Ainsi, le nouveau réseau est obtenu en appliquant les étapes suivantes :

1. La première étape permet d'initialiser les paramètres du réseau pour la deuxième étape pour former une bonne performance en reconnaissance et classification d'images. Les dernières couches entièrement connectées de VGG16 sont supprimées, car elles ont été utilisées pour classer les 1 000 classes d'ImageNet.
2. L'adaptation est effectuée afin de résoudre le nouveau problème de classification en gelant tous les poids des couches pré-entraînées.
3. L'extension du VGG16 est faite sur le modèle en ajoutant deux couches entièrement connectées et une couche pour la sortie. La couche de la sortie donne la classification pour nos deux classes : présence ou absence du mildiou. Ainsi, toutes les couches du nouveau modèle sont entraînées sur les nouvelles images (ensemble de données cibles) et la fonction d'activation sigmoïde pour trouver la probabilité que les données appartiennent à une classe.

La section suivante décrit le processus expérimental pour valider notre approche.

4.4 Expérimentation et résultats

4.4.1 Détails de l'entraînement

Notre ensemble de données est constitué de 126 images saines et malades. Dans ce travail, cet ensemble de données a été divisé en un ensemble d'entraînement, de validation et de test non vus par le modèle d'apprentissage. Sa description est donnée dans la section 1.6. Dans la littérature, plusieurs subdivisions du jeu de données sont citées dans la littérature. Par exemple,

les 2/3 (respectivement 80%) de la taille du jeu de données sont utilisés pour l'entraînement et les 1/3 (respectivement 20%) pour la validation.

	Entraînement/ Validation	Test
Malades	70	18
Saines	29	9
Total	99	27

TABLE 4.1 – Ensemble de données

Nous retiendrons dans ce travail, pour les données d'entraînement et de validation 80% des images (soit 99 images) et les 20% des images restantes (soit 27 images) sont utilisés pour le test (Table 4.1). Pour mesurer la performance du classificateur, plusieurs subdivisions ont été appliquées : 80/20 (respectivement 70/30, et 60/40) signifie 80% (respectivement 70%, et 60%) du jeu de données d'entraînement et 20% (respectivement 30%, et 20%) du jeu de données de validation.

Les images en entrée du réseau sont redimensionnées en 150x150 pixels pour extraire les caractéristiques importantes. Le processus d'optimisation choisi est la descente de gradient stochastique (SGD) qui fournit une approche itérative pour minimiser l'erreur d'apprentissage. La Table 4.2 résume les paramètres d'apprentissage initiaux de la phase d'entraînement.

4.4.2 Résultats de l'expérimentation

Durant l'expérimentation, l'ensemble des programmes a été développé avec le framework Keras [C⁺ 15]. Ils sont exécutés sur un ordinateur portable Intel Core i5 avec 2,7 GHz et 8 Go de mémoire RAM. À l'aide des mesures d'évaluation, la meilleure performance du classificateur sur l'ensemble de test a été obtenue par la configuration 80/20. L'ensemble des résultats sont présentés dans la Table 4.3 en pourcentage (%).

Pendant la phase d'entraînement, l'arrêt anticipé (voir la sous-section 3.5.2) est introduit pour éviter le surapprentissage et aider le modèle à trouver la valeur de convergence pour la perte de validation la plus faible possible. En pratique, cela se fait par le biais de la patience qui définit le nombre maximum d'époques consécutives du modèle sur l'ensemble de validation sans amélioration et qui est tolérée avant l'arrêt de l'entraînement. Cette valeur est obtenue à partir de la 30^{ème} époques sur les 60 définie avec une plus petite perte de validation de 19%. Pour un utilisateur agricole, la méthodologie d'apprentissage automatique réalisée pour classer la maladie du mildiou peut être déployée sur un dispositif mobile. Le système d'aide est ainsi validé par des tests sur le terrain et le retour d'information de l'utilisateur agricole.

Paramètres	Valeur
Taux d'apprentissage (Learning rate)	1e-4
Moment (Momentum)	0.9
Epoque	100

TABLE 4.2 – Les paramètres d'entraînement (hyper-paramètres)

4.5. Conclusion

Configuration	Accuracy	Précision	Rappel	F1-score
80/20	95.00	94.50	90.50	91.75
70/30	93.50%	91.00	87.50	88.66
60/40	94.00%	93.00	88.00	89.67

TABLE 4.3 – Résultats d’expérimentations

4.5 Conclusion

Dans cette étude, nous avons introduit un réseau de neurones convolutif pré-entraîné VGG16 pour l’identification de la maladie du mildiou chez le mil. Nous avons montré l’efficacité de l’apprentissage par transfert sur un petit jeu de données. L’architecture pré-entraînée VGG16 a été modifiée sur ses couches hautes pour apprendre nos données et effectuer une classification binaire. La performance observée du modèle durant les expérimentations atteint une justesse de 95%.

Dans les perspectives de ce travail, notre objectif est d’identifier la maladie mildiou observable chez diverses cultures comme le coton ou la pomme de terre. Notre solution peut s’intégrer à un dispositif embarqué (appareil photo numérique, un téléphone intelligent, drone) pour aider les agriculteurs à identifier les maladies des plantes. De plus, nous aimerons utiliser d’autres réseaux de neurones pour affiner et réduire le taux d’erreur du classificateur.

Il est communément admis que les réseaux de neurones profonds entrent dans la catégorie des modèles à boîtes noires. Cependant, grâce à une technique de visualisation de l’image, il est possible de sélectionner la zone infectée et voir l’étendue. Cette information peut aider un agriculteur à comprendre et communiquer aux autres la prise de décision d’un modèle profond pour la classification effectuée. Ce point, nommé l’explicabilité de l’intelligence artificielle sera étudié dans le chapitre suivant. À ce niveau, nous explorons une autre application de l’apprentissage profond en agriculture, en particulier l’identification et la localisation des insectes parasites.

Chapitre 5

Réseau de neurones convolutif explicable pour la reconnaissance des insectes nuisibles

5.1 Introduction

Les attaques fongiques des cultures représentent une menace pour la sécurité alimentaire, car elles réduisent considérablement le rendement et la qualité des cultures. Déterminer les causes de leur apparition permet une amélioration des performances des systèmes agricoles et alimentaires. Une attention particulière est accordée aux insectes ravageurs comme un cas responsable de ces attaques. La détection, l'identification et la localisation efficaces des ravageurs en temps opportun constituent de véritables défis dans le contexte de l'agriculture de précision. Résoudre ces défis permettrait de prendre des mesures appropriées pour lutter contre les dommages causés tout en contribuant à la réduction du taux de pesticides utilisés.

Nous assistons aujourd'hui aux différents succès de l'intelligence artificielle pour présenter des avancées dans diverses tâches automatiques. Son sous-domaine qui est l'apprentissage profond a été très bénéfique et fait l'objet de plusieurs projets de recherches. Les tâches de reconnaissance des formes, de traitement automatique des langues, de reconnaissance vocale ou de diagnostic médical sont largement influencées par ses progrès. Grâce aux réseaux de neurones profonds dits convolutifs, le développement des systèmes intelligents a permis d'améliorer la précision des modèles de classification d'images dans tous les domaines y compris l'agriculture sur son volet gestion de la santé des cultures.

Les réseaux de neurones convolutifs se composent principalement de couches convolutives et d'autres couches de traitement qui extraient des cartes de caractéristiques sur un ensemble d'images. Ensuite, des catégories d'étiquettes sont exprimées en sortie du réseau pour désigner le résultat de la classification. Ces couches renferment un nombre important de paramètres d'apprentissage et une organisation hiérarchique qui sont exploités par les algorithmes d'apprentissage. Il est difficile, voire compliqué d'expliquer ou de comprendre les décisions prises par les modèles d'apprentissage pour obtenir les résultats. Malgré leurs résultats satisfaisants des réseaux de neurones profonds lors des tâches de classification des images, l'un des inconvénients est la difficulté de comprendre ce que le réseau a appris. En conséquence, ils sont considérés comme des boîtes noires (avec très peu de transparence) dans la mesure où les différentes couches qui les composent peuvent être difficiles à déchiffrer. De ce fait, il est

souvent difficile et compliqué d’expliquer ou de comprendre les décisions qu’ils prennent pour prédire les sorties. Par conséquent, si nous comprenions ce que le modèle a appris, il est possible de garantir la qualité des résultats obtenus. Cette capacité est cruciale et importante pour les utilisateurs dans les domaines comme l’imagerie médicale ou le suivi de la santé des exploitations agricoles.

À présent, nous assistons à une croissance des exigences utilisateurs dans la transparence des algorithmes de l’apprentissage profond. En premier lieu, il s’agira d’offrir une compréhension du processus d’extraction des caractéristiques au niveau des couches dans un environnement interprétable par les utilisateurs. En second lieu, il faudrait aider les utilisateurs à faire confiance dans les recommandations des modèles profonds. Troisièmement, nous devons permettre aux utilisateurs néophytes de distinguer un réseau de neurones fort par rapport à un réseau de neurones faible, même lorsque les deux réseaux font des prédictions similaires. Il est difficile de répondre à ces attentes si l’utilisateur n’a pas accès à une explication satisfaisante de leur processus de décision. Des efforts ont donc été faits pour prendre en compte les exigences de la transparence, de la clarté et la capacité d’explication qui sont utiles à l’évolution des systèmes d’intelligence artificielle. La Defense Advanced Research Projects Agency (DARPA) des États-Unis a lancé une initiative à cet égard appelée eXplainable Artificial Intelligence (XAI) [GA19] avec l’objectif de généraliser le développement de plateformes explicables (Figure 1.19). En général, l’idée derrière l’explicabilité des modèles d’apprentissage consiste à fournir aux utilisateurs un aperçu des prédictions du modèle. À cet effet, il a émergé un nombre croissant de recherches sur les techniques qui visualisent les prédictions [BADDS⁺20, DR20, GA19]. Ils permettent d’expliquer les mécanismes complexes ou le fonctionnement des algorithmes et montrer les possibilités de les décloisonner. Par exemple, certaines techniques ont permis de visualiser les importantes régions d’une image en utilisant les effets de la déconvolution, la carte d’activation de classe (CAM), et la rétropropagation. Ces techniques apportent des améliorations qualitatives de la prédiction par des modifications du gradient.

La méthode proposée dans ce travail consiste à identifier les insectes ravageurs dans les cultures à l’aide un réseau de neurones convolutif (CNN) explicable. Une telle approche est utile pour la surveillance de la santé des cultures. Elle met en évidence les importantes régions de l’image d’entrée que la méthode de l’explicabilité a déterminé comme étant importantes pour la prédiction. En effet, nous démontrons que l’approche de l’explicabilité améliore efficacement la qualité de prédiction d’un modèle. La plupart des explications ont une forte ressemblance avec les données utilisées pour entraîner le modèle. Afin de fournir aux agriculteurs, des explications qualitatives des images par la localisation de la région infectée et interpréter les résultats obtenus, **il est nécessaire d’évaluer quantitativement le modèle développé**. Par conséquent, les principales contributions de la méthode proposée sont les suivantes :

1. Cette étude est une tentative qui étudie ce que les CNN apprennent durant le diagnostic de la culture. Nous construisons un processus d’interprétation du modèle par l’analyse de la culture à partir d’une image en entrée.
2. D’un point de vue informatique, cette étude fournit des cartes de visualisations d’un CNN appliqué à l’analyse de la culture. En général, ces cartes fournies par les méthodes de l’explicabilité (méthodes xai) sont sensiblement différentes pour la localisation d’objet. Nous nous servons de cette différence pour construire un schéma de localisation des insectes parasites.
3. Nous utilisons la théorie de l’information mutuelle pour montrer qu’il est possible de

combiner les visualisations des méthodes de l'explicabilité et créer des visualisations plus fines de la cible qui facilite sa localisation.

4. Lors de la classification d'images, nos visualisations donnent un aperçu des faiblesses du CNN montrant la présence de l'insecte ravageur pour des prédictions apparemment déraisonnables ou à faible probabilité.

Ce travail est organisé comme suit. La section 5.2 passe en revue les travaux existants liés à la détection des insectes ravageurs y compris les techniques de l'explicabilité. La méthode proposée est présentée dans la section 5.3. La section 5.4 présente les données expérimentales et la mise en œuvre de la méthode proposée. La section 5.5 explique les résultats, les discussions et les différentes analyses des résultats. Enfin la dernière section présente la conclusion de l'étude et les travaux futurs.

5.2 Travaux connexes

Les insectes ravageurs sont des organismes qui endommagent les cultures et provoquent des pertes importantes. Les variabilités saisonnières, les conditions météorologiques et le changement climatique extrêmes dans le monde augmentent aussi le risque de leur apparition. En même temps, ils représentent un danger pour les agriculteurs et une menace pour la sécurité alimentaire [FAO20]. Cependant, il existe une grande variété de ces insectes vivant dans des environnements complexes avec une constante évolution tout au long de la saison. Par conséquent, la capacité de les détecter est d'une grande importance pour la bonne santé et le développement des cultures. Bien que les méthodes traditionnelles soient faciles à comprendre, la plupart d'entre elles sont testées sur des images à faible densité [DLJ⁺19, LWD⁺20]. Il devient donc important de rechercher et de développer des principes directeurs permettant de promouvoir la qualité et la sécurité des récoltes nécessaires à la production ou à la consommation. Autrement dit, la détection à temps des insectes nuisibles peut contribuer à éviter des pertes financières aux agriculteurs et réduire l'utilisation des pesticides qui est un problème pour la santé publique et l'environnement [DBD16, FAO20, LBSM⁺20]. L'apprentissage profond dispose d'une structure hiérarchique formée de plusieurs couches dont les caractéristiques extraites à l'aide des diverses méthodes sont plus efficaces. Ces dernières années, les réseaux de neurones convolutifs sont couramment utilisés pour les tâches de classification d'image et montrent un résultat prometteur.

