

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

المدرسة الوطنية العليا للفلاحة

Ecole Nationale Supérieure Agronomique El-Harrach -Alger

THESE

En vue de l'obtention du diplôme de Doctorat en Sciences Agronomiques

Spécialité : Machinisme Agricole

THEME

**Intégration des techniques de la vision artificielle dans
l'identification des grains de quelques variétés de blé en
Algérie**

Présentée et soutenue publiquement par M. LAABASSI Karim

à l'ENSA le: 31/05/2022

Devant le jury composé de :

MEFTI Mohamed	Professeur, ENSA	Président
KACI Ferhat	Professeur, ENSA	Directeur de thèse
BENKHERBACHE Nadjat	MCA, Université de M'sila	Examineur
DJEFFAL Abdelhamid	Professeur, Université de Biskra	Examineur
MAHMOUDI Said	Professeur, UMONS, Belgique	Invité

Année universitaire 2021-2022

Remerciements

Je tiens avant tout à remercier Dieu le tout puissant de m'avoir donné la force et la volonté pour achever ce modeste travail.

Mes remerciements vont également à Mr KACI Ferhat, Professeur à l'ENSA d'avoir accepté de m'encadrer, pour son aide, son encouragement et son soutien tout au long de la réalisation de ce travail, et aussi pour sa grande patience.

Je remercie Mr MEFTI Mohamed, Professeur à l'ENSA d'avoir accepté de présider le jury de soutenance.

Je remercie Docteur BENKHERBACHE Nadjat (Université de M'Sila), Professeur DJEFFAL Abdelhamid (Université de Biskra) et Professeur MAHMOUDI Saïd (Université de MONS, Belgique) d'avoir accepté d'évaluer ce travail de recherche et faire partie du jury.

Je remercie tous mes amis et collègues qui m'ont prêté main et aidé pour réaliser ce travail, je cite sans exclusion : *Pr Ettoumi Y.F, Dr Claes G, Dr Chebli A, Dr Ferroukhi Sd, Dr Boukraa S, Pr Khelifi L, Dr Omari C, Dr Brahim M, Dr Laraba S, Dr Mokrane K, Dr Benaliya S, Dr Etsouri S, Dr Delli R, Dr Laribi, Mr Goucem S. L'équipe de l'U-MONS : Dr Mahmoudi Sd, Dr Belarbi MA, Dr Debouch O. ; Ing Amiali S. (CNCC), Dr Kherif O (ITGC), Mr LAABASSI A., Mr LAABASSI Mohamed, Mr. LAABASSI Md S, Mr. Adels D'Bichi et Mr. ChemsEddine Laagoune.*

Je remercie aussi ma famille, ma mère, mes frères et mes sœurs, ma femme et mes petits enfants pour leur soutiens et motivation.

Je remercie tous les enseignants et les travailleurs de l'ENSA.

Je remercie toutes les personnes professionnelles, au niveau de tous les centres et les instituts techniques qui travaillent pour aboutir à un meilleur présent et futur de l'agriculture algérienne.

Je remercie aussi toute personne ayant contribué de près ou de loin à la réalisation de ce travail.

Dédicaces

A ma famille :

Abla, Yahia El-Cherif et Amina Melak

A la mémoire de mon père

A Ma Mère et à mes Frères et ma Sœur

@ mes amis.

A mon Pays.

.

Karim LAABASSI

[وَعِنْدَهُ مَفَاتِحُ الْغَيْبِ لَا يُعْلَمُهَا إِلَّا هُوَ وَيَعْلَمُ مَا فِي الْبَرِّ وَالْبَحْرِ وَمَا تَسْقُطُ

مِنَ وَرَقَةٍ إِلَّا يُعْلَمُهَا وَلَا حَبَّةٍ فِي ظُلُمَاتِ الْأَرْضِ وَلَا رَطْبٍ وَلَا يَابِسٍ إِلَّا فِي

كِتَابٍ مُّبِينٍ] (59) سورة الانعام

Résumé

Les solutions technologiques et l'innovation sont un levier incontournable pour le développement agricole. Elles assurent l'une des voies les plus efficaces pour accroître et pour maintenir la production à des niveaux quantitatifs et qualitatifs acceptables, ceci par la maîtrise technologique des différents aspects de l'agriculture. On assiste depuis le début de cette dernière décennie à l'ère de la transition technologique et digitale qui est pilotée par l'intelligence artificielle ; cette dernière a permis de générer des solutions technologiques à des problèmes agronomiques complexes, tout en augmentant l'efficacité du machinisme et de l'automatisation conventionnelle.

La maîtrise technologique a pris aujourd'hui, grâce à des systèmes de contrôle intelligent, un nouveau cadre conceptuel. Elle est basée sur la capture et l'exploration des quantités massives de données et l'implémentation des solutions technologiques basées sur l'intelligence artificielle. Le cap des processus de production est fixé sur une démarche de qualité générative de produits traçables.

Ce travail est un exemple de l'élaboration et l'implémentation d'une solution technologique pour le problème de la reconnaissance de l'identité des grains de blé, qui a permis par conséquent de réaliser l'identification variétale dans toute la chaîne des grains.

Nous avons redéfini ce problème de l'identification des grains de blé d'un contexte de perception humaine vers un contexte de perception machine avec la technique de classification d'images en exploitant une solution basée sur la Computer Vision (CV) et Le machine Learning (ML).

Cette solution technologique nous a permis de tester deux approches de classification d'images (la classique et la Deep learning) et d'aboutir, par une approche adaptée à l'Intégration des techniques de la vision artificielle dans l'identification des grains de quelques variétés de blé, à un résultat satisfaisant (95-98% de précision) par l'approche de Deep Learning.

Mots clés : Blé ; Chaîne De Grain ; Identification Variétale ; Machine Learning ; Computer Vision ; Deep Learning ; Vision Artificielle ; Solution Technologique.

Abstract

Technological solutions and innovation are an essential lever for agricultural development. They provide one of the most effective ways to increase and maintain production at acceptable quantitative and qualitative levels, through technological mastery of the various aspects of agriculture. Since the beginning of this last decade, we have been witnessing the era of technological and digital transition which is driven by artificial intelligence; the latter has made it possible to generate technological solutions to complex agronomic problems, while increasing the efficiency of conventional machinery and automation.

Technological mastery has taken today, thanks to intelligent control systems, a new conceptual framework. It is based on the capture and exploration of massive amounts of data and the implementation of technological solutions based on artificial intelligence. The course of production processes is set on a generative quality approach for traceable products.

This work is an example of the development and implementation of a technological solution for the problem of wheat grain identity recognition which consequently allowed varietal identification throughout the grain chain.

We have redefined this problem of wheat grain identification from a human perception context to a machine perception context with the image classification technique by exploiting a solution based on Computer Vision (CV) and Machine Learning (ML).

This technological solution allowed us to test two image classification approaches (Classic and Deep Learning) and to succeed, through an approach adapted to the Integration of artificial vision techniques in the identification of grains of a few wheat varieties, with a satisfactory result (95-98% accuracy) by the Deep Learning approach.

Keywords: Wheat; grain chain; varietal identification; Machine Learning; Computer Vision; Deep Learning; Artificial Vision; Technological Solution.

تلخيص

إن الحلول التكنولوجية والابتكار روافع أساسية لتطوير الفلاحة في منظورها الشامل، إذ تمثل أكثر السبل نجاعة لزيادة وإدارة الإنتاج الفلاحي وتثبيتته في مستويات مقبولة، آمنة وصامدة في وجه تحديات الامن الغذائي في كنف وتعقيدات التنمية المستدامة. ويتم تجسيد هذا الشأن عن طريق الاستفادة من التحكم التكنولوجي في جميع جوانب الفلاحة وخاصة في سلاسل الإنتاج ذات الصلة بها. لأجل ذلك نشهد منذ بداية العقد الماضي عصر التحول التكنولوجي والرقمي المرتكز على الذكاء الاصطناعي، فقد أتاح هذا الأخير توجهها جديدا وإمكانية فعلية لإنشاء حلول ناجعة لمشكلات فلاحية أكثر تعقيدا من ذي قبل مع زيادة كفاءة ودقة الآلية والالتمة التقليدية بتقنيات الرؤية الحاسوبية والتعلم الآلي.

في هذا النسق نتناول في اطروحتنا دراسة إمكانية ادماج وتكيف تقنيات الرؤية الحاسوبية والتعلم الآلي في إشكالية التعرف على هوية حبوب القمح من أجل تحديد الصنف المنتمية اليه، تعد عملية التعرف هذه اساسا لعدة تطبيقات معتمدة على الادراك البشري ولعل أهمها ما هو مرتبط بالجودة وعمليات المراقبة، التصديق والتصنيف في مختلف مستويات سلسلة انتاج الحبوب.

اثناء هذه الدراسة قمنا بتحديد وإعادة صياغة الاشكال المطروح بما يتوافق مع تقنية تصنيف الصور، تعد هذه الأخيرة أهم التقنيات المعتمد عليها في الذكاء الاصطناعي عن طريق تقنيات الرؤية الحاسوبية والتعلم الآلي؛ فقد سمحت لنا هذه الدراسة انشاء واعتماد منهجية تنفيذية للمقاربتين الأبرز في هذا المجال ويتعلق الامر بالمقاربة السطحية عن طريق الرؤية الحاسوبية والتعلم الآلي التقليديين، ومن جهة أخرى المقاربة العميقة عن طريق التعلم العميق. هذا الأخير أتاح لنا نتائج مرضية (بدقة بين 95 و 98%).

كلمات مفتاحية: قمح؛ سلسلة الحبوب؛ تحديد الأصناف؛ التعلم الآلي؛ الرؤية الحاسوبية؛ التعلم العميق؛ الرؤية الإصطناعية؛ حل تكنولوجي.

Liste des abréviations

AI	<i>Artificial Intelligence</i>
ANN	<i>Artificial Neural Network</i>
BW	<i>Bread Wheat</i>
CCLS	<i>Coopérative des céréales et légumes secs</i>
CF	<i>Colour Features</i>
CfsSubsetEval	<i>Correlation-based Feature Subset Selection</i>
CNCC	<i>Centre national de contrôle et de certification des semences et des plants</i>
CNN	<i>Convolutional Neural Network</i>
CP-NN	<i>Counter Propagation Neural Network</i>
CV	<i>Computer Vision</i>
DA	<i>Discriminant Analysis</i>
DL	<i>Deep learning</i>
DSIFT	<i>Dense Scale Invariant Features</i>
Faster R-CNN	<i>Faster Region Convolutional Neural Network</i>
GIR	<i>Grain Identity Recognition</i>
ICA	<i>Imperialist Competitive Algorithm</i>
ISTA	<i>International Seed Testing Association</i>
ITGC	<i>Institut technique des grandes culture</i>
K-NN	<i>K-nearest neighbours</i>
LDA	<i>linear discriminant analysis</i>
M.C.T.	<i>Features combinations (morphologic, color, texture)</i>
MF	<i>Morphological Features</i>
ML	<i>Machine learning</i>
MLF-NN	<i>multi-layer feed-forward Neural Network</i>
MLP	<i>Multi-Layer Perceptron</i>
NIR	<i>near infrared</i>
OCDE	<i>The Organization for Economic Co-operation and Development</i>
QDA	<i>Quadratic Discriminant Analysis</i>
SLC	<i>Specie Level Classification</i>
SMO	<i>Sequential Minimal Optimization</i>
SVM	<i>Support vector machine</i>
TF	<i>Texture Features</i>
UPOV	<i>International Union for the Protection of New Varieties of Plants</i>
VGA	<i>Visual Grain Analysis</i>
VLC	<i>Variety Level Classification</i>

Liste des figures

FIGURE 1: LE TEST DE PURETE AU LABORATOIRE	3
FIGURE 2: (A) IMAGES AU MICROSCOPE ELECTRONIQUE COLOREES ARTIFICIELLEMENT DE DIVERS GRAINS DE POLLEN POUR MONTRER LEUR CARACTERE DISTINCTIF ET LEUR VARIETE DE TAILLE ET DE FORME - LE POLLEN DE GRAMINEES EST ENERCLE. (B) CELLULES EPIDERMiques FOSSILISEES TROUVEES (WHEATBP,2021)	11
FIGURE 3: EVOLUTION DU BLE. (A) VOIE SIMPLIFIEE MENANT AU BLE CULTIVE. (B) ÉPIS ET GRAINES D'ESPECES DE BLE SELECTIONNEES. (VERGAUWEN ET DE SMET, 2017)	12
FIGURE 4: CYCLE BIOLOGIQUE DU BLE, ENTRE LE DEVELOPPEMENT ET LA CROISSANCE LA PLANTE EVOLUE DANS L'ESPACE ET DANS LE TEMPS. (AHDB 2021)	13
FIGURE 5:DIFFERENTES ECHELLES DE DEVELOPPEMENTS DE LA PLANTE DE BLE (HYLES ET AL., 2020)	14
FIGURE 6: LES STADES DE DEVELOPPEMENT SELON (ADBH, 2021)	16
FIGURE 7 : STRUCTURE D'UN GRAIN DE BLE MURE (A : COUPE LONGITUDUNALE, B : COUPE TRANSVERSALE)	17
FIGURE 8: DIAGRAMME DE CONSTITUTION DU GRAIN DE BLE	17
FIGURE 9: LES STADE DE DEVELOPPEMENT DU FLEURON (W : WEEKS = SEMAINES) (ZHENG ET AL., 2016).	19
FIGURE 10: ASPECT CLES DE LA FLORAISON DU BLE (WHEATBP, 2020)	19
FIGURE 11: UN FLEURON PENDANT LA PERIODE DE FLORAISON ET LES DEBUTS DE DEVELOPPEMENT DES GRAINS (OKADA ET AL., 2018)..	20
FIGURE 12: COUPE TRANSVERSALE DU GRAIN FRAIS AU DEBUT DU REMPLISSAGE DU GRAIN.....	21
FIGURE 13: COUPE LONGITUDINALE DU GRAIN FRAIS A TRAVERS LE SILLON VENTRAL.....	21
FIGURE 14: GRAIN DE BLE A LA FIN DU STADE LAITEUX	21
FIGURE 15: LES DIFFERENTES ENVELOPPES DU GRAIN AU STADE GS85	22
FIGURE 16: GRAIN DE BLE A LA FIN DU STADE PATEUX.....	22
FIGURE 17: GRAIN DE BLE AU STADE GS89	23
FIGURE 18 : UN EPILLET INDIVIDUEL 30 JOURS APRES LA POLLINISATION	23
FIGURE 19 : SURMATURITE DES GRAINS DE BLE DUR.....	24
FIGURE 20: DIMENSIONS AXIALES DANS UN GRAIN DE BLE (PONCE-GARCIA ET AL., 2017)	27
FIGURE 21: PRINCIPALES ETAPES DE LA CHAINE DES GRAINS	31
FIGURE 22: LES COMPOSANTS DE L'ECHANTILLON DANS L'ANALYSE DE PURETE CONFORMEMENT AUX REGLES DE L'AOSA	37
FIGURE 23: LES TROIS NIVEAU DE VISION PAR ORDINATEUR (ZAITOUN AND AQEL 2015)	46
FIGURE 24: OBJECTIFS ET OPERATIONS DE COMPUTER VISION (F.-F. LI, KARPATHY, AND JOHNSON 2016)	47
FIGURE 25: METHODES D'EXTRACTION DES ATTRIBUTS (MUTLAG ET AL. 2020).....	51
FIGURE 26: LES DOMAINES DE MACHINE LEARNING (KRZYK 2018)	53
FIGURE 27: BOTANISTS' (LEFT) VERSUS COMPUTER VISION (RIGHT) DESCRIPTION OF FLOWERS	55
FIGURE 28: LES ATTRIBUTS ET LA FORMATION DES NOUVELLES IMAGES EN DEEP LEARNING (SLADOJEVIC ET AL. 2016)	56
FIGURE 29: SHALLOW ET DEEP LEARNING APPROCHES ET L'APPLICATIONS POUR L'IDENTIFICATION DES PLANTES.....	57
FIGURE 30: EXEMPLE D'UN SCHEMA DE PHENOTYPAGE DES PLANTES (L. LI, ZHANG, AND HUANG 2014)	58
FIGURE 31: L'ARCHITECTURE GLOBALE DU RESEAU NEURONAL CONVOLUTIF (CNN). (ALOM ET AL., 2019)	62
FIGURE 32: FLUX DU TRAVAIL D'EXPERIMENTATION	68