Grâce aux efforts de l'ImageNet Large Scale Visual Recognition Challenge (LSVRC), plusieurs architectures de réseaux de neurones profonds ont émergé. La précision de la classification d'images de ces réseaux profonds dépasse celle de l'apprentissage automatique traditionnel. Nous évoquons les architectures AlexNet, GoogleNet, VGGNet, ResNet comme des standards au niveau des réseaux de neurones profonds.

Dawei et al. ont proposé une méthode de classification pour dix types d'insectes ravageurs et deux types de mauvaises herbes en utilisant l'apprentissage par transfert [DLJ⁺19]. Leur expérimentation a fourni une performance de 93,84 % de justesse (accuracy) pour la reconnaissance des insectes ravageurs et 98,92 % de justesse pour les mauvaises herbes.

Chen et al. prédisent la présence des insectes ravageurs et des maladies sur le coton en fonction du climat et de l'effet atmosphérique de la terre à l'aide des réseaux à mémoire long-court terme (LSTM) [CXZ⁺20]. Le réseau LSTM ainsi conçu est nommé Bi-Long Short-Term Memory (LSTM). Il se compose en deux LSTM parallèles pour former le modèle de classification.

Les résultats expérimentaux ont montré que le Bi-LSTM a obtenu une bonne prédiction de performance de l'aire sous la courbe (Area Under the Curve ou AUC) de 95%.

Plusieurs réseaux de neurones convolutifs sont discutés dans les travaux de [FYKP17, LW20], pour reconnaître les insectes ravageurs sur les tomates. Dans [FYKP17], les auteurs offrent une « méta-architecture » qui repose sur trois familles de détecteurs et de localisateurs des maladies et des ravageurs : le réseau de neurones convolutifs basé sur une répartition géographique plus rapide (Faster Region-based Convolutional Neural Network ou Faster R-CNN), le réseau pleinement convolutif par région (Region-based Fully Convolutional Network ou R-FCN) et le détecteur multi-boîtes à coup unique (Single Shot Multibox Detector ou SSD). Chaque structure est combinée avec des réseaux de neurones renommés comme VGGNet ou ResNet. Les images utilisées dans le système sont collectées dans des conditions saisonnières variées. L'expérimentation a montré une performance de 83% comme moyenne des précisions (mean Average Precision) avec le cas Faster R-CNN pour VGG-16. Malgré la complexité des certaines classes pour les ravageurs en raison de la forte variation intra-classe et de certaines similitudes inter-classes Faster R-CNN avec VGG-16 montre de meilleurs résultats. L'utilisation de la technique d'augmentation des données a également permis à leur système d'améliorer la précision moyenne pour chaque cas par rapport à un système entraîné sans utiliser l'augmentation des données.

Dans [FYLP18], les auteurs ont proposé d'améliorer le modèle de détection des maladies et ravageur de la tomate dont la tâche de classification est confrontée à un problème bien connu, le déséquilibre des données d'entraînement. La méthode proposée affine leur précédente étude [FYKP17] et obtient 13.70% de gain supplémentaire, c'est-à-dire 96.25 % de moyenne des précisions (mean Average Precision). Ce résultat est dû à la conception d'unité d'intégration de diagnostic qui conserve les échantillons correctement classés afin d'éliminer les échantillons mal classés indiquant à tort la présence d'un insecte ravageur ou d'une maladie.

Liu et Jun ont amélioré le modèle de détection YoloV3 lors de l'identification des tomates dans [LW20]. Ils ont validé leur expérience en comparant les modèles SDD, Faster-R-CNN et le modèle original YoloV3 en mesurant la moyenne des précisions (mean Average Precision). Le nouveau modèle construit est plus performant que les trois autres avec une justesse (accuracy) de 92,39 % et un faible temps d'exécution pour la détection. Une série de transformations du modèle original a permis d'obtenir cette performance : la conversion des couches entièrement connectées pour réduire le nombre de paramètres du réseau par la création des multiples couches combinées (couche 2 et couche 4 convolutives sont combinées en couche 5).

Alves et al ont proposé ResNet34* un nouveau modèle qui est basé sur ResNet34 pour classifier dix types d'insectes ravageurs du coton. Ce modèle de classification utilise la technique de l'apprentissage par transfert pour atteindre finalement une valeur de 98% pour le F-score [ASB20] par rapport aux méthodes standards comme Resnet50, Alexnet ou ResNet34. Le ResNet34* remplace la dernière couche linéaire du modèle original ResNet34 par 7 couches supplémentaires.

Yafen Li et al ont affiné le modèle GoogleNet pour améliorer la tâche classification de dix types d'insectes ravageurs [LWD⁺20]. L'expérimentation a atteint une justesse de classification de 6,22% supérieure à celles des autres modèles ResNet ou VGGNet, et 5,91% de plus supérieurs à celle du modèle original, soit un score de 98,91%. L'ensemble de données initial a subi des prétraitements avant l'étape d'entraînement : la suppression de l'arrière-plan de l'image et l'augmentation de l'ensemble de données. L'implémentation de leur approche comprenait différentes combinaisons de paramètres d'apprentissage, notamment les optimiseurs, et

5.3. Méthode proposée

de taux d'apprentissage dans une validation croisée à 5 plis (5-folds cross-validation). Nous observons que tous ces travaux améliorent la précision des modèles, mais ils présentent des limites en ce qui concerne l'exposition de l'explicabilité en cas d'échec. Avec un peu de recul, il est important de comprendre comment les modèles arrivent à leur conclusion, car les réseaux neurones profonds ont longtemps été considérés comme des boîtes noires. Le concept de l'intelligence artificielle explicable (XAI) a été proposé pour rendre transparent et compréhensible le fonctionnement ou la prédiction modèle par les humains (utilisateurs). Cette notion de confiance est particulièrement importante dans les contextes où une décision peut avoir des conséquences. Les approches XAI pour répondre à cet objectif sont discutées dans la section 1.3.5.

5.3 Méthode proposée

La classification et la localisation d'insectes ravageurs dans les cultures comprend quatre principales étapes : (i) l'extraction des caractéristiques (ii) la classification (iii) la visualisation de la prédiction et (iv) la localisation du ravageur. Nous expliquons chaque élément constitutif dans les sous-sections qui suivent. La méthode proposée pour les réalisations de ces phases est représentée par Figure 5.1).

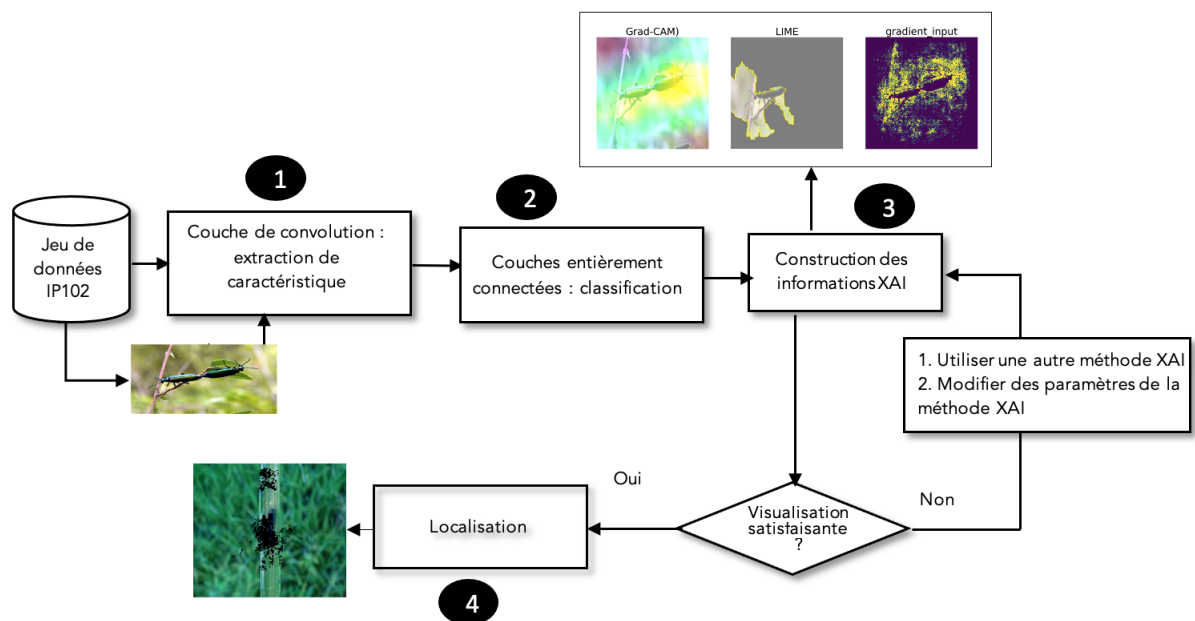


FIGURE 5.1 – Système d'identification des parasites basé sur l'explicabilité du réseau de neurones convolutifs

5.3.1 Extraction des caractéristiques

L'opération d'extraction de caractéristiques consiste à mettre en évidence les propriétés d'une donnée d'entrée (texte, son ou image) afin de faciliter la découverte d'éléments discriminants (chaînes de caractères, graphes, etc.) nécessaires à la classification. Cette opération

permet aux réseaux de neurones convolutifs d'apprendre à identifier les caractéristiques de l'image à l'aide de couches cachées. Les premières couches (couches basses) peuvent apprendre par exemple à détecter les contours de l'image et les dernières couches (couches hautes) à découvrir les motifs pour l'ensemble des images. Pendant l'extraction, les informations et les connaissances sont partagées entre les différentes couches du réseau. Étant donné la nature profonde des couches du réseau, l'augmentation des données et l'apprentissage par transfert sont combinés pour rendre efficace ce processus d'extraction. L'augmentation de données va enrichir notre jeu de données tandis que l'apprentissage par transfert facilite la rapidité de l'entraînement et le réentraînement de tous les poids du réseau. Le CNN pré-entraîné inception-v3 est utilisé comme modèle de base (Figure 5.2). Son architecture est décrite en détail dans [SVI⁺15].

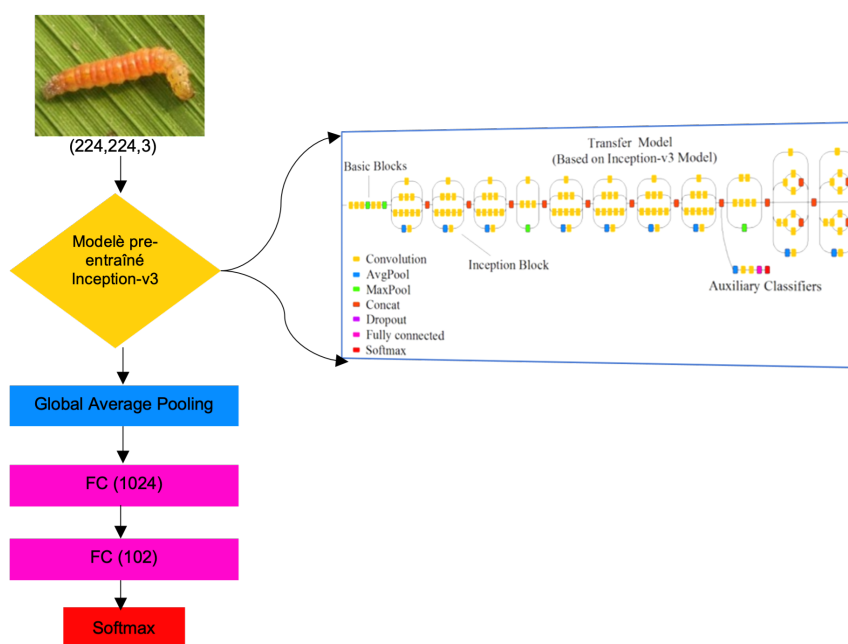


FIGURE 5.2 – Ajustement de l'architecture d'inception-v3 pour la classification des insectes ravageurs.

5.3.2 Classification

Les caractéristiques extraites par le modèle pré-entraîné sont introduites dans le module de classification pour prédire la classe correspondante de l'entrée parmi les 102 classes de IP102 (voir 1.6.2). Nous utilisons l'approche ajustement fin ou fine-tuning, qui est une approche de l'apprentissage par transfert pour adapter le modèle CNN pré-entraîné inceptionv3 à notre étude. Les couches hautes du modèle sont remplacées par trois couches supplémentaires : une couche globale average pooling, deux couches entièrement connectées (fully connected). La dernière couche entièrement connectée prend la fonction d'activation softmax. Le schéma global du modèle proposé est présenté à la Figure 5.2. Nous avons fixé les valeurs des couches fully connected à 1024 et à 102 neurones. Le modèle final reçoit une image d'entrée à trois canaux d'une résolution de 224 x 224 pixels et renvoie un vecteur à 102 dimensions, ce qui correspond aux 102 classes de IP102. Le résultat de la classification peut être déterminé en fonction de la valeur maximale de la probabilité des 102 classes grâce à la fonction d'activation.