FIGURE 33: FLUX DE TRAVAIL DE LA COLLECTE DES GRAINS (LAABASSI ET AL. 2021)	75
FIGURE 34: FLUX DE LA COLLECTE DES IMAGES DES GRAINS DE BLE	79
FIGURE 35 : DIFFERENTES SITUATIONS DE DISPERSION, LA QUELLE EST LA PLUS ADEQUATE	80
FIGURE 36: DIFFERENTES FACES DE LA GRAINE DE BLE, LAQUELLE FAUT-IL CHOISIR ?	81
FIGURE 37: UNE FEUILLE MILLIMETRIQUE POUR DEFINIR L'ECHELLE SUR LES IMAGES PRINCIPALES.....	82
FIGURE 38: ORGANISATION DES DOSSIERS DE LA BASE DE DONNEES IMAGES	83
FIGURE 39: IMAGES DES GRAINS DE CHAQUE VARIETE.....	84
FIGURE 40: PROTOCOLE EXPERIMENTALE SHALLOW	90
FIGURE 41: LES DIFFERENTES ETAPES DU FLUX DE TRAVAIL DE L'EXTRACTION DES ATTRIBUTS MORPHOLOGIQUES	91
FIGURE 42: REPRESENTATION GRAPHIQUE DU PROTOCOLE DE L'EXPERIMENTATION <i>DEEP LEARNING CLASSIFICATION</i>	118
FIGURE 43: PROGRESSION DU PROTOCOLE EXPERIMENTAL ET L'ENCHAINEMENT DE L'APPRENTISSAGE SUR 03 ANNEES.....	119
FIGURE 44: EVOLUTION DES PERFORMANCES DES CNN EN FONCTION DE LA QUANTITE DES DONNEES (ALOM ET AL. 2019)	119
FIGURE 45: DIAGRAMME SCHEMATIQUE DU MODELE INCEPTIONV3 (MAHDIANPARI ET AL. 2018).....	121
FIGURE 46 DIAGRAMME SCHEMATIQUE DU MODELE MOBILENET (HOWARD ET AL. 2017).....	121
FIGURE 47: DIAGRAMME SCHEMATIQUE DU MODELE XCEPTION (MAHDIANPARI ET AL. 2018).....	122
FIGURE 48: DIAGRAMME SCHEMATIQUE DU MODELE RESNET (MAHDIANPARI ET AL. 2018).....	122
FIGURE 49: DIAGRAMME SCHEMATIQUE DU MODELE DENSENET (MAHDIANPARI ET AL. 2018).....	122
FIGURE 50: PRINCIPE DE LA TECHNIQUE DE TRANSFERT LEARNING	123
FIGURE 51: EXTRAIT DU CODE DE TRANSFERT LEARNING POUR L'ARCHITECTURE INCEPTION V3.....	124
FIGURE 52 : EXTRAIT DE CODE DE INCEPTIONV3 ET MOBILENET (FINE TUNED).....	125
FIGURE 53: COURBES D'APPRENTISSAGES ET LES PERFORMANCES DES DIFFERENTS CNNs.....	127
FIGURE 54: LA VARIATION DES COURBE D'ACCURACY ET LOSS DU MODELE INCEPTIONV3 MODIFIE.....	130
FIGURE 55: LA VARIATION DES COURBE D'ACCURACY ET LOSS DU MODELE MOBILENET MODIFIE.....	130

Liste des tableaux

TABLEAU 1: LES PRINCIPAUX ASPECTS DE LA QUALITE NECESSITANT UNE EVALUATION A DIFFERENTS STADES DE LA CHAINE DE GRAIN (WRIGLEY, 2017).	32
TABLEAU 2: IDENTITE DU MATERIEL VEGETAL UTILISE	73
TABLEAU 3: INFORMATIONS RELATIVES AUX GRAINS COLLECTES	74
TABLEAU 4: ATTRIBUTS MORPHOLOGIQUES CALCULES RECENSES DE LA BIBLIOGRAPHIE	92
TABLEAU 5 : METHODES D'EVALUATION DES ATTRIBUTS	94
TABLEAU 6: METHODE DE RECHERCHE POUR LA SELECTION ET LE CLASSEMENT DES ATTRIBUTS	95
TABLEAU 7 : RESUME DES RESULTATS DE LA CLASSIFICATION DU NIVEAU « ESPECE »	96
TABLEAU 8: RESULTAT DE LA COMPARAISON ENTRE LES DIFFERENTS ALGORITHMES DE CLASSIFICATION (SLC)	96
TABLEAU 9: OPTIMISATION (FINE TUNING) DU MLPCLASSIFIER (SLC)	97
TABLEAU 10: OPTIMISATION (FINE TUNING) DU <i>FUNCTIONS.MULTILAYERPERCEPTRON</i> (SLC)	98
TABLEAU 11: FINE TUNING DU RANDOMFOREST (SLC)	98
TABLEAU 12: RESULTATS DETAILLES DE <i>FUNCTIONS.MPLCLASSIFIER</i>	99
TABLEAU 13: RESULTATS DETAILLES DE <i>FUNCTIONS.MULTILAYERPERCEPTRON</i>	99
TABLEAU 14: RESULTATS DETAILLES DE <i>TREE.RANDOMFOREST</i>	99
TABLEAU 15: MATRICES DE CONFUSION DES TROIS CLASSIFICATEURS (SLC-MANIPULATION A)	100
TABLEAU 16: RESULTATS DE LA SELECTION DES ATTRIBUTS (SLC)	101
TABLEAU 17: VISUALISATION DES PERFORMANCE DE CHAQUE CLASSIFICATEUR EN FONCTION DES ATTRIBUTS ET DES CLASSES	102
TABLEAU 18: VISUALISATION DES PERFORMANCE DE CHAQUE CLASSIFICATEUR EN FONCTION DES CLASSES	102
TABLEAU 19: MATRICES DE CONFUSION AVEC LES ATTRIBUTS DE SMART GRAIN	103
TABLEAU 20: MATRICES DE CONFUSION « FACTEURS DE FORMES CALCULES »	103
TABLEAU 21 : MATRICES DE CONFUSION « CfsSUBSETEVAL »	104
TABLEAU 22: LES MATRICES DE CONFUSION DES CLASSIFIERATTRIBUTEVAL, INFOGAINATTRIBUTEVAL ET ONERATTRIBUTEVAL	104
TABLEAU 23: LES MATRICES DE CONFUSION DE RELEFATTRIBUTEVAL	105
TABLEAU 24: RESULTATS DE LA CLASSIFICATION EN NIVEAU VARIETE	106
TABLEAU 25: RESULTATS DE LA COMPARAISON ENTRE LES DIFFERENTS ALGORITHMES DE CLASSIFICATION (VLC)	107
TABLEAU 26: RESULTATS DETAILLES PAR CLASSE DU CLASSIFICATEUR <i>FUNCTIONS.SMO</i>	108
TABLEAU 27: MATRICE DE CONFUSION DU CLASSIFICATEUR <i>FUNCTIONS.SMO</i>	108
TABLEAU 28: RESULTATS DETAILLES PAR CLASSE DU CLASSIFICATEUR <i>FUNCTION.MLPCLASSIFIER</i>	109
TABLEAU 29: MATRICE DE CONFUSION DU CLASSIFICATEUR <i>FUNCTION.MLPCLASSIFIER</i>	109
TABLEAU 30: RESULTATS DETAILLES PAR CLASSE DU CLASSIFICATEUR <i>FUNCTION.MULTILAYERPERCEPTRON</i>	109
TABLEAU 31: MATRICE DE CONFUSION DU CLASSIFICATEUR <i>FUNCTION.MULTILAYERPERCEPTRON</i>	109
TABLEAU 32 : RESULTATS DE LA SELECTION DES ATTRIBUTS (VLC)	110
TABLEAU 33 : VISUALISATION DES RESULTATS OBTENUES DE LA CLASSIFICATIONS VLC EN INTERA-CATEGORIE	111
TABLEAU 34: VISUALISATION INTRA CLASSE DE LA PRECISION DE CHAQUE CLASSIFICATEUR	111
TABLEAU 35: LES MATRICES DE CONFUSION ISSUES DE LA CLASSIFICATION PAR LES ATTRIBUTS DE SMARTGRAIN	112