5.3.3 Production des cartes visuelles

La satisfaction de l'utilisateur est une condition minimale de l'acceptation d'un système d'apprentissage automatique. Le test de satisfaction (Hoffman et al., 2019 ; Mohseni et al., 2020) se définit comme étant la mesure avec laquelle les utilisateurs estiment qu'ils comprennent le système ou le processus qui leur est expliqué. Cependant, une visualisation claire des résultats d'un modèle de classification peut être une option minimale pour son acceptation. Cette conception soulève quelques discussions dans la recherche en XAI, car il s'agit de choisir une des méthodes satisfaisantes pouvant justifier une modélisation ou une prédiction comme la tâche de classification.

Dans ce contexte, nous proposons plusieurs « cartes visuelles explicables » pour établir une transparence dans les prédictions des modèles. L'explication visuelle représente le résultat prédit sous forme de carte thermique ou de chaleur (heatmap), de carte d'attention, qui montrent l'ampleur d'un phénomène sous forme de couleur ou de variation des pixels en termes d'intensité, de teinte, de contraste ou de saturation. C'est une technique bien utile pour rendre les modèles plus compréhensibles et explicables.

Dans la mesure où le résultat d'une seule méthode est peu suffisant pour l'utilisateur, plusieurs cartes visuelles E_i sont construites à l'aide de diverses méthodes comme les suivantes : **Visualization based Back-Propagation (Grad-Cam, Integrated Gradient, Gradient Input)**, **Visualization Based Local Approximation (LIME)**, **Visualization Based Perturbation (Occlusion Sensitivity)**. Leurs capacités de visualisations sont comparées pour la détection des ravageurs dans les cultures étant donné une entrée D prédit comme classe C .

$$\text{Visualize explanation } (D, C) = \{E_1, E_2, \dots, E_n\}$$

Cette génération des cartes visuelles a pour objectif final de choisir une représentation satisfaisante qui localise la cible. Le choix de la méthode est si difficile parce qu'il s'agit de faire un compromis entre la performance du modèle fourni d'une part et d'autre part la raison de la nature locale / globale des explications. Sa conception ne doit pas se limiter à l'utilisation de ces méthodes, mais de juger leurs utilités et la structure des explications. La démarche adoptée pour lever cette limitation consiste à fusionner ou combiner les représentations locales E_i de la donnée d'entrée D prédite comme classe C afin de former la représentation finale $E_{D,C}$ optimale comme ce qui suit :

$$E_{D,C} = f_{\theta}(E_1, E_2, \dots, E_n)$$

où f représente une fonction de transformation des données à partir d'un opérateur de calcul θ .

Bien que les méthodes explicables soient assez générales dans le sens où elles peuvent être utilisées pour expliquer les fonctions d'activations dans n'importe quelle couche d'un réseau de neurones profonds, dans ce travail, nous nous concentrons sur l'explication des décisions de la couche de sortie uniquement. La section suivante décrit en détail ce processus d'unification des explications visuelles.

5.3.4 Combinaison des méthodes de l'explicabilité pour une localisation fine des insectes ravageurs dans les cultures

Dans ce travail, nous proposons d'utiliser des outils mathématiques pour construire une image combinée qui désigne le résultat final de l'explicabilité. Cette opération vise à augmenter la capacité du modèle de classification par une localisation plus fine d'un objet. Pour ce faire, nous faisons appel à des opérations telles que l'union ou le chevauchement des valeurs d'intensité de pixels correspondantes des images. Dans la littérature, d'autres techniques de transformations peuvent être employées comme l'addition, la soustraction, la moyenne pour construire une image combinée. Le défi à relever est un alignement des images pour mettre en évidence les importantes régions relatives à l'information recherchée.

Le principal problème que pose un alignement d'images est que, l'information contenue dans chaque image utilisée peut être perdue dans le processus de fusion. Afin de remédier à ce problème, l'information mutuelle est utilisée comme mesure pour déterminer la combinaison favorable par paire d'images. L'information mutuelle (MI) obtenue fournit un classement sur le degré de similitude pouvant exister entre les images. Par ailleurs, la MI aide à réduire l'incertitude dans la visualisation et la localisation des insectes ravageurs.

Fondamentalement, l'information mutuelle est une mesure de dépendance statistique entre deux variables aléatoires X et Y . Elle mesure la quantité d'information apportée par une variable X étant donné une valeur connue de l'autre variable Y (Figure 5.3). La dépendance entre les variables est importante lorsque la mesure augmente tandis que la valeur nulle signifie que les variables sont indépendantes.

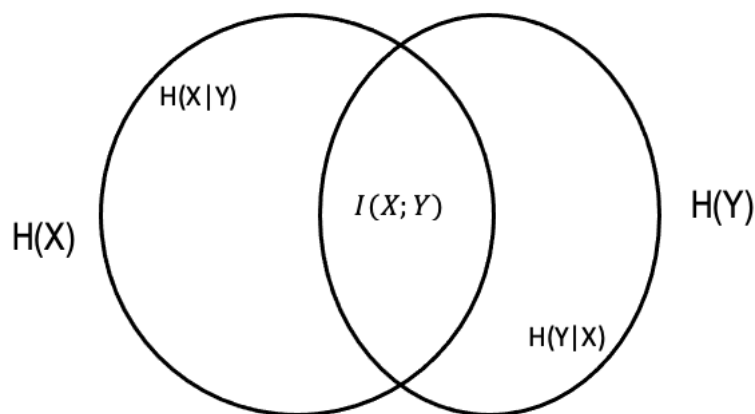


FIGURE 5.3 – Relation entre l'entropie, l'entropie conjointe et l'information mutuelle

L'information mutuelle entre deux variables aléatoires X et Y peut être formulée de la manière suivante :

$$I(X ; Y) = H(X) - H(X | Y)$$

avec

$$H(X) = - \sum_{x \in X} p_X(x) \log p_X(x)$$

$$H(Y) = - \sum_{y \in Y} p_Y(y) \log p_Y(y)$$

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p_{XY}(x, y) \log(p_{XY}(x, y))$$

où $I(X ; Y)$ est l'information mutuelle pour X et Y , $H(X)$ est l'entropie de X et $H(X | Y)$ est l'entropie conditionnelle de X sachant Y connue (observée). Cette mesure est symétrique, l'échange de X avec Y fournit le même résultat.

Lorsque l'information mutuelle entre plusieurs images est faible, nous déduisons que les méthodes ayant fourni ces résultats ne sont pas proches dans la résolution du problème considéré. Cet état de fait est possible du moment où la localisation d'un objet cible (insecte ravageur) peut être biaisée (ou confuse) uniquement dans certaines configurations : la sortie d'une méthode se différencie grandement de celles des autres méthodes. À cet effet, nous avons croisé (aligné) les cartes visuelles explicatives des différentes méthodes XAI, par chevauchement (overlap) ou union (comme montré par la Figure 5.4). Le chevauchement est l'intersection (correspondant à l'opérateur AND) des espaces d'images tandis que l'union (correspondant à l'opérateur OU) représente la fusion de ces espaces, c'est-à-dire un remplissage des espaces non définis avec des zéros. La disjonction exclusive (correspondant à l'opérateur XOR) est définie par l'union moins l'intersection.

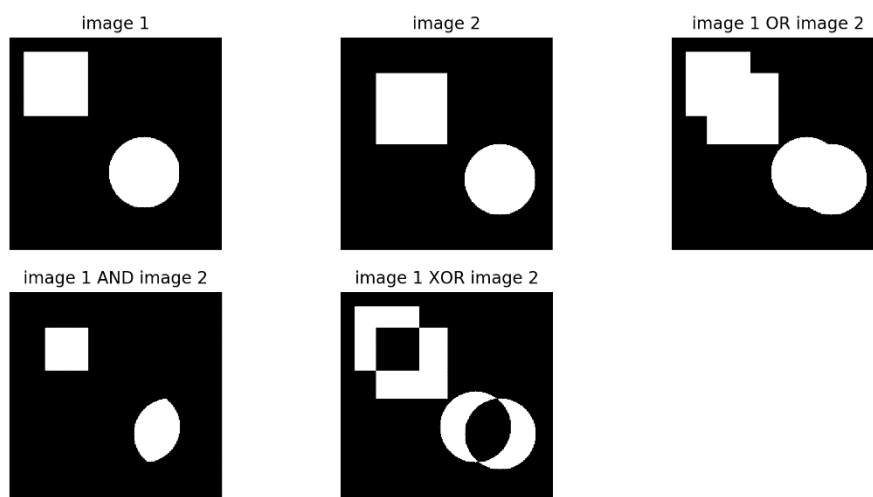


FIGURE 5.4 – Alignement des images en une seule image en utilisant les opérateurs OR, AND, and XOR

L'importance de ces opérations vient du fait suivant : certaines images issues de l'application des méthodes XAI peuvent être porteuses d'informations absentes ou faiblement décrites par d'autres méthodes qui n'ont pas réussi à localiser correctement l'objet cible.

Par exemple, l'union permet de tirer profit de la complémentarité des méthodes en utilisant ensemble les informations issues des caractéristiques de chacune de ces méthodes. Ainsi, des besoins de résolution de problème (comme la localisation d'objets dans une image) non couverts par une méthode A , peuvent être satisfaits par une méthode B .

L'intérêt de l'intersection est de vérifier la similarité entre les résultats fournis par les méthodes

employées. Son objectif est de montrer la concordance des méthodes et éventuellement la fiabilité des résultats obtenus. De ce fait, dans des circonstances analogues, une méthode *A* peut-être remplacée par une méthode *B* similaire.

Le processus d'alignement est appliqué sur des images de même dimension. Il est composé en trois étapes définies comme suit :

Etape 1 : elle consiste à convertir des images en couleurs en des images de niveaux de gris d'intensité.

Etape 2 : elle comprend la détection des bords d'images en utilisant un seuil permettant d'obtenir des images en blanc sur noir. La valeur du seuil est déterminée grâce à des techniques de seuillage qui partitionne une image dans le premier plan (en avant) et à l'arrière-plan. Cette technique modifie la représentation d'une image en une autre plus facile à traiter. D'une manière générale, les méthodes de seuillage remplacent chaque pixel de l'image source avec un pixel noir si l'intensité est inférieure à la valeur du seuil ou blanc si l'intensité de pixel est supérieure à cette valeur du seuil. Le procédé d'Otsu [Ots79, YSLC12] est une technique de seuillage d'image utilisée pour la segmentation d'une image en niveau de gris bimodal. C'est une technique robuste pour la détection, le suivi et la reconnaissance des objets. Dans ce travail, la combinaison des algorithmes d'inversion binaire (Inverse-Binary Thresholding) et d'Otsu, convenaient le mieux à nos images pour extraire l'objet cible, à savoir le ravageur de l'arrière-plan. L'algorithme d'Otsu repose sur des informations statistiques dérivées de l'histogramme (comme l'entropie). Il calcule automatiquement la valeur optimale du seuil qui sépare les pics de l'histogramme.

Etape 3 : elle permet la combinaison entre les paires des cartes visuelles issues de deux transformations précédentes en une seule image en utilisant l'opération d'union, de chevauchement, ou de différence. Le résultat de la fusion constitue une description fine de la localisation du ravageur. Cette combinaison (voir la section 5.6) est guidée par l'information mutuelle qui donne un score à chaque paire. Un score élevé signifie une meilleure combinaison.

5.4 Expérimentation

5.4.1 Paramètres d'apprentissage

Dans ce travail, l'optimiseur Adam¹ a été sélectionné pour optimiser la pondération des poids de la fonction de combinaison. L'entraînement du modèle est effectué avec un taux d'apprentissage de $1e-4$ et un ensemble de paramètres supplémentaires $\beta_1 = 0,9$ et $\beta_2 = 0,999$. Seule la valeur du taux d'apprentissage a été mise à jour avec $1e-4$. Pour classifier les caractéristiques extraites, nous utilisons la fonction d'activation softmax avec 102 unités de valeurs. La distance entre les prédictions du modèle et le vrai label est définie par la fonction de perte, l'entropie croisée catégorique. Le reste des paramètres est présenté dans la Table 5.1.

Dans notre jeu de données, nous avons en moyenne 737 instances par classes. Nous améliorons l'entraînement du modèle profond par séries de transformations géométriques des données : rotation aléatoire, retournement horizontal et vertical, etc. Enfin, nous prétraitons les images en

1. Algorithme d'estimation du mouvement adaptatif (Adaptive Movement Estimation en anglais)

Paramètre	Valeur
Taille de l'image (en pixel)	224x224
Taille du lot	64
Taux d'apprentissage	1e-4
Epoque	60
Optimiseur	Adam
Fonction de perte	entropie croisée catégorique
Plage de cisaillement	0.2
Remise à l'échelle	1./255
Plage de zoom	0.2
Plage de rotation	25
Retournement vertical	Vrai
Retournement horizontal	Vrai

TABLE 5.1 – Valeurs des hyper-paramètres utilisés dans les expériences.

normalisant chaque pixel entre 0 et 1. Toutes les images sont aussi redimensionnées à la taille de 224x224 pixels. Nous définissons le reste des paramètres supplémentaires dans la Table 5.1.

5.4.2 Implémentation

Nos implémentations ont été réalisées à l'aide de la bibliothèque Python. Les traitements accomplis pendant l'alignement d'images ainsi que les calculs de l'information mutuelle sont implémentés avec les librairies open source OpenCV² et Scikit Learn³. Les réseaux de neurones proposés ont été mis en œuvre à l'aide du Framework TensorFlow en utilisant 1 GPU Nvidia, 50G de mémoire pour l'entraînement et le test.

5.5 Résultats et discussions

Les résultats du modèle et des méthodes de visualisation sont décrits dans cette section. Nous dressons une étude comparative en termes de *justesse* de classification et de *F1 – score* pour notre CNN inception-v3 et celui du travail original. Les méthodes de visualisation telles que Grad-CAM, LIME, etc. afin de localiser les régions de la culture où apparaissent les insectes ravageurs.

2. <https://opencv.org/>

3. <https://scikit-learn.org/stable/index.html>

5.5.1 Résultat de la classification

Pour la classification d’images, le CNN ajusté inception-v3 a été appliqué avec succès. La Table 5.2 présente les valeurs de justesse de la classification et F1 sur l’ensemble test. Avec un meilleur score à 1 et un pire score à 0, le CNN inception-v3 fournit les valeurs de 52,99% pour la justesse et 48,77% pour le score F1 après 60 époques.

La Table 5.2 fournit aussi les résultats comparatifs avec le travail original effectué par [WZL⁺19]. Ces résultats démontrent que l’utilisation de inception-v3 a son intérêt en raison du traitement parallèle (bloc inception) et de la profondeur optimale de son architecture, ce qui améliore l’apprentissage des caractéristiques de l’image. Les CNN du travail de [WZL⁺19] sont entraînés par la technique de l’augmentation des données qui a donné de meilleurs résultats par rapport au système d’entraînement sans augmentation de données. Ces résultats justifient, l’apport de l’augmentation des données dans l’apprentissage d’un modèle prédiction.

Travaux	Modèle CNN	Justesse (%)	F1-score (%)
Travail original [WZL ⁺ 19]	AlexNet	41.8	34.1
	GoogleNet	43.5	32.7
	VGG16	48.2	38.7
	ResNet50	49.2	40.1
Travail proposé	Inception-v3	52.99	48.77

TABLE 5.2 – Comparaison de nos résultats avec ceux du travail original

5.5.2 Analyse qualitative

Les cartes de visualisation sont présentées par la Figure 5.5. Chaque ligne commence par une image originale suivie d’une carte d’explication générée par les méthodes XAI comme Grad-CAM [SCD⁺16], LIME [RSG16], Gradient Input [SGSK17], Integrated Gradients [STY17], Occlusion Sensitivity [ZF13] pour expliquer les décisions du modèle. La comparaison met en évidence les régions les plus importantes dans chaque image. Les implémentations de ces méthodes XAI sont fournies dans la bibliothèque `tf_explain`⁴ et `lime`⁵ disponibles dans le dépôt GitHub.

En observant les cartes, les méthodes XAI identifient les régions de présence de l’insecte ravageur et créent des tâches de segmentation pour évaluer leur sensibilité pour la classe prédite.

Nous observons que les régions mises en évidence varient d’une méthode à l’autre, ce qui a une influence sur les attentes de l’utilisateur. Par exemple, les régions détectées par la méthode Grad-CAM ou Integrated Gradient sont plus informatives que celles détectées par les autres méthodes (Figure 5.5). Malgré la mauvaise prédiction (probabilité faible) de certaines images, les méthodes Grad-CAM, LIME et Integrated Gradients, sont parvenues à identifier d’importantes régions. Cependant, l’analyse des occlusions ne parvient pas à détecter les régions multiples, les résultats de ces méthodes souffrent d’une spécificité insuffisante qui est l’arrière-plan bruité.

4. <https://github.com/sicara/tf-explain>

5. <https://github.com/marcotcr/lime>

5.5. Résultats et discussions

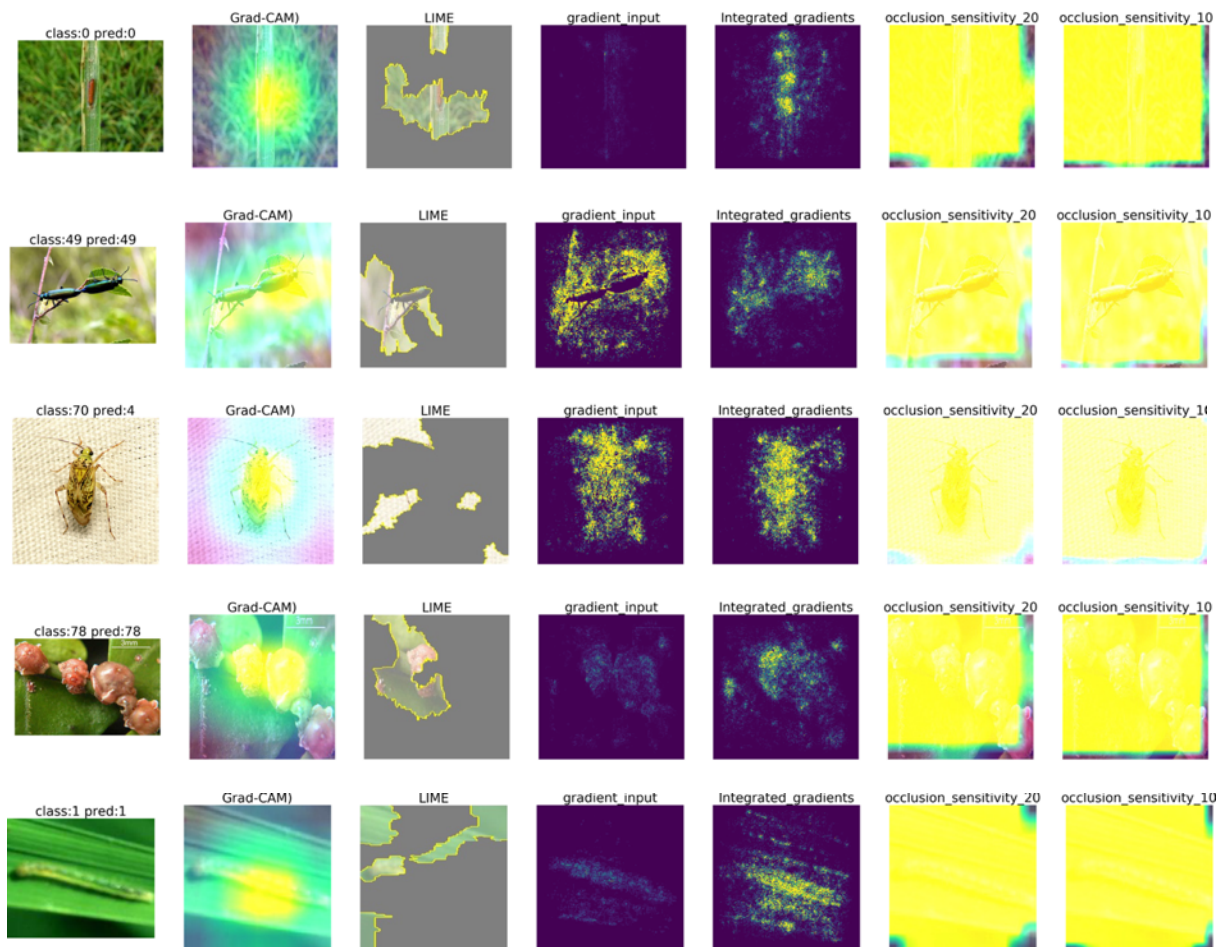


FIGURE 5.5 – Exemples de résultats de classification explicables. « class » désigne la vérité terrain, « pred » classe prédite.

Nous pouvons vérifier visuellement et progressivement que certaines améliorations sont obtenues au niveau des couches de neurones en augmentant les effets de perturbation avec la méthode (Figures 5.6 et 5.7). Dans certaines situations, les méthodes de l’explicabilité peuvent échouer (ou générer des explications inappropriées) lorsque les infections ne couvrent pas une grande surface (c’est le cas par exemple de la première ligne sur la Figure 5.5).

La visualisation du CNN montre la possibilité d’ouvrir la boîte noire de l’apprentissage profond afin de réduire les obstacles liés à leur utilisation. Néanmoins, les méthodes LIME ou occlusion nécessitent de multiples inférences pour obtenir une carte visuelle par image d’entrée (comme montré par la Figure 5.7). Ces modifications rendent l’analyse coûteuse en calcul et lente pour de nombreux contextes, surtout dans une analyse en temps réel ou le déploiement de dispositifs mobiles.

Les résultats indiquent que même si les méthodes de visualisation génèrent des résultats significatifs, les humains jouent toujours le rôle le plus important dans l’appréciation des résultats. Ils relient les résultats produits par l’ordinateur avec ceux des professionnels des plantes à expliquer pourquoi le modèle a fait une prédiction particulière. Ces explications peuvent être utilisées pour communiquer à d’autres utilisateurs (potentiellement des non experts).

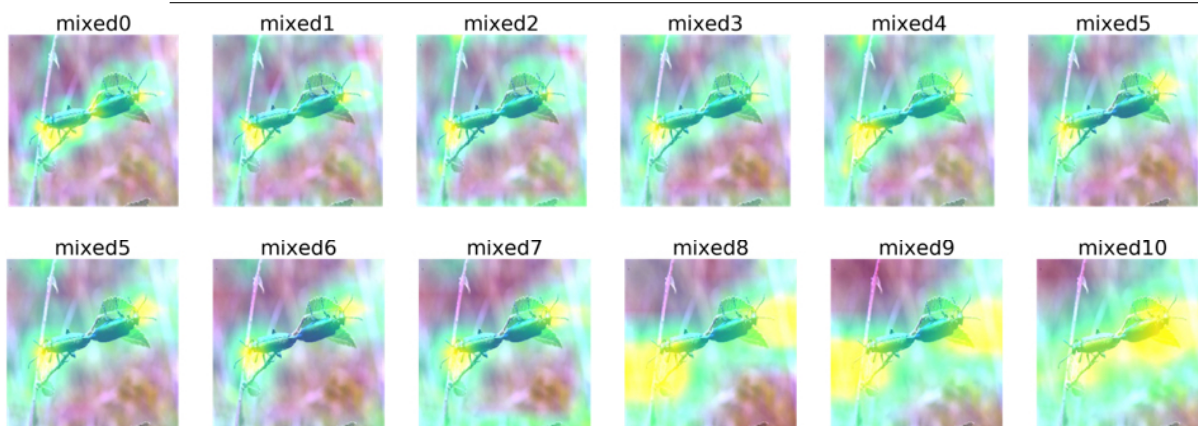


FIGURE 5.6 – Carte d'explication d'une observation avec Grad-CAM sur les couches profondes mixed0 à mixed10

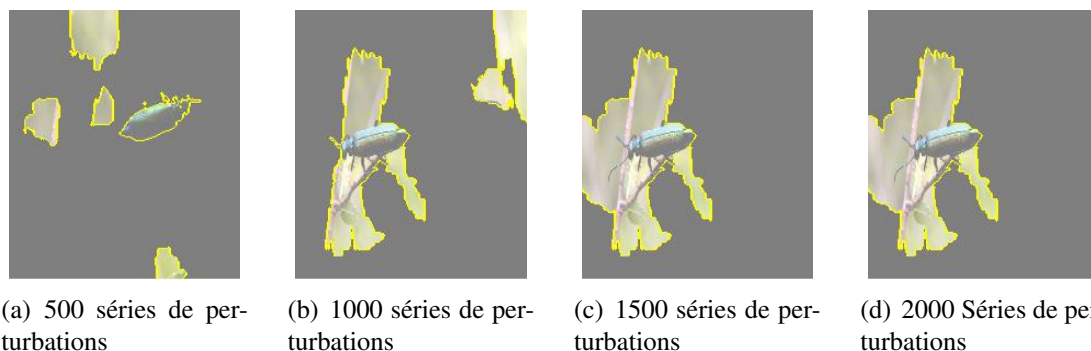


FIGURE 5.7 – Paramètres d'apprentissage affectant la visualisation LIME.

Il est essentiel de garder à l'esprit que si les cartes visuelles sont l'un des moyens les plus courants et les plus naturels de présenter une explication de modèle, elles sont également sujettes à interprétation par le praticien. Les praticiens peuvent interpréter l'explication différemment, en particulier lorsque la méthode d'explication est inconnue.

5.6 Combinaison des méthodes de l'explicabilité

Les figures 5.8 et 5.9 présentent les divers résultats de la combinaison ou l'alignement des cartes visuelles obtenues. Sur ces figures, les régions découvertes par l'opération d'alignement sont représentées par une tâche en couleur noire. Ses résultats indiquent que la localisation d'objet guidée par la théorie de l'information mutuelle est efficace et efficiente. L'alignement qui détient un score élevé dans le calcul de l'information mutuelle représente au mieux la cible. Les résultats d'alignement permettent de mieux situer la région de la culture sujette à une attaque fongique. Les opérateurs employés comme AND, OR, et XOR sont considérés pour mesurer le chevauchement entre les paires d'images. Lorsque l'information mutuelle est la plus élevée entre deux images, alors cet alignement est considéré comme positif ou certain (Figure 5.9) par rapport aux autres combinaisons. Lorsqu'elle est faible entre deux images, l'alignement est dit négatif ou incertain (Figure 5.10).