TABLEAU 36: LES MATRICES DE CONFUSION ISSUES DE LA CLASSIFICATION PAR LES ATTRIBUTS DE FORMES CALCULES	112
TABLEAU 37: LES MATRICES DE CONFUSION ISSUES DE LA CLASSIFICATION PAR LES ATTRIBUTS SELECTIONNES PAR CfsSUBSETÉVAL	113
TABLEAU 38: LES MATRICES DE CONFUSION ISSUES DE LA CLASSIFICATION PAR LES ATTRIBUTS SELECTIONNES PAR RELIEFFATTRIBUTEVAL	113
TABLEAU 39: RESULTATS DE L'ETAPE DE CONSTRUCTION DU MODELE POUR LA VLC.....	126
TABLEAU 40 : RESULTATS DU TEST DES DIFFERENTS CNNs AVEC LES DONNEES DE COLLECTE DE 2017	128
TABLEAU 41 : MATRICE DE CONFUSION DU MODELE XCEPTION	129
TABLEAU 42 : MATRICE DE CONFUSION DU MODELE INCEPTIONV3	129
TABLEAU 43: MATRICE DE CONFUSION DU MODELE MOBILENET	129
TABLEAU 44: MATRICE DE CONFUSION DU MODELE DENSENET201	129
TABLEAU 45: LES RESULTATS DU TEST DE INCEPTIONV3 ET MOBILENET AVEC LES DONNEES DE COLLECTE 2017-2019	130
TABLEAU 46: MATRICE DE CONFUSION DU MODELE INCEPTIONV3 MODIFIE.....	131
TABLEAU 47: <i>CLASSIFICATION REPORT</i> DU MODELE INCEPTIONV3 MODIFIE	131
TABLEAU 48: MATRICE DE CONFUSION DU MODELE MOBILENET MODIFIE	131
TABLEAU 49: <i>CLASSIFICATION REPORT</i> DU MODELE MOBILENET MODIFIE	131

Table des matières

LISTE DES ABREVIATIONS	I
LISTE DES FIGURES	II
LISTE DES TABLEAUX	IV
TABLE DES MATIERES	VI
INTRODUCTION GENERALE	1
PARTIE BIBLIOGRAPHIQUE	7
CHAPITRE 1 : LE CARACTERE BIOLOGIQUE	8
I. INTRODUCTION	9
II. LE BLE, GENERALITES	10
III. ECHELLES DE DEVELOPPEMENT DU BLE	13
IV. CARYOPSE, LE GRAIN DU BLE	16
IV.1. ASPECTS GENERAUX	16
IV.2. LE DEVELOPPEMENT ET LA CROISSANCE DU CARYOPSE	18
IV.2.1 <i>Stades de Fondation et de Construction</i>	18
V. LE GRAIN DE BLE, UN ENSEMBLE DE TRAITS	24
V.1. COULEUR	24
V.2. TEXTURE	25
V.3. LES TRAITS MORPHOLOGIQUE ET LES DESCRIPTEURS GEOMETRIQUES	26
VI. CONCLUSION	27
CHAPITRE 2 : OPERATION DE L'IDENTIFICATION DES GRAINS	29
I. INTRODUCTION	30
II. L'IDENTITE ET L'ETAT DES GRAINS, LES ENJEUX DE LA QUALITE DANS LA CHAINE DES GRAINS	31
II.1. LA SELECTION ET LA REPRODUCTION (BREEDING)	33
II.2. LA MULTIPLICATION ET LA PRODUCTION DES GRAINS.....	34
II.3. LA RECEPTION DES GRAINS.....	35
III. TEST DE SEMENCE	36
III.1. TEST DE PURETE	37
III.2. PURETE ANALYTIQUE.....	38
III.3. PURETE GENETIQUE OU TEST DE VARIETE	39

IV.	CONCLUSION	40
	CHAPITRE 3 : IDENTITE BIODIGITALE DU GRAIN	42
I.	INTRODUCTION	43
II.	LA PERCEPTION MACHINE (MACHINE PERCEPTION)	45
III.	L'INTELLIGENCE ARTIFICIELLE.....	45
IV.	COMPUTER VISION	46
IV.1.	VISION DE BAS NIVEAU (LOW LEVEL VISION)	47
IV.1.1	Acquisition d'images :.....	48
IV.1.2	Traitement d'image	49
IV.2.	VISION DE NIVEAU INTERMEDIAIRE.....	49
IV.2.1	La segmentation d'image	49
IV.2.2	L'extraction des attributs (Feature Extraction).....	50
IV.3.	VISION DE HAUT NIVEAU (RECONNAISSANCE D'OBJET)	52
V.	MACHINE LEARNING.....	53
V.1.	APPRENTISSAGE NON SUPERVISE	53
V.2.	APPRENTISSAGE SUPERVISE	54
V.3.	RESEAUX DE NEURONES ARTIFICIELS (ANN)	54
VI.	LA REPRESENTATION MACHINE.....	55
VII.	TRAVAUX EN RELATION AVEC L'ETUDE.....	56
VII.1.	SHALLOW APPROCHES POUR LA CLASSIFICATION DES GRAINS DE BLE	58
VII.2.	THE DEEP LEARNING APPROCHES.....	61
VII.3.	IDENTIFICATION VARIETALE PAR LE DEEP LEARNING.....	63
VIII.	CONCLUSION :.....	65
	CONCLUSION DE LA PARTIE BIBLIOGRAPHIE :	66
	PARTIE EXPERIMENTALE	68
	INTRODUCTION EXPERIMENTALE	68
	CHAPITRE 1 : LA COLLECTE DES DONNEES « THE DATA COLLECTION »	70
I.	INTRODUCTION	71
II.	LA COLLECTE DES GRAINS	73
II.1.	LE MATERIEL VEGETAL	73
II.2.	DEMARCHE ET MODALITES DE LA COLLECTE DES GRAINS	74

II.3.	CONSTAT ET PROSPECTIVE.....	75
III.	LA COLLECTE DES IMAGES.....	76
III.1.	LE SYSTEME D'ACQUISITION.....	77
III.2.	DEMARCHE DE L'ACQUISITION DES IMAGES.....	78
III.2.1	<i>Distribution et arrangement des grains</i>	79
III.2.2	<i>Choix de l'arrière-plan</i>	82
III.2.3	<i>Définition de l'échelle :</i>	82
III.2.4	<i>Paramétrage du scan et annotation des images :</i>	83
III.3.	OBTENTION DES IMAGES DE GRAINS INDIVIDUELS.....	83
IV.	CONCLUSION	84
	CHAPITRE 2 : EXPERIMENTATION N° 1 SHALLOW APROCHE.....	85
	MATERIELS ET METHODES	86
I.	INTRODUCTION	87
II.	LA BASE DE DONNEES DE L'APPRENTISSAGE :.....	89
III.	PROTOCOLE EXPERIMENTALE SHALLOW	89
IV.	L'EXTRACTION DES ATTRIBUTS MORPHOLOGIQUES.....	91
V.	CLASSIFICATION D'IMAGES ET ANALYSE DE DONNEES.....	92
V.1.	LES CLASSIFICATEURS (CLASSIFIER)	92
V.2.	LA SELECTION DES ATTRIBUTS.....	94
V.3.	L'EVALUATION DE LA CLASSIFICATION	95
	RESULTATS ET DISCUSSIONS	96
VI.	LA CLASSIFICATION EN NIVEAU « ESPECE »	96
VI.1.	MANIPULATION A -SLC, LA CLASSIFICATION SANS SELECTION D'ATTRIBUTS.....	96
VI.1.1	<i>Pré-sélection des classificateurs en SLC</i>	96
VI.1.2	<i>Optimisation d'hyperparamètres des classificateurs (MnpA)</i>	97
VI.1.3	<i>Résultats Détaillés de la Manipulation A :</i>	99
VI.2.	MANIPULATION B-SLC, LA CLASSIFICATION PAR SELECTION D'ATTRIBUTS	100
VI.2.1	<i>Résultats de la sélection des attributs</i>	101
VI.2.2	<i>Performances des classificateurs (SLC)</i>	102
VI.2.3	<i>Matrices de confusion de chaque classification</i>	103
VII.	LA CLASSIFICATION EN NIVEAU « VARIETE »	106
VII.1.	MANIPULATION A- VLC, SANS SELECTION D'ATTRIBUTS	106

VII.1.1	<i>Performances et Pré -Choix des classificateurs :</i>	106
VII.1.2	<i>Optimisation d'hyperparamètres des classificateurs (VLC)</i>	107
VII.1.3	<i>Résultats détaillés de la classification</i>	108
VII.2.	MANIPULATION B- VLC	110
VII.2.1	<i>Résultats de la Sélection des attributs (VLC)</i>	110
VII.2.2	<i>Performances des classifications en (VLC)</i>	110
VII.2.3	<i>Matrices de confusion issues des différentes combinaisons d'attributs</i>	112
VIII.	DISCUSSION	114
IX.	CONCLUSION DE L'EXPERIMENTATION N°1	116
	CHAPITRE 3: EXPÉRIMENTATION N°2 DEEP LEARNING APPROCHES	117
I.	INTRODUCTION	118
II.	LA BASE DE DONNEES IMAGES	119
III.	CLASSIFICATEUR DEEP LEARNING	120
IV.	LA TECHNIQUE DE « TRANSFERT LEARNING »	123
V.	STRUCTURE DES MODELES	124
V.1.	ARCHITECTURE	124
V.2.	HYPERPARAMATERS	125
VI.	RESULTATS ET DISCUSSION	126
VI.1.	PHASE DE BUILD MODEL	126
VI.2.	TEST DU MODELE	128
VI.3.	PERFORMANCE DES MODELES CNNs SUR LA BASE DE DONNEES 2017-2019	130
VI.4.	PHENO-DEEP DISCUSSION	131
VII.	CONCLUSION DE L'EXPERIMENTATION N°2	133
VIII.	CONCLUSION GENERALE	134
	RÉFÉRENCES BIBLIOGRAPHIQUES	136

Introduction générale

La culture du blé, comme les autres cultures céréalières, fait partie des grandes cultures à finalité stratégique ; cette dernière est principalement liée à la nutrition humaine et à la sécurité alimentaire. En raison de ses caractéristiques uniques, son adaptabilité agronomique, sa facilité de stockage et ses qualités nutritionnelles (Wrigley, 2009) la culture du blé a contribué à la sédentarité de l'humanité et a donné naissance aux premières civilisations (Gustafson *et al.*, 2009). Le résultat de l'évolution des activités relatives de la culture de blé ainsi que les autres céréales, a été l'émergence de la chaîne de valeurs « la chaîne de grains » (Wrigley, Batey and Miskelly, 2017). Aussi, un écosystème de production d'aspect agronomique, industriel et technologique s'est installé autour de ces grains blé pour leur cultivation et transformation (Rust, Davis and Goldberg, 1957; Bonneuil and Thomas, 2012) ; Des règles de management de qualité en vue de la réduction des pertes ont été établies (Delwiche, 2010).

Selon le rapport de l'ONU sur les *Perspectives de la population dans le monde de 2019* la population mondiale passera de 7,7 milliards à 9.7 milliards en 2050. Chaque année une quantité colossale de blé est produite à l'échelle mondiale, pour assurer l'approvisionnement en blé, elle est en 2021 à 775 millions de tonnes. Le maintien de ce régime de production en cadence évolutive passe obligatoirement par la maîtrise des différents aspects de la chaîne de grains à savoir la production, la technologie et la gestion de la qualité, afin d'atteindre un rendement optimal et un produit de qualité.

Un des principaux soucis dans cette chaîne de valeur est de préserver l'identité et la qualité des grains. Ce souci principal est lié aux pertes qualitatives et quantitatives dues à la traçabilité ; la maîtrise de ces concepts est importante pour la réduction des pertes et gaspillages de denrées alimentaires, ce qui est l'une des cibles essentielles des *objectifs de développement durable* (ODD) ; cette maîtrise constitue également un moyen d'atteindre d'autres cibles des ODD, en particulier celles afférentes à la sécurité alimentaire, à la nutrition et à la durabilité environnementale (FAO, 2019).

La notion *identité et état du grain* est étroitement liée à la notion de la qualité et de l'évaluation de cette dernière dans la chaîne de grain (Shewry, 1996; Wrigley, 2017a). Une évaluation efficace de la qualité du grain fournit des connaissances essentielles sur les défauts et les avantages des grains analysés. Ça peut concerner quelques grains cultivés par un sélectionneur comme base pour

un nouveau génotype ou bien de grosses quantités de marchandise en vrac. Ce qui arrive au grain après l'analyse dépend en grande partie des résultats analytiques. Ainsi, un bon management depuis la production jusqu'à la transformation dépend aussi de l'efficacité de l'analyse (Wrigley, 2017a).

Aujourd'hui et à l'échelle mondiale, la chaîne d'événements à valeur ajoutée (qui détermine les aspects essentiels de la qualité du grain) est devenue normative; elle commence par les croisements et la sélection de nouvelle variété pour avoir des caractères de qualité meilleure; ensuite c'est la propagation des lignées élites lors d'essais sur le terrain sur divers sites, dont l'objectif est de sélectionner les meilleures lignées et les inscrire aux droits d'obteneur (UPOV, 2010; Korir *et al.*, 2013; ISTA, 2018a), l'évènement suivant est de multiplier les semences pour une diffusion commerciale (programme de multiplication); c'est ainsi que les grandes opérations de semis et de gestion à la ferme commencent, suivies par les opérations de poste-récolte, le transport et le stockage des grains; enfin vient l'étape de la transformation et l'utilisation de ces grains et leurs dérivés pour l'alimentation humaine et animale.

Dans l'industrie des semences, le programme de certification des semences est un processus important en raison de son impact économique certain (Copeland and McDonald, 2001). La certification est une méthode et des règles officiellement reconnues (L'OCDE, 2009; AOSA, 2018; ISTA, 2018d). La certification fournit les informations de qualité nécessaires à l'étiquetage des semences à vendre. Ce programme comprend un processus de test des semences (Elias *et al.*, 2012). Lorsqu'un lot de semences est considéré pour la certification, l'échantillon soumis est testé pour la pureté, la viabilité, la germination et la teneur en graines de mauvaises herbes nocives. Les tests importants concernent la viabilité et la pureté des semences. Le test de pureté comprend deux aspects, la pureté physique et la pureté variétale (Copeland and McDonald, 2001). Selon Elias *et al.* (2012) plusieurs facteurs peuvent affecter la pureté des graines, parmi lesquels: le mélange mécanique et l'étiquetage incorrect, et d'autres. Ils modifient l'intégrité génétique du cultivar de blé d'origine, ce qui rend l'identification variétale impérative pour déterminer le niveau de pureté d'un lot de semences. L'analyse est effectuée par un organisme de certification privé ou par un laboratoire national de semences officiel (Copeland & McDonald, 2001). L'analyse physique est également un test courant à la réception de la récolte pour la consommation (Howitt and Miskelly, 2017).

Durant la dernière décennie (2009- 2019), la quantité moyenne de semence brute réceptionnée à chaque campagne est d'environ 2,5 millions de tonnes; des milliers d'échantillons prélevés des

lots de semence (en fonction du grade de la semence) sont soumis à l'analyse au laboratoire pour le contrôle de qualité ; en moyenne deux millions de quintaux sont contrôlés et une moyenne de 450 mille quintaux sont refusés chaque année.

La principale cause des refus de certification de semence est liée à la pureté ; en moyenne 80% des refus (période 2009-2019) sont dues au mélange spécifique (la présence de d'autres espèces). Les quantités refusées pour un mélange variétal ne sont définies qu'en test à postériori au champ en contrôle en végétation.

Le test de pureté physique (Figure 1) est une méthode non destructive et une analyse visuelle directe des caractéristiques des grains (ISTA, 2018b). L'opération est basée totalement sur la perception humaine et sur sa propre expérience, le pivot central étant la **reconnaissance de l'identité des grains (Le GIR (Grain Identity Recognition))** qui est basée sur **l'analyse visuelle des grains (VGA)**.

En règle générale, l'expert effectue une **classification au niveau des espèces (SLC)** en discriminant manuellement le contenu de l'échantillon en fractions comprenant les semences pures, les semences d'autres cultures, les semences de mauvaises herbes et la matière inerte.

En ce qui concerne les tests de pureté variétale, l'expert effectue des examens basés sur l'observation des aspects visuels des grains ; cette tâche correspond à **la classification au niveau variétal (VLC)**. Généralement dans ce test (VLC), l'identification est basée sur des clés de détermination impliquant la plante entière, durant son cycle biologique, la croissance morphologique, son comportement et d'autres clés pour différencier les variétés de blé.