Dans les cas de figures citées, nous observons que l'alignement des images 1 et 3 qui sont

5.6. Combinaison des méthodes de l'explicabilité

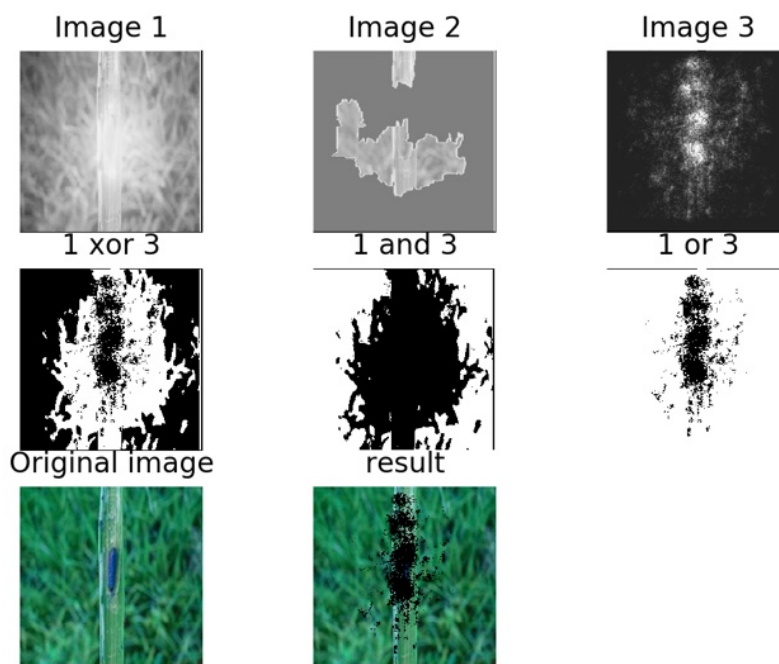


FIGURE 5.8 – Alignement des images avec les mesures de l'information mutuelle (MI) suivantes : $MI(\text{image 1 et 2}) = 0.33$; $MI(\text{image 1 et 3}) = 0.56$; $MI(\text{image 2 et 3}) = 0.398$. Les images image 1, image 2 et image 3 correspondent respectivement aux résultats des méthodes Grad-CAM, LIME et Integrated Gradient

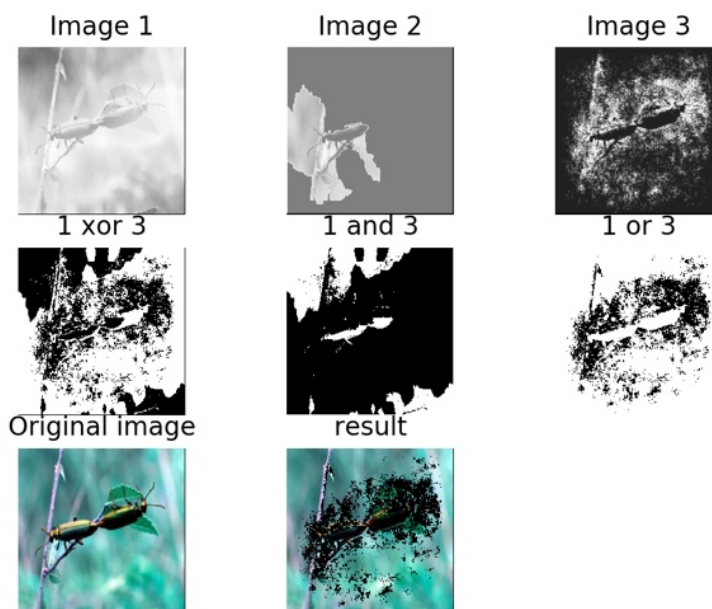


FIGURE 5.9 – Alignement des images avec les mesures de l'information mutuelle (MI) suivantes : $MI(\text{image 1 et 2}) = 0.20$; $MI(\text{image 1 et 3}) = 0.36$; $MI(\text{image 2 et 3}) = 0.25$. Les images image 1, image 2 et image 3 correspondent respectivement aux résultats des méthodes Grad-CAM, LIME et Integrated Gradient

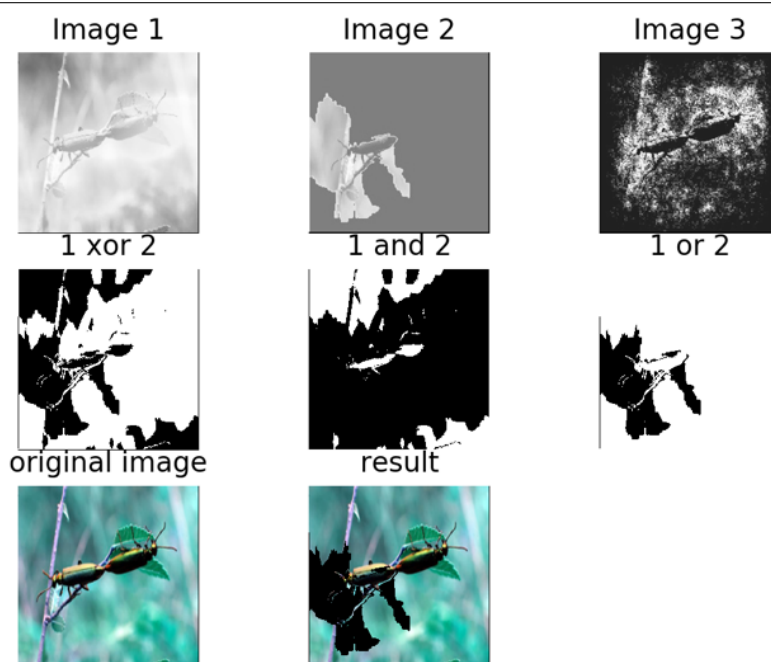


FIGURE 5.10 – Un exemple d’alignement négatif avec MI (image 1 et 2) = 0.20

respectivement les résultats des méthodes Grad-CAM et Integrated Gradient donne en premier lieu des informations les plus significatives de la localisation de l’insecte ravageur comparativement aux autres. Les transformations d’images avec les opérateurs binaires OR, AND, et XOR sont des évaluations logiques qui nous permettent d’extraire les objets du premier plan de l’image. Parmi les opérateurs logiques employés, celui qui localise le mieux le ravageur dans l’arrière-plan de l’image est l’opérateur OR. L’opérateur OR prend deux arguments et effectue l’opération OR sur chaque bit de la valeur de l’opérande. Nous avons donc un moyen d’approfondissement de la recherche d’information dans les paires d’images alignées. Lorsque l’objet cible est présent dans le plan de l’image, cet opérateur met un 1 et un 0 s’il est absent de ce plan. Le résultat fourni mérite des analyses plus poussées permettant une compréhension plus fine des mécanismes sous-jacents à cette observation et la marge d’erreur d’appréciation associée.

Notre étude a montré que les méthodes de visualisation pour localiser des insectes ravageurs ouvrent une nouvelle voie pour les études de la science des plantes, dans un contexte pluridisciplinaire où les informaticiens et les scientifiques des plantes travaillent en synergie grâce aux modèles d’apprentissage automatique. Grâce à l’accompagnement de ces experts du domaine, nous pouvons apprécier si les caractéristiques découvertes ont une signification physiologique concrète afin d’effectuer un ajustement adéquat aux égards des réalités du terrain.

5.7 Conclusion

Cependant, les modèles d’apprentissage profond sont généralement considérés comme des boîtes noires et il est difficile de comprendre ce que le modèle a appris. L’approche proposée dans ce chapitre utilise des méthodes de visualisation pour comprendre ce que le modèle a appris, ce qui non seulement garantit la pertinence des résultats générés, mais aussi améliore leur qualité. Si des biais présents dans les données affectent le processus d’entraînement, ces tech-

5.7. Conclusion

niques de visualisation permettent d'éclaircir les impacts sur le modèle et de visualiser ces biais. Nous avons formulé une proposition de renforcement de l'explicabilité par la combinaison des résultats des méthodes de visualisation. Cette combinaison guidée par théorie de l'information mutuelle a permis de mettre en évidence les régions les plus problématiques dans chaque image contenant les ravageurs. Elle constitue une preuve visuelle claire mise à disposition de l'utilisateur pour une meilleure perception de la façon dont un ravageur est identifiable dans l'image. En effet, certaines méthodes ont montré des potentialités d'explicabilité plus avantageuses par rapport à d'autres. Par exemple, Grad-CAM suggère des éléments visuels offrant une meilleure explication que la méthode Occlusion Sensitivity. Lorsque l'utilisateur regarde les résultats fournis, la région où l'insecte apparaît est facilement visible grâce à l'observation de la carte visuelle explicable, ce qui facilite une bonne identification du problème considéré. Ce travail est mené sur le jeu de données de référence IP102 qui contient une variété d'insectes nuisibles avec différentes caractéristiques dans des environnements réels. IP102 est l'un des plus grands jeux de données contenant plus de 75 000 images pour 102 classes et 737 images en moyenne par classe. Pour la tâche de classification réalisée, l'apprentissage par transfert et la technique d'augmentation des données ont contribué à l'amélioration de la précision du modèle. Grâce à cela, le CNN pré-entraîné inception-v3 a atteint une précision de 52,99% sur l'ensemble de test, ce qui est meilleur que le résultat du travail original considéré.

Pour étendre le travail actuel à court terme, nous envisageons d'étudier de manière plus détaillée les éléments pouvant contribuer à une meilleure compréhension de la mauvaise classification des images mal caractérisées. Par conséquent, il peut être possible d'identifier les modifications à appliquer afin de fournir des informations supplémentaires pour obtenir une explication claire du modèle de classification. À moyen terme, nous songeons à une utilisation comparative avec d'autres architectures d'apprentissage profond telles que DenseNet, et InceptionResNet qui sont susceptibles de fournir aussi des résultats de classification tout aussi intéressants sur le même jeu de données.

À travers nos discussions, nous comprenons qu'il est nécessaire de connaître comment les modèles prédictifs arrivent à leurs conclusions pour leur déploiement. Même si les méthodes de visualisation génèrent des résultats significatifs, les humains jouent toujours le rôle le plus important dans l'appréciation des résultats. Ils relient les résultats produits par l'ordinateur avec ceux des professionnels des plantes. Ils sont capables d'expliquer pourquoi le modèle a fait une prédiction particulière. Néanmoins, ces professionnels peuvent interpréter l'explication donnée différemment, en particulier lorsque la méthode est inconnue.

Conclusion et perspectives

L'exploitation du numérique dans agriculture permet de développer des outils d'optimisation du travail des agriculteurs. Elle assure la réalisation d'un nouveau schéma productif dont le but est d'obtenir un meilleur rendement. Cependant, l'apprentissage profond apporte de réelles contributions grâce à des dispositifs automatiques intelligents et à des modèles prédictifs utilisant les réseaux de neurones profonds.

À la suite d'une étude bibliométrique, les travaux analysés ont prouvé que, l'apprentissage profond apparaît comme une voie d'amélioration de la précision des opérations agricoles et fournit de précieux conseils aux agro-industrielles par la valorisation des connaissances issues de l'exploitation des données agricoles. En conséquence, la recherche a donné un aperçu des applications possibles et des travaux de recherche et de R&D en cours. Nous avons donc discuté des principales contributions des chercheurs impliqués dans le développement d'une agriculture de précision. Il nous a été possible de proposer de nouvelles possibilités d'exploration des images agricoles pour une analyse de données plus efficace. Ces propositions exploitent la capacité des réseaux de neurones convolutifs dans les tâches de classification sur des images. À travers nos études, il est évident que les réseaux de neurones convolutifs améliorent certaines pratiques agricoles pour rendre l'agriculture davantage durable et précise. La vision d'une agriculture de précision permet aux agriculteurs d'obtenir plus de résultats avec peu de charges de travail. Pour cela, les systèmes sont conçus pour déterminer les importants paramètres liés à la résolution des tâches d'exploitation agricole (comme le rendement des cultures, l'irrigation, l'occupation du sol, la surveillance des cultures, le désherbage, etc.).

Les travaux étudiés présentent des contributions pour le traitement d'images lors de la surveillance des cultures. De nos jours, les développements technologiques offrent à l'humanité des moyens pour produire suffisamment de nourriture. Cependant, de nombreuses menaces pèsent sur les cultures affectant la qualité et la quantité des productions. Ces menaces sont le résultat d'un changement climatique, des maladies des plantes, des insectes ravageurs, etc. Relever ces défis est un problème majeur pour les scientifiques et les phytopathologistes notamment avec les techniques de l'apprentissage profond. Dans ce contexte, quatre problématiques basées sur l'apprentissage profond, et plus particulièrement les réseaux de neurones convolutifs ont été introduites : (i) la définition d'un cadre méthodologique des futurs systèmes agricoles, (ii) la détection des attaques fongiques par la technique d'apprentissage par transfert (ii) le partage des caractéristiques apprises dans le réseau de neurones convolutifs, et (iii) l'implication de l'acteur humain dans le processus d'apprentissage automatique. Ces problématiques ont été traitées en proposant quatre (4) contributions qui vont servir de méthodes d'apprentissage dans l'agriculture. Nos méthodologies présentent des opportunités d'analyse, d'interprétation et de visualisation des images afin de gagner en transparence et compréhension des modèles prédictifs.