Figure 1: Le test de pureté au laboratoire

Il convient donc de noter qu'une même tâche est effectuée dans les deux tests de pureté (physique et variétale), il s'agit de la reconnaissance d'identité de grain (GIR) par une approche de classification à plusieurs niveaux taxonomiques (VLC et SLC), en exploitant toutes les clés visuelles du grain qui fournissent la plus haute distinction entre les variétés même si elles peuvent différer en raison des changements dans les conditions de croissance (Howitt et Diane, 2017) ; par conséquent, cette identification est la tâche clé pour SLC et VLC. À l'exception des méthodes basées sur l'ADN, sur des réactions chimiques ou d'autres méthodes destructives, le GIR basé sur l'analyse visuelle des grains (VGA) est d'un grand avantage économique et d'une grande utilité opérationnelle sur le terrain. Outre le fait que l'analyse de la pureté est une méthode à faible débit, la VLC et dans certains cas la SLC basé sur la VGA peuvent être difficiles voire impossibles à réaliser en raison de la grande similitude entre les caractéristiques des espèces *Triticum spp* (Chiara Delogu, 2013); ces caractéristiques peuvent également être affectées par les conditions de croissance.

Dans ce contexte, pour entreprendre le problème ci-dessus, une nouvelle approche est suggérée pour effectuer le GIR, par un processus de classification automatique en intégrant une solution technologique avancée. L'approche devrait avoir les critères suivants : elle est basée sur la perception artificielle qui peut remplacer la perception humaine, elle préserve les avantages du VGA en utilisant le grain entier et elle est capable à effectuer une VLC avec précision. Les approches automatiques pourraient améliorer considérablement les conditions de travail des tests de pureté, ainsi que d'avoir d'autres avantages liés à sa mise en œuvre (Génération des données).

La méthode qu'on a visée, est basée sur la technologie de Computer Vision (CV) et de Machine Learning (ML) (E.R. Davies, 2012; Makridakis, Spiliotis and Assimakopoulos, 2018). Le CV et le ML ont déjà été appliqués avec succès dans de nombreux secteurs et chaînes de production pour les mêmes tâches de reconnaissance et de classification d'objets, dans l'industrie agro-alimentaire (Dutta Gupta and Ibaraki, 2014; Vithu and Moses, 2016), dans des recherches avancées sur céréales et autres cultures (Davies, 2009; Pridmore, 2015; Patrício and Rieder, 2018; Tripodi *et al.*, 2018). Récemment, (Wäldchen and Mäder, 2018) ont réalisé une revue systématique de la littérature, dans le but d'une analyse approfondie et d'une comparaison des études primaires sur les approches de Computer Vision pour l'identification des espèces végétales.

La classification d'images est l'une des techniques de Machine Learning (ML) les plus utilisées pour les tâches d'identification et de reconnaissance (Du and Sun, 2006; Patrício and Rieder, 2018).

Selon la définition de Davies (2012) et la réflexion conceptuelle sur les systèmes de Machine Learning ML de Le Cun (LeCun, Bengio and Hinton, 2015) deux approches sont principalement utilisées pour la classification d'images à savoir les modèles classiques (Shallow models) et les modèles Deep Learning.

Pour être clair, on a défini et analysé le problème central de notre étude comme étant un problème de *reconnaissance de l'identité du grain (GIR) au niveau variétal (VLC)* par la technique de classification d'images impliquant la CV et ML.

En ce qui concerne notre travail, ce dernier est divisé en deux parties et chaque partie contient des chapitres comme suit :

Partie I qui est la partie bibliographie elle comprends 3 chapitres

Chapitre 1 : Le caractère biologique

Dans ce chapitre nous avons abordé l'aspect biologique du grain et son origine en tant que matière biologique issue de *l'interaction génomes-environnement*. La question de *l'échelle de développement* a été abordée pour mettre en contexte le concept « l'origine des différentes caractéristiques du grain », qui sont le résultat du processus de formation, de différenciation et de développement durant le cycle biologique. L'aspect apparent (visuel) du grain a été traité avec importance car, la CV est basée sur les traits qualitatif et quantitatif des grains. Le background théorique construit dans ce chapitre renferme une partie importante des informations relatives aux grains, qui permettent l'analyse et l'interprétation des résultats.

Chapitre 2 : l'opération de l'identification des grains (La perception humaine)

Ce chapitre traite l'opération faisant l'objet de l'intégration de la CV et le ML, cette intégration a été revue dans ce chapitre, l'objectif étant de comprendre les concepts de base liés à l'opération de l'identification et les domaines d'exploitation de cette opération, un focal sur la technologie de la semence et l'analyse de la pureté.

Chapitre 3 : identité biodigitale des grains de blé (La perception machine)

Ce chapitre est une brève adaptation des concepts théoriques de la solution technologique proposée à la problématique ; il comprend une introduction à l'intelligence artificielle et ces deux branches exploitées dans notre travail, s'agissant de la Computer Vision et le Machine Learning (perception machine), ainsi qu'une revue détaillée des travaux connexes à l'opération d'identification et de classification des grains.

Partie II expérimentation en 3 chapitres :

Chapitre 1 : La collecte des données

Ce chapitre est consacré à l'opération « la collecte des données » et à la construction de la base d'apprentissage selon une démarche *et une stratégie élaborée* (basée sur la variabilité et la diversification des données).

Ce chapitre décrit en finalité l'approche de collecte des données à suivre dans un système d'information sur la qualité des grains.

Chapitre 2 : l'expérimentation N°1 l'approche Shallow Learning :

Dans ce chapitre nous avons décrit notre première tentative à réaliser la classification des grains par la CV classique, qu'on a appelé l'Approche Shallow (approche superficielle). Cette dernière est basée sur l'extraction préalable des caractéristiques (morphologiques dans notre cas) et la classification par des algorithmes d'apprentissage Machine. Les résultats et les conclusions sur la faisabilité et les limites de cette technique nous ont conduit à exploiter une autre méthode basée sur le Deep Learning.

Chapitre 3 : l'expérimentation N° 2 l'approche Deep Learning :

Dans ce chapitre nous avons décrit notre deuxième tentative à réaliser la classification des grains par la classification d'images par des réseaux de neurones CNNs (*Convolutional Neural Network*). Cette technique nous a donné des résultats confirmant la faisabilité de l'identification des grains à l'échelle variétale.

Quelques contributions de notre travail de recherche :

Ce travail constitue une approche et des modalités de l'utilisation de la vision artificielle pour la classification des grains au niveau variétale. Ce travail va ouvrir des horizons pour l'utilisation de cette solution à tous les niveaux de la chaîne de grains (contrôle, inspection, certification, phénotypage...).

Partie Bibliographique

Chapitre 1 : Le caractère biologique

I. Introduction

La cultivation, la propagation et l'amélioration des plantes à graines ont été des phases importantes du développement de l'agriculture et par conséquent celui de l'humanité suite à la sédentarisation (Serna-Saldivar, 2010). Les grains de blé, comme les autres céréales, sont une importante source de nourriture ; ces grains partagent une caractéristique essentielle qui est à l'origine de leurs importances dans le régime alimentaire de l'humanité ; cette caractéristique est leurs capacités à rester intacts, pendant une longue période, en tant que réserve de nutriment (Gustafson *et al.*, 2009). Grâce à cette capacité on a pu les stocker et les transporter à longues distance, parfois dans des conditions très sévères. Cette période s'étale du moment où la plante est à la fin du stade de maturité de ces grains jusqu'à la germination ou jusqu'à la consommation.

Au stade actuel de l'évolution, toute en restant fidèle à la « caractéristique essentielle » cité plus haut, les caractéristiques recherchées dans la chaîne des grains, se focalisent principalement sur l'aspect « quantité et qualité ». Les grains de blé renferment des informations de nature et origines diverses qui peuvent définir cet aspect de « quantité et qualité », ces informations dépendent du **caractère biologique** du grain, des pratiques agronomiques, de la solution technologique déployée à chaque niveau de la chaîne des grains et l'environnement dans lequel le tout interagit.

A travers ce chapitre nous avons abordé l'aspect biologique du grain de blé et vu son importance dans la construction du background agronomique. Ceci est nécessaire à décortiquer l'intégration de la technique de la Vision Artificielle dans l'identification de l'identité des grains de blé pour des applications diverses, comme le contrôle et la certification de semences.

Sous le **concept du caractère biologique**, nous avons considéré que l'identité du grain est liée à son origine en tant que matière biologique issue du processus de *formation et de développement* durant le cycle biologique. La question de l'échelle de développement a été abordée pour mettre en contexte le concept et localiser « l'origine des différents traits du grain ». L'aspect apparent du grain a été traité avec importance car, la Computer Vision est basée sur des traits d'identification qualitatifs et quantitatifs des grains. Ce background théorique renferme donc une partie importante des informations utilisées dans l'analyse et l'interprétation des résultats.