Cependant, le traitement automatisé des données agricoles n'est pas une tâche facile, car elle

présente plusieurs défis scientifiques et technologiques qui ouvrent de nouvelles manières de voir et d'analyser le monde. Nous nous sommes focalisés sur les réseaux de neurones convolutifs qui ont exhibé des performances intéressantes, en particulier dans le domaine de la vision par ordinateur. Ils vont donc permettre l'extraction de connaissances à partir de données pour aider contextuellement les agriculteurs dans le processus de prise de décision avec les éléments de support explicatifs. En conséquence, les agriculteurs peuvent choisir des méthodes appropriées pour élaborer de pertinentes stratégies de protection des cultures afin de favoriser la durabilité des systèmes agricoles considérés.

Les travaux de recherche décrits dans ce manuscrit nous ont permis de cerner la dimension du traitement des données (agricoles) avec l'apprentissage profond, qui tire sa force dans les théories mathématiques au service du traitement des données avec une considérable puissance des calculs. Il est difficile de définir clairement, quelles représentations en couches du réseau de neurones profond, il faudrait pour un apprentissage efficient. Les architectures standards convolutives comme VGGNet, ResNet, Inception servent de bases pour vérifier de nombreux travaux de recherches ainsi que le transfert d'apprentissage pour partager les connaissances apprises sur les données tout en limitant les temps de calcul.

Par essence, les méthodes d'apprentissage profond s'appuient sur des jeux de données larges pour l'entraînement. C'est ainsi que l'apprentissage profond combiné avec l'augmentation des données apparaît comme une solution alternative en présence de la faiblesse des données. Aujourd'hui, même, si, les pratiques agricoles pilotées par les systèmes intelligents gagnent du terrain, de nombreuses difficultés existent ouvrant la voie à des perspectives futures :

La robotique et l'intelligence artificielle (IA) permettent de réduire les défis les plus importants ou compliqués pour les humains. Ainsi, les nouvelles techniques doivent s'appuyer sur des objets connectés pour répondre à la demande de cultiver plus avec moins de charges pour les producteurs et rendre l'agriculture accessible à tous.

Les algorithmes développés lors de nos travaux de recherches peuvent être déployés sur les objets connectés ou embarqués sur des engins (drones et tracteurs) ou des humains (lunettes connectées, smartphones, appareils photos) dans les pratiques agricoles.

Le coût d'investissement dans l'agriculture numérique peut s'avérer énorme. Nombreux sont les agriculteurs qui ne disposent pas d'un budget suffisant pour un investissement. La création des portails collaboratifs à vocation agricole pour mettre à la disposition des agriculteurs, des données multiples (cartes de fertilité, cartes visuelles explicables, images satellitaires et relevés météorologiques) est nécessaire dans ce cas.

Les algorithmes d'apprentissage pour détecter les attaques fongiques, gérer les intrants ou cartographier le rendement des cultures doivent conduire à la réalisation des nouveaux services exploitables en condition réelle.

L'utilisation d'images de drones dans l'agriculture est devenue un point chaud. Les images qu'ils fournissent sont en haute définition et ils ont une grande flexibilité en termes d'acquisition de données. En conséquence, leur utilisation répond à des perspectives de l'agriculture dans des opérations ciblées comme la pulvérisation.

Les nouvelles recherches en agriculture de précision devront favoriser l'échange entre les systèmes agricoles et les utilisateurs. L'objectif de cette intégration est de construire une IA où un utilisateur peut faire confiance à sa recommandation et justifier son utilisation (ou ses prédictions). Une telle solution aura la capacité de fournir aux utilisateurs les informations et connaissances utiles à leurs exploitations pour des fins d'amélioration de leurs rendements.

Bibliographie

- [AAMB18] Javad Abbasi Aghamaleki and Sina Moayed Baharlou. Transfer learning approach for classification and noise reduction on noisy web data. *Expert Systems with Applications*, 105 :221–232, September 2018.
- [ABA17] Jihen Amara, Bassem Bouaziz, and Alsayed Algergawy. A deep learning-based approach for banana leaf diseases classification. *Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband*, 2017.
- [ABC⁺19] Vijay Arya, Rachel K. E. Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Q. Vera Liao, Ronny Luss, Aleksandra Mojsilović, Sami Mourad, Pablo Pedemonte, Ramya Raghavendra, John Richards, Prasanna Sattigeri, Karthikeyan Shanmugam, Moninder Singh, Kush R. Varshney, Dennis Wei, and Yunfeng Zhang. One Explanation Does Not Fit All : A Toolkit and Taxonomy of AI Explainability Techniques. *arXiv :1909.03012 [cs, stat]*, September 2019.
- [AC17] Massimo Aria and Corrado Cuccurullo. Bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4) :959–975, November 2017.
- [ACÖG18] Marco Ancona, Enea Ceolini, Cengiz Öztireli, and Markus Gross. Towards better understanding of gradient-based attribution methods for Deep Neural Networks, February 2018.
- [AHRM14] Auzi Asfarian, Yeni Herdiyeni, Aunu Rauf, and Kikin Hamzah Mutaqin. A computer vision for rice disease identification to support integrated pest management. *Crop Protection*, (61) :103–104, 2014.
- [AKS⁺19] Marko Arsenovic, Mirjana Karanovic, Srdjan Sladojevic, Andras Anderla, and Darko Stefanovic. Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry*, 11(7), 2019.
- [ALG⁺19] Antoine Affouard, Jean-Christophe Lombardo, Hervé Goëau, Pierre Bonnet, and Alexis Joly. PI@ntNet, April 2019.
- [ALY⁺21] Pei An, Junxiong Liang, Kun Yu, Bin Fang, and Jie Ma. Deep structural information fusion for 3d object detection on lidar-camera system. *Computer Vision and Image Understanding*, page 103295, 2021.
- [ALYS17] S. Atharva, X. Liu, X. Yang, and D. Shi. A patch-based convolutional neural network for remote sensing image classification. *Neural Networks. Volume 95*, pages 19–28, 2017.

- [ARS18] Muhammad Jamal Afridi, Arun Ross, and Erik M. Shapiro. On automated source selection for transfer learning in convolutional neural networks. *Pattern Recognition*, 73 :65–75, January 2018.
- [ASAV13] Sai Arivazhagan, R Newlin Shebiah, S Ananthi, and S Vishnu Varthini. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International : CIGR Journal*, 15(1) :211–217, 2013.
- [ASB20] Adão Nunes Alves, Witenberg S.R. Souza, and Díbio Leandro Borges. Cotton pests classification in field-based images using deep residual networks. *Computers and Electronics in Agriculture*, 174 :105488, July 2020.
- [ASM17] Halimatu Sadiyah Abdullahi, Ray E. Sheriff, and Fatima Mahieddine. Convolution neural network in precision agriculture for plant image recognition and classification. In *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, pages 1–3, Luton, August 2017. IEEE.
- [AZH⁺21] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning : Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1) :1–74, December 2021.
- [BADDS⁺20] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bernetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable Artificial Intelligence (XAI) : Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58 :82–115, June 2020.
- [BBM⁺15] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE*, 10(7) :e0130140, July 2015.
- [BBSF19] Philippe Borianne, Frederic Borne, Julien Sarron, and Emile Faye. Deep mangoes : from fruit detection to cultivar identification in colour images of mango trees, 2019.
- [BCDZ17] Alessio Benavoli, Giorgio Corani, Janez Demšar, and Marco Zaffalon. Time for a Change : A Tutorial for Comparing Multiple Classifiers Through Bayesian Analysis. *Journal of Machine Learning Research*, 18(77) :1–36, 2017.
- [BCSB20] Utpal Barman, Ridip Dev Choudhury, Diganto Sahu, and Golap Gunjan Barman. Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease. *Computers and Electronics in Agriculture*, 177, October 2020.
- [BEF19] Kevin Bascol, Rémi Emonet, and Elisa Fromont. Improving Domain Adaptation By Source Selection. In *ICIP 2019 - IEEE International Conference on Image Processing*, Taipei, Taiwan, September 2019. IEEE.
- [BHC18] M Dian Bah, Adel Hafiane, and Raphael Canals. Deep learning with unsupervised data labeling for weed detection in line crops in uav images. *Remote Sensing*, 10(11), 2018.

- [BLSB04] Matthew R. Boutell, Jiebo Luo, Xipeng Shen, and Christopher M. Brown. Learning multi-label scene classification. *Pattern Recognition*, 37(9) :1757–1771, September 2004.
- [BMS17] Simone Bianco, Davide Mazzini, and Raimondo Schettini. Deep Multibranch Neural Network for Painting Categorization. In Sebastiano Battiato, Giovanni Gallo, Raimondo Schettini, and Filippo Stanco, editors, *Image Analysis and Processing - ICIAP 2017*, volume 10484, pages 414–423. Springer International Publishing, Cham, 2017.
- [C⁺15] François Chollet et al. Keras. <https://keras.io>, 2015.
- [Car97] Rich Caruana. Multitask Learning. *Machine Learning*, 28(1) :41–75, July 1997.
- [CCZ⁺20] Junde Chen, Jinxiu Chen, Defu Zhang, Yuandong Sun, and Y. A. Nanekaran. Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture*, 173 :undefined–undefined, 2020.
- [CFHR17] Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint Distribution Optimal Transportation for Domain Adaptation. *arXiv :1705.08848 [cs, stat]*, May 2017.
- [CFT14] Nicolas Courty, Rémi Flamary, and Devis Tuia. Domain Adaptation with Regularized Optimal Transport. In Toon Calders, Floriana Esposito, Eyke Hüllermeier, and Rosa Meo, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 8724, pages 274–289. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.
- [CH16] Gong Cheng and Junwei Han. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117 :11–28, July 2016.
- [CHL⁺20a] Ching-Ju Chen, Ya-Yu Huang, Yuan-Shuo Li, Chuan-Yu Chang, and Yueh-Min Huang. An AIoT based smart agricultural system for pests detection. *IEEE Access*, 8 :180750–180761, 2020.
- [CHL20b] Yuh-Shyan Chen, Chih-Shun Hsu, and Chia-Ling Lo. An entire-and-partial feature transfer learning approach for detecting the frequency of pest occurrence. *IEEE Access*, 8 :92490–92502, 2020.
- [CKKT19] Solemane Coulibaly, Bernard Kamsu-Foguem, Dantouma Kamissoko, and Daouda Traore. Deep neural networks with transfer learning in millet crop images. *Computers in Industry*, 108 :115–120, June 2019.
- [CLX18] Yidong Chai, Hongyan Liu, and Jie Xu. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. *Knowledge-Based Systems*, 161 :147–156, 2018.
- [CR20] Dae-Kyoo Kim Chengjuan Ren, and Dongwon Jeong. A Survey of Deep Learning in Agriculture : Techniques and Their Applications. *Journal of Information Processing Systems*, 16(5) :1015–1033, October 2020.
- [CSHB18] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++ : Generalized gradient-based visual explanations for deep convolutional networks. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar 2018.

- [CTH⁺09] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. NUS-WIDE : A real-world web image database from National University of Singapore. In *CIVR '09*, 2009.
- [CX16] Guo Chen and Lu Xiao. Selecting publication keywords for domain analysis in bibliometrics : A comparison of three methods. *Journal of Informetrics*, 10(1) :212–223, 2016.
- [CXZ⁺20] Peng Chen, Qingxin Xiao, Jun Zhang, Chengjun Xie, and Bing Wang. Occurrence prediction of cotton pests and diseases by bidirectional long short-term memory networks with climate and atmosphere circulation. *Computers and Electronics in Agriculture*, 176 :105612, September 2020.
- [CZ14] C. Zhang and Z. Zhang. Improving multiview face detection with multi-task deep convolutional neural networks. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1036–1041, March 2014.
- [CZC⁺17] Xi Cheng, Youhua Zhang, Yiqiong Chen, Yunzhi Wu, and Yi Yue. Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, 141 :351–356, September 2017.
- [DB09] Nicola De Bellis. *Bibliometrics and citation analysis : from the science citation index to cybermetrics*. scarecrow press, 2009.
- [DBD16] A. Dey, Debasmita Bhounik, and K. N. Dey. Automatic Detection of Whitefly Pest using Statistical Feature Extraction and Image Classification Methods. 2016.
- [DCDR18] Steven Debaere, Kristof Coussement, and Tom De Ruyck. Multi-label classification of member participation in online innovation communities. *European Journal of Operational Research*, 270(2) :761–774, October 2018.
- [DCP⁺19] Jingcheng Du, Qingyu Chen, Yifan Peng, Yang Xiang, Cui Tao, and Zhiyong Lu. ML-Net : Multi-label classification of biomedical texts with deep neural networks. *Journal of the American Medical Informatics Association : JAMIA*, June 2019.
- [Dem06] Janez Demšar. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, pages 1–30, January 2006.
- [DFSC19] Bharath Bhushan Damodaran, Rémi Flamary, Vivien Seguy, and Nicolas Courty. An Entropic Optimal Transport loss for learning deep neural networks under label noise in remote sensing images. *Computer Vision and Image Understanding*, page 102863, November 2019.
- [DK17] Finale Doshi-Velez and Been Kim. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv :1702.08608 [cs, stat]*, March 2017.
- [DKC⁺21] P. Deepalakshmi, Prudhvi T. Krishna, Siri S. Chandana, K. Lavanya, and Parvathaneni Naga Srinivasu. Plant leaf disease detection using CNN algorithm. *International Journal of Information System Modeling and Design*, 12(1, SI) :1–21, January 2021.
- [DLJ⁺19] Wang Dawei, Deng Limiao, Ni Jiangong, Gao Jiyue, Zhu Hongfei, and Han Zhongzhi. Recognition pest by image-based transfer learning. *Journal of the Science of Food and Agriculture*, 99(10) :4524–4531, August 2019.

- [DR20] Arun Das and Paul Rad. Opportunities and Challenges in Explainable Artificial Intelligence (XAI) : A Survey. *arXiv :2006.11371 [cs]*, June 2020.
- [DSB17] Derek Doran, Sarah Schulz, and Tarek R. Besold. What Does Explainable AI Really Mean? A New Conceptualization of Perspectives. *arXiv :1710.00794 [cs]*, October 2017.
- [EC21] Benjamin Ejzenberg and Anna Choury. Objectif ia : initiez-vous à l’intelligence artificielle, 2021.
- [EMA⁺20] Borja Espejo-Garcia, Nikos Mylonas, Loukas Athanasakos, Spyros Fountas, and Ioannis Vasilakoglou. Towards weeds identification assistance through transfer learning. *Computers and Electronics in Agriculture*, 171, April 2020.
- [EVGW⁺07] Mark. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. 2007.
- [EVGW⁺12] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012). 2012.
- [FAO20] FAO. Plant pests and diseases : FAO in Emergencies. <http://www.fao.org/emergencies/emergency-types/plant-pests-and-diseases/en/>, 2020.
- [FK18] G.A. Fotso Kanga. A deep heterogeneous feature fusion approach for automatic land-use classification. *Information Sciences, Issue*, 467 :199–218, 2018.
- [Fou20] Fotis Foukalas. Cognitive IoT platform for fog computing industrial applications. *Computers & Electrical Engineering*, 87 :106770, October 2020.
- [Fuk80] Kuniyuki Fukushima. Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4) :193–202, April 1980.
- [FYKP17] Alvaro Fuentes, Sook Yoon, Sang Kim, and Dong Sun Park. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. *Sensors*, 17(9) :2022, September 2017.
- [FYLP18] Alvaro F. Fuentes, Sook Yoon, Jaesu Lee, and Dong Sun Park. High-Performance Deep Neural Network-Based Tomato Plant Diseases and Pests Diagnosis System With Refinement Filter Bank. *Frontiers in Plant Science*, 9 :1162, August 2018.
- [GA19] David Gunning and David Aha. Darpa’s explainable artificial intelligence (xai) program. *AI Magazine*, 40(2) :44–58, Jun. 2019.
- [Gan05] Jean-Marc A Gandonou. Essays on precision agriculture technology adoption and risk management. volume 227. University of Kentucky Doctoral Dissertations, 2005.
- [GBS⁺18] Sambuddha Ghosal, David Blystone, Asheesh K. Singh, Baskar Ganapathysubramanian, Arti Singh, and Soumik Sarkar. An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, 115(18) :4613–4618, May 2018.
- [GBY⁺18] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining Explanations : An Overview of Interpretability of Machine Learning. pages 80–89, May 2018.

- [GKCA17] Kasthurirangan Gopalakrishnan, Siddhartha K. Khaitan, Alok Choudhary, and Ankit Agrawal. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials*, 157 :322–330, December 2017.
- [GOO⁺18] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, Pablo Martinez-Gonzalez, and Jose Garcia-Rodriguez. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, 70 :41–65, September 2018.
- [GULG16] Guillermo L. Grinblat, Lucas C. Uzal, Mónica G. Larese, and Pablo M. Granitto. Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, 127 :418–424, September 2016.
- [Gup19] Jyoti Gupta. The Role of Artificial intelligence in Agriculture Sector. <https://customerthink.com/the-role-of-artificial-intelligence-in-agriculture-sector/>, October 2019.
- [GV15] Eva Gibaja and Sebastián Ventura. A Tutorial on Multilabel Learning. *ACM Comput. Surv.*, 47(3) :52 :1–52 :38, April 2015.
- [GXF] Yuanyue Ge, Ya Xiong, and Pål J. From. Instance Segmentation and Localization of Strawberries in Farm Conditions for Automatic Fruit Harvesting. 52(30) :294–299.
- [GZY⁺20] Yan Guo, Jin Zhang, Chengxin Yin, Xiaonan Hu, Yu Zou, Zhipeng Xue, and Wei Wang. Plant Disease Identification Based on Deep Learning Algorithm in Smart Farming. *Discrete Dynamics in Nature and Society*, 2020 :1–11, August 2020.
- [HDN19] Lam Si Tung Ho, Vu Dinh, and Cuong V. Nguyen. Multi-task learning improves ancestral state reconstruction. *Theoretical Population Biology*, 126 :33–39, April 2019.
- [Hir05] J. E. Hirsch. An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences*, 102(46) :16569–16572, November 2005.
- [HLvW18] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. *arXiv :1608.06993 [cs]*, January 2018.
- [HLW⁺17] Jun Huang, Guorong Li, Shuhui Wang, Zhe Xue, and Qingming Huang. Multi-label classification by exploiting local positive and negative pairwise label correlation. *Neurocomputing*, 257 :164–174, September 2017.
- [HRDR21] Waleej Haider, Aqeel-Ur Rehman, Nouman M. Durrani, and Sadiq Ur Rehman. A generic approach for wheat disease classification and verification using expert opinion for knowledge-based decisions. *IEEE Access*, 9 :31104–31129, 2021.
- [HS15] David P. Hughes and Marcel Salathe. An open access repository of images on plant health to enable the development of mobile disease diagnostics. November 2015.
- [HYG⁺19] Zhi-Fen He, Ming Yang, Yang Gao, Hui-Dong Liu, and Yilong Yin. Joint multi-label classification and label correlations with missing labels and feature selection. *Knowledge-Based Systems*, 163 :145–158, January 2019.

- [HZC⁺17] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. MobileNets : Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv :1704.04861 [cs]*, April 2017.
- [HZRS15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv :1512.03385 [cs]*, December 2015.
- [IL65] Alekseï Grigor'evich Ivakhnenko and Valentin Grigorévich Lapa. *Cybernetic Predicting Devices*. CCM Information Corporation, 1965.
- [Iva68] A. G. Ivakhnenko. The Group Method of Data of Handling ; A rival of the method of stochastic approximation. *Soviet Automatic Control*, 13 :43–55, 1968.
- [Iva71] A. G. Ivakhnenko. Polynomial theory of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-1(4) :364–378, 1971.
- [Kan58] L. Kantorovitch. On the Translocation of Masses. *Management Science*, 5(1) :1–4, October 1958.
- [KHC20] Mohamed Kerkech, Adel Hafiane, and Raphael Canals. Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Computers and Electronics in Agriculture*, 174, July 2020.
- [KIP⁺08] Johanna Koskinen, Matti Isohanni, Henna Paajala, Erika Jääskeläinen, Pentti Nieminen, Hannu Koponen, Pekka Tienari, and Jouko Miettunen. How to use bibliometric methods in evaluation of scientific research ? an example from finnish schizophrenia research. *Nordic journal of psychiatry*, 62 :136–43, 02 2008.
- [KLSS17] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 14(5) :778–782, May 2017.
- [KP18] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture : A survey. *Computers and Electronics in Agriculture*, 147 :70–90, April 2018.
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [LBBH98] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, November 1998.
- [LBD⁺89] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4) :541–551, December 1989.
- [LBM⁺18] Konstantinos Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. Machine Learning in Agriculture : A Review. *Sensors*, 18(8) :2674, August 2018.
- [LBMS18] Philipp Lottes, Jens Behley, Andres Milioto, and Cyrill Stachniss. Fully convolutional networks with sequential information for robust crop and weed detection

- in precision farming. *IEEE Robotics and Automation Letters*, 3(4) :2870–2877, Oct 2018.
- [LBSM⁺20] Marjorie Le Bars, Fatoumata Sidibe, Elisabeth Mandart, Jacques Fabre, Philippe Le Grusse, and Cheick Hamalla Diakite. évaluation des risques liés à l’utilisation de pesticides en culture cotonnière au Mali. *Cahiers Agricultures*, 29 :4, 2020.
- [LKAL16] Martin Längkvist, Andrey Kiselev, Marjan Alirezaie, and Amy Loutfi. Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks. *Remote Sensing*, 8(4) :329, April 2016.
- [LKF10] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pages 253–256, May 2010.
- [LMR17] Khalil Laghmari, Christophe Marsala, and Mohammed Ramdani. Classification multi-labels graduée : Apprendre les relations entre les labels ou limiter la propagation d’erreur ? In *EGC*, 2017.
- [LW20] Jun Liu and Xuewei Wang. Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. *Frontiers in Plant Science*, 11 :898, June 2020.
- [LWD⁺20] Yanfen Li, Hanxiang Wang, L. Minh Dang, Abolghasem Sadeghi-Niaraki, and Hyeonjoon Moon. Crop pest recognition in natural scenes using convolutional neural networks. *Computers and Electronics in Agriculture*, 169 :105174, February 2020.
- [LWL⁺19] Hefei Ling, Ziyang Wang, Ping Li, Yuxuan Shi, Jiazhong Chen, and Fuhao Zou. Improving person re-identification by multi-task learning. *Neurocomputing*, 347 :109–118, June 2019.
- [LWW⁺21] Zheng Li, Chaofeng Wang, Jun Wang, Shihui Ying, and Jun Shi. Lightweight adaptive weighted network for single image super-resolution. *Computer Vision and Image Understanding*, 211 :103254, 2021.
- [LWZ⁺19] Rui Li, Rujing Wang, Jie Zhang, Chengjun Xie, Liu Liu, Fangyuan Wang, Hongbo Chen, Tianjiao Chen, Haiying Hu, Xiufang Jia, Min Hu, Man Zhou, Dengshan Li, and Wancai Liu. An effective data augmentation strategy for cnn-based pest localization and recognition in the field. *IEEE Access*, 7 :160274–160283, 2019.
- [LYG⁺10] Guo-Zheng Li, Mingyu You, Lei Ge, Jack Y. Yang, and Mary Qu Yang. Feature Selection for Semi-supervised Multi-label Learning with Application to Gene Function Analysis. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology*, BCB ’10, pages 354–357, New York, NY, USA, 2010. ACM.
- [LYZ⁺17] Yang Lu, Shujuan Yi, Nianyin Zeng, Yurong Liu, and Yong Zhang. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, 267 :378–384, December 2017.
- [MC17] Maryam Rahnemoonfar and Clay Sheppard. Deep Count : Fruit Counting Based on Deep Simulated Learning. *Sensors*, 17(4) :905, April 2017.

- [MCDG21] Omid Mohamad Nezami, Akshay Chaturvedi, Mark Dras, and Utpal Garain. Pick-object-attack : Type-specific adversarial attack for object detection. *Computer Vision and Image Understanding*, 211 :103257, 2021.
- [MGCV18] Jose M. Moyano, Eva L. Gibaja, Krzysztof J. Cios, and Sebastián Ventura. Review of ensembles of multi-label classifiers : Models, experimental study and prospects. *Information Fusion*, 44 :33–45, November 2018.
- [MHS16] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, 7 :1419, September 2016.
- [MLY⁺17] Andrew Maxwell, Runzhi Li, Bei Yang, Heng Weng, Aihua Ou, Huixiao Hong, Zhaoxian Zhou, Ping Gong, and Chaoyang Zhang. Deep learning architectures for multi-label classification of intelligent health risk prediction. *BMC Bioinformatics*, 18(S14), December 2017.
- [NGMN19] Manuel Nunes, Enrico Gerding, Frank McGroarty, and Mahesan Niranjan. A comparison of multitask and single task learning with artificial neural networks for yield curve forecasting. *Expert Systems with Applications*, 119 :362–375, April 2019.
- [NJS⁺19] Koushik Nagasubramanian, Sarah Jones, Asheesh K. Singh, Soumik Sarkar, Arti Singh, and Baskar Ganapathysubramanian. Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods*, 15(1), August 2019.
- [Ots79] Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1) :62–66, January 1979.
- [PAW⁺17] Michael P. Pound, Jonathan A. Atkinson, Darren M. Wells, Tony P. Pridmore, and Andrew P. French. Deep Learning for Multi-task Plant Phenotyping. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2055–2063, Venice, Italy, October 2017. IEEE.
- [PKA19] Victor Partel, Charan Kakarla, and Yiannis Ampatzidis. Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. *Computers and Electronics in Agriculture*, 157 :339–350, February 2019.
- [Pla17] Planet : Understanding the amazon from space, 2017.
- [PPZM18] Rafael B. Pereira, Alexandre Plastino, Bianca Zadrozny, and Luiz H. C. Merschmann. Correlation analysis of performance measures for multi-label classification. *Information Processing & Management*, 54(3) :359–369, May 2018.
- [PSA⁺19] Artzai Picon, Maximilian Seitz, Aitor Alvarez-Gila, Patrick Mohnke, Amaia Ortiz-Barredo, and Jone Echazarra. Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Computers and Electronics in Agriculture*, 167, December 2019.
- [PVG⁺11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos,