II. Le Blé, généralités

Au cours des dix derniers millénaires, la domestication des céréales (Tanno and Willcox, 2006) notamment le blé (*Triticum* spp.), le riz (*Oryza sativa* L.) et le maïs (*Zea mays* L.) fut l'une des démonstrations les plus spectaculaires impliquant la manipulation par l'homme des processus évolutifs de spéciation, de sélection naturelle et d'adaptation (Gustafson *et al.*, 2009).

Selon (Pitt, 2014) le « blé » est une construction ethnobotanique (Portères, 1966), un collectif de plantes défini par l'homme englobant une diversité taxonomique, génétique et morphologique. Elle appartient à la famille des graminées (Poaceae, tribu Triticeae), mais sa taxonomie est extrêmement compliquée. Néanmoins, le terme blé décrit un certain nombre d'espèces et de sous-espèces du genre *Triticum* (Michael J Gooding, 2009), les plus importantes aujourd'hui sont le blé tendre (*T. aestivum* subsp. *aestivum*), qui représente plus de 90 % de la production mondiale de blé, et le blé dur (*T. turgidum* subsp. *durum*), qui est responsable de 5 % supplémentaires. Cependant, selon (Pitt, 2014) le blé englobe une suite botanique beaucoup plus vaste, comprenant deux genres, *Triticum* et *Aegilops* et environ 600 espèces (Zohary et Hopf 2000, Morrison 2001 in Pitt, 2014). L'amélioration du blé est étroitement associée à la caractérisation et à la compréhension de l'évolution du blé et à la diversification génétique de diverses espèces et les apparentées au blé. L'évolution du genre *Triticum* résulte principalement de l'hybridation inter- et intra-spécifique, de la polyploïdisation et de la formation récurrente du blé et de ses parents sauvages (Gustafson *et al.*, 2009).

Des preuves botaniques, génétiques et archéologiques récentes ont montré qu'une petite zone centrale du croissant fertile - près des cours supérieurs du Tigre et de l'Euphrate, dans l'actuel sud-est de la Turquie - nord de la Syrie - fut le berceau de la culture céréalière (Bonjean, 2001). Mais bien avant, les analyses des grains de pollen (Figure 2) montrent que les graminées étaient présentes il y a 55 millions d'années (MA), au Paléocène, bien que leur première apparition puisse bien avoir été antérieure à cela (Kellogg 2001 in Gooding, 2009).

Les séquences génétiques (Huang *et al* 2002, Gill *et al* 2004 in Gooding, 2009) indiquent que les trois principales céréales (blé, riz et maïs) ont divergé d'un ancêtre commun d'environ 40 MA. On estime que la divergence entre le blé et l'orge et le blé et le seigle s'est produite à 10–14 MA et 7 MA, respectivement. La divergence la plus récente est celle des graminées de chèvre (*Aegilops* spp.) de *Triticum*, qui semblent avoir rayonné de 2,5– 4,5 MA. (Michael J. Gooding, 2009)

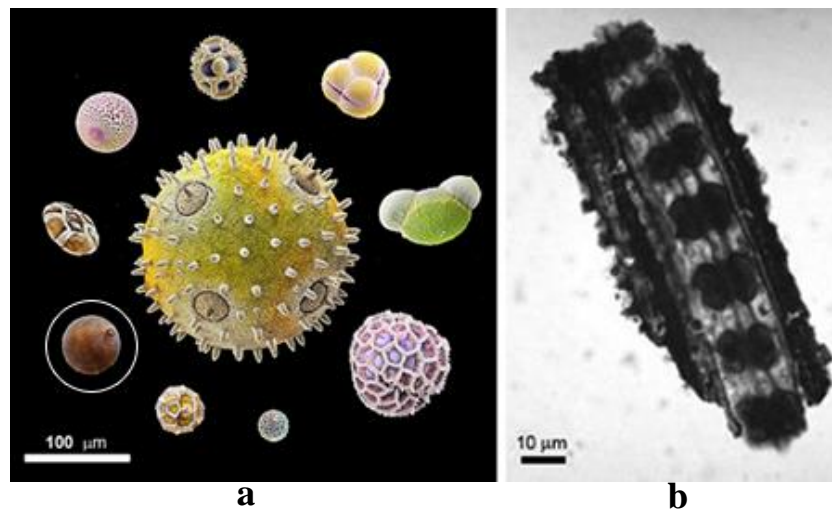


Figure 2: (a) Images au microscope électronique colorées artificiellement de divers grains de pollen pour montrer leur caractère distinctif et leur variété de taille et de forme - le pollen de graminées est encerclé. (b) Cellules épidermiques fossilisées trouvées (WheatBP, 2021)¹

D'après (Vergauwen and De Smet, 2017) l'ancêtre tétraploïde du blé dur d'aujourd'hui est né il y a environ 500 000 ans, tandis que l'ancêtre hexaploïde du blé tendre d'aujourd'hui est né d'un récent événement de polyploïdisation il y a environ 8 000 ans au Moyen-Orient . Le même auteur affirme que trois espèces de graminées ancestrales sont à l'origine du Blé hexaploïde : le *Triticum urartu* (fournissant le génome **A**), un proche de l'actuel *Aegilops speltoides* (fournissant le génome **B**) et le *Aegilops tauschii* (fournissant le génome **D**) (Figure 3A).

Comme illustré dans la figure 3A, la combinaison de *T. urartu* et d'un proche parent d'*A. speltoides* a donné naissance à *T. turgidum ssp. dicoccoides* (emmer sauvage), qui était cultivé sous le nom de *Triticum turgidum subsp. con.* Cette dernière espèce est l'ancêtre de *T. turgidum ssp. durum* et, en combinaison avec *A. tauschii*, a donné naissance au *T. aestivum* hexaploïde (Figure 3 A,B). Un parent de *T. urartu*, à savoir *Triticum monococcum ssp. aegilopoides* (également appelé *T. boeoticum*), était cultivé comme petit épeautre (appelé *T. monococcum ssp. monococcum*).

¹ http://bio-gromit.bio.bris.ac.uk/cerealgenomics/WheatBP/Documents/DOC_Evolution.php