- D. Courneau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.
- [PY10] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. *IEEE Trans. on Knowl. and Data Eng.*, 22(10) :1345–1359, October 2010.
- [RBM⁺17] Amanda Ramcharan, Kelsea Baranowski, Peter McCloskey, Babuali Ahmed, James Legg, and David Hughes. Using Transfer Learning for Image-Based Cassava Disease Detection. *Frontiers in Plant Science*, 8 :1852, October 2017.
- [RCFT18] Ievgen Redko, Nicolas Courty, Rémi Flamary, and Devis Tuia. Optimal Transport for Multi-source Domain Adaptation under Target Shift. *arXiv :1803.04899 [stat]*, March 2018.
- [RDS⁺14] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *arXiv :1409.0575 [cs]*, September 2014.
- [RDS⁺15] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3) :211–252, December 2015.
- [Res17] Research and Markets. AI in Agriculture Market 2017-2025 - Focus on Precision Farming, Drone Analytics, Agriculture Robots & Livestock Monitoring - Research and Markets. <https://www.businesswire.com/news/home/20171221005505/en/AI-in-Agriculture-Market-2017-2025—Focus-on-Precision-Farming-Drone-Analytics-Agriculture-Robots-Livestock-Monitoring—Research-and-Markets>, December 2017.
- [RFFT17] Paulo E. Rauber, Samuel G. Fadel, Alexandre X. Falcao, and Alexandru C. Telea. Visualizing the Hidden Activity of Artificial Neural Networks. *IEEE Transactions on Visualization and Computer Graphics*, 23(1) :101–110, January 2017.
- [RHGS15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn : Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28 :91–99, 2015.
- [Ros57] F. Rosenblatt. The perceptron - A perceiving and recognizing automaton. Technical Report 85-460-1, Cornell Aeronautical Laboratory, Ithaca, New York, January 1957.
- [RPR18] Aravind Krishnaswamy Rangarajan, Raja Purushothaman, and Aniiirudh Ramesh. Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Computer Science*, 133 :1040–1047, 2018.
- [RSG16] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?" : Explaining the Predictions of Any Classifier. *arXiv :1602.04938 [cs, stat]*, August 2016.

- [RSG18] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors : High-precision model-agnostic explanations. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.
- [Rud17] Sebastian Ruder. An Overview of Multi-Task Learning in Deep Neural Networks. *arXiv :1706.05098 [cs, stat]*, June 2017.
- [RvHB⁺20] Thijs Ruigrok, Eldert van Henten, Johan Booij, Koen van Boheemen, and Gert Kootstra. Application-specific evaluation of a weed-detection algorithm for plant-specific spraying. *Sensors*, 20(24), 2020.
- [RZ18] Bharath Ramsundar and Reza Bosagh Zadeh. *TensorFlow for Deep Learning : From Linear Regression to Reinforcement Learning*. O'Reilly Media, Beijing, first edition edition, 2018.
- [SCD⁺16] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM : Visual Explanations from Deep Networks via Gradient-based Localization. *arXiv :1610.02391 [cs]*, October 2016.
- [SCJJ19] Uday Pratap Singh, Siddharth Singh Chouhan, Sukirty Jain, and Sanjeev Jain. Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease. *IEEE Access*, 7 :43721–43729, 2019.
- [SDMC20] Arun Kumar Sangaiah, Jerline Sheebha Anni Dhanaraj, Prabu Mohandas, and Aniello Castiglione. Cognitive IoT system with intelligence techniques in sustainable computing environment. *Computer Communications*, 154 :347–360, March 2020.
- [SGAS17] A. Shahin, Y. Guo, K. Amin, and A.A. Sharawi. White blood cells identification system based on convolutional deep neural learning networks. In *Computer Methods and Programs in Biomedicine*, pages 1–12. 2017.
- [SGSK17] Avanti Shrikumar, Peyton Greenside, Anna Shcherbina, and Anshul Kundaje. Not Just a Black Box : Learning Important Features Through Propagating Activation Differences. *arXiv :1605.01713 [cs]*, April 2017.
- [SHK⁺14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15 :1929–1958, 2014.
- [SLJ⁺14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions. *arXiv :1409.4842 [cs]*, September 2014.
- [SM19] Wojciech Samek and Klaus-Robert Müller. Towards explainable artificial intelligence. *Lecture Notes in Computer Science*, page 5–22, 2019.
- [SS19] Y. Seo and K.-s Shin. Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with Applications*, February. *Volume*, 116 :328–339, 2019.
- [SSC17] Y. Shiqi, J. Sen, and X. Chunyan. Convolutional neural networks for hyperspectral image classification. *Neurocomputing*, 219 :88–98, 2017.
- [SSL20] Amitojdeep Singh, Sourya Sengupta, and Vasudevan Lakshminarayanan. Explainable Deep Learning Models in Medical Image Analysis. *Journal of Imaging*, 6(6) :52, June 2020.

- [STY17] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic Attribution for Deep Networks. *arXiv :1703.01365 [cs]*, June 2017.
- [SVI⁺15] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. *arXiv :1512.00567 [cs]*, December 2015.
- [SVZ14] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep Inside Convolutional Networks : Visualising Image Classification Models and Saliency Maps. *arXiv :1312.6034 [cs]*, April 2014.
- [SWM17] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable Artificial Intelligence : Understanding, Visualizing and Interpreting Deep Learning Models. *arXiv :1708.08296 [cs, stat]*, August 2017.
- [SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*, 2014.
- [TBSN12] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja. Robust visual tracking via multi-task sparse learning. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2042–2049, June 2012.
- [TG20] Erico Tjoa and Cuntai Guan. A Survey on Explainable Artificial Intelligence (XAI) : Towards Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2020.
- [TIT⁺21] Muhammad Tufail, Javaid Iqbal, Mohsin Islam Tiwana, Muhammad Shahab Alam, Zubair Ahmad Khan, and Muhammad Tahir Khan. Identification of tobacco crop based on machine learning for a precision agricultural sprayer. *IEEE Access*, 9 :23814–23825, 2021.
- [TK07] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification : An overview. In *Int J Data Warehousing and Mining*, 2007.
- [TN07] D. V. Tran and N. Nguyen. The concept and implementation of precision farming and rice integrated crop management systems for sustainable production in the twenty-first century. 2007.
- [TR19] K. Thenmozhi and U. Srinivasulu Reddy. Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, 164, September 2019.
- [VCS20] Shradha Verma, Anuradha Chug, and Amit Prakash Singh. Exploring capsule networks for disease classification in plants. *Journal of Statistics and Management Systems*, 23(2) :307–315, 2020.
- [Vil09] Cédric Villani. *Optimal Transport : Old and New*. Number 338 in Grundlehren Der Mathematischen Wissenschaften. Springer, Berlin, 2009.
- [vW10] Nees Jan van Eck and Ludo Waltman. Software survey : VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2) :523–538, August 2010.
- [Was18] Ritika Wason. Deep learning : Evolution and expansion. *Cognitive Systems Research*, 52 :701–708, December 2018.

- [WB17] J. Wehrmann and R.C. Barros. Movie genre classification : A multi-label approach based on convolutions through time. *Applied Soft Computing, Volume*, 61 :973–982, 2017.
- [WGVB17] Sjaak Wolfert, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. Big Data in Smart Farming – A review. *Agricultural Systems*, 153 :69–80, May 2017.
- [WGZ18] A. Wosiak, K. Glinka, and D. Zakrzewska. Multi-label classification methods for improving comorbidities identification. *Computers in Biology and Medicine*, 100 :279–288, September 2018.
- [WLW⁺18] Wei Weng, Yaojin Lin, Shunxiang Wu, Yuwen Li, and Yun Kang. Multi-label learning based on label-specific features and local pairwise label correlation. *Neurocomputing*, 273 :385–394, January 2018.
- [WLX⁺21] Rujing Wang, Liu Liu, Chengjun Xie, Po Yang, Rui Li, and Man Zhou. Agri-Pest : A Large-Scale Domain-Specific Benchmark Dataset for Practical Agricultural Pest Detection in the Wild. *Sensors*, 21(5) :1601, February 2021.
- [WRH17] Sheng Wang, Ashwin Raju, and Junzhou Huang. Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 620–623, Melbourne, Australia, April 2017. IEEE.
- [WWX⁺20] Fangyuan Wang, Rujing Wang, Chengjun Xie, Po Yang, and Liu Liu. Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition. *Computers and Electronics in Agriculture*, 169, February 2020.
- [WYZZ09] Changhu Wang, Shuicheng Yan, Lei Zhang, and Hong-Jiang Zhang. Multi-label sparse coding for automatic image annotation. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1643–1650, 2009.
- [WZL⁺19] Xiaoping Wu, Chi Zhan, Yukun Lai, Ming-Ming Cheng, and Jufeng Yang. IP102 : A Large-Scale Benchmark Dataset for Insect Pest Recognition. In *IEEE CVPR*, pages 8787–8796, 2019.
- [XFLH17] Z. Xia, X. Feng, J. Lin, and A. Hadid. Deep convolutional hashing using pairwise multi-label supervision for large-scale visual search. In *Signal Processing : Image Communication, Volume*, volume 59, pages 109–116. 2017.
- [Xin20] Xinhuanews. [http ://french.xinhuanet.com/2020-01/20/c_138720980.htm](http://french.xinhuanet.com/2020-01/20/c_138720980.htm), January 2020.
- [XRvD20] Ning Xie, Gabrielle Ras, Marcel van Gerven, and Derek Doran. Explainable Deep Learning : A Field Guide for the Uninitiated. *arXiv :2004.14545 [cs, stat]*, April 2020.
- [YGYN14] Kyosuke Yamamoto, Wei Guo, Yosuke Yoshioka, and Seishi Ninomiya. On Plant Detection of Intact Tomato Fruits Using Image Analysis and Machine Learning Methods. *Sensors*, 14(7) :12191–12206, July 2014.
- [YL18] Xi Yin and Xiaoming Liu. Multi-Task Convolutional Neural Network for Pose-Invariant Face Recognition. *IEEE Transactions on Image Processing*, 27(2) :964–975, February 2018.

- [YLCL18] Yao-Yuan Yang, Yi-An Lin, Hong-Min Chu, and Hsuan-Tien Lin. Deep Learning with a Rethinking Structure for Multi-label Classification. *arXiv :1802.01697 [cs, stat]*, February 2018.
- [YSLC12] Xiaolu Yang, Xuanjing Shen, Jianwu Long, and Haipeng Chen. An Improved Median-based Otsu Image Thresholding Algorithm. *AASRI Procedia*, 3 :468–473, January 2012.
- [YXA18] Xiaohui Yuan, Lijun Xie, and Mohamed Abouelenien. A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. *Pattern Recognition*, 77 :160–172, May 2018.
- [ZF13] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. *arXiv :1311.2901 [cs]*, November 2013.
- [ZF14] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8689, pages 818–833. Springer International Publishing, Cham, 2014.
- [ZHZ19] Liheng Zhong, Lina Hu, and Hang Zhou. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221 :430–443, February 2019.
- [Zim08] Cindy Zimmerman. The Five “R’s” of Precision — Precision. <https://precision.agwired.com/2008/11/11/the-five-rs-of-precision/>, 2008.
- [ZLB97] P. Zwaenepoel and J.M. Le Bars. L’agriculture de précision. *Ingénieries eau-agriculture-territoires*, (12) :p. 67 – p. 79, 1997.
- [ZLL⁺18] Nanyang Zhu, Xu Liu, Ziqian Liu, Kai Hu, Yingkuan Wang, Jinglu Tan, Min Huang, Qibing Zhu, Xunsheng Ji, Yongnian Jiang, Ya Guo, 1. Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi 214122, China, 2. School of Internet of Things, Jiangnan University, Wuxi 214122, China, 3. Chinese Academy of Agricultural Engineering, Beijing 100125, China, 4. Department of Bioengineering, University of Missouri, Columbia, MO 65211, USA, and 5. Jiangsu Zhongnong IoT Technology Co., LTD, Yixing 214200, China. Deep learning for smart agriculture : Concepts, tools, applications, and opportunities. *International Journal of Agricultural and Biological Engineering*, 11(4) :21–28, 2018.
- [ZLLL17] Jianqing Zhu, Shengcai Liao, Zhen Lei, and Stan Z. Li. Multi-label convolutional neural network based pedestrian attribute classification. *Image and Vision Computing*, 58 :224–229, February 2017.
- [ZSJS18] Jiang Zhu, Yonghui Song, Dingde Jiang, and Houbing Song. A New Deep-Q-Learning-Based Transmission Scheduling Mechanism for the Cognitive Internet of Things. *IEEE Internet of Things Journal*, 5(4) :2375–2385, August 2018.
- [ZYC⁺18] Ni Zhuang, Yan Yan, Si Chen, Hanzi Wang, and Chunhua Shen. Multi-label Learning Based Deep Transfer Neural Network for Facial Attribute Classification. *Pattern Recognition*, 80 :225–240, August 2018.