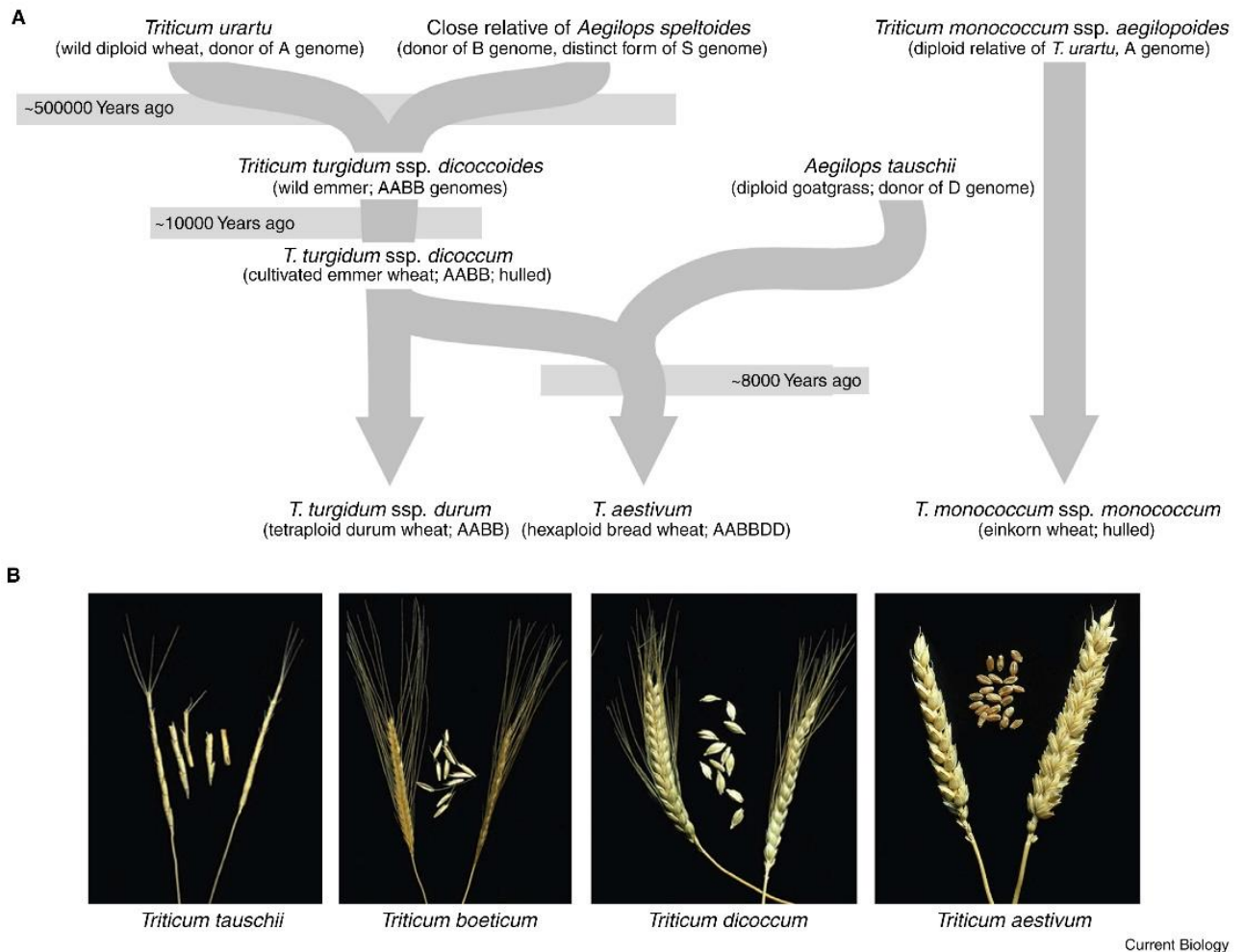
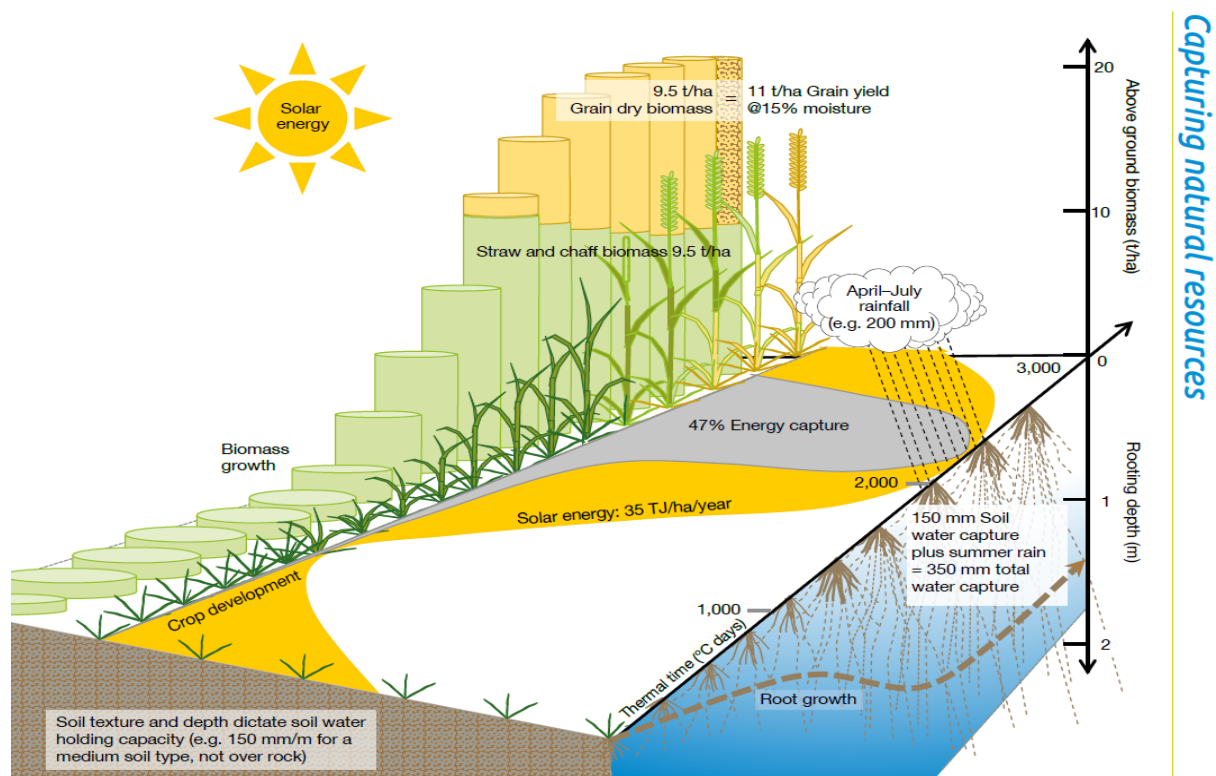


Figure 3: Evolution du Blé. (A) Voie simplifiée menant au blé cultivé. (B) Épis et graines d'espèces de blé sélectionnées. (Vergauwen et De Smet, 2017)

Selon Bonjean, (2001), jusqu'à la fin du 19^e siècle, tous les blés cultivés existaient sous forme de variétés locales très hétérogènes. Les cultivars de blé étaient des mélanges morphologiquement uniformes de lignées consanguines et de ségrégations hybrides, produits de faibles niveaux de croisements aléatoires au sein d'une race locale. Toute sélection artificielle visait principalement un rendement accru, une plus grande taille de graine, une meilleure qualité de farine et une adaptation à un plus large éventail de régimes climatiques et agricoles (Feldman et al., 1995).

III. Echelles de développement du blé

Dans un cycle biologique, le développement et la croissance d'une plante sont des processus liés, mais distincts (Figure 4). La croissance est définie comme étant l'augmentation permanente de volume qui est due plutôt à l'accumulation de la biomasse, tandis que le développement est la progression du cycle de vie de la plante, en matière de l'initiation et la différenciation des organes ; ainsi il s'agit de la progression des étapes par lesquelles passent les cellules, les organes et les plantes au cours de leur cycle de vie. Souvent, mais pas toujours, la croissance et le développement se produisent simultanément. (McMaster, 2009; Hyles *et al.*, 2020)



Le développement comprend, selon Hyles *et al.* (2020) des phases distinctes tel qu'illustré dans la Figure 4: Cycle biologique du blé, entre le développement et la croissance la plante évolue dans l'espace et dans le temps. (AHDB, 2021)

figure 5. Selon Meier *et al.*, (2009) et Hyles *et al.*, (2020) plusieurs échelles de développement ont été proposées pour décrire les phases clés du cycle biologique du blé. Les différentes échelles montrent les phases de développement de la plante du blé en optimum culture. Selon (Kong, Guo and Sun, 2015) ces différentes phases clés ne sont en effet que l'exécution jalonnée, sous l'influence de différents facteurs, du code génétique stocké dans les supports de l'information génétique, cette dernière est partagée par trois génomes nécessaire à la vie de la cellule végétale,

il s'agit de l'ADN, le génome nucléaire 95% de la totalité, le génome des mitochondries (environ 1%) et le génome des chloroplastes (environ 4%). Au sein des cellules végétales, il y a, en effet, une coopération entre ces trois génomes afin de déterminer les différents caractères de la plante.

Classiquement la croissance et le développement sont divisés en plusieurs stades à savoir: **la germination** conduit aux plantules, le premier stade de croissance de la plante, suivi du **tallage** (formation de pousses latérales), de **l'hivernation** (la chute des températures en automne incite le plant de blé à développer un niveau élevé de résistance au froid, subissent également un changement d'une forme de croissance verticale à une forme de croissance prostrée), **jointure** (les nœuds de la tige sont d'abord détectés au-dessus du sol), **la montaison** (le gonflement de la gaine de la feuille ultime, la feuille étendard, lorsque l'épi en développement se développe à l'intérieur), **l'épiaison** (à mesure que la tige continue de s'allonger, la tête est poussée hors de la gaine de la feuille étendard) et **la floraison**, **La maturation**, ou développement du grain, se divise en lait, pâte molle, pâte dure et maturité physiologique, stade où le poids du grain est maximum (Paulsen, 1997 in (Edwards, 2010). En matière d'échelles, Feekes a proposé en 1941 une échelle décimale (stades 1 à 11), représentant le cycle de vie du blé depuis le tallage, l'allongement de la tige, l'épiaison et la floraison, jusqu'à la maturation (Figure. 5B.).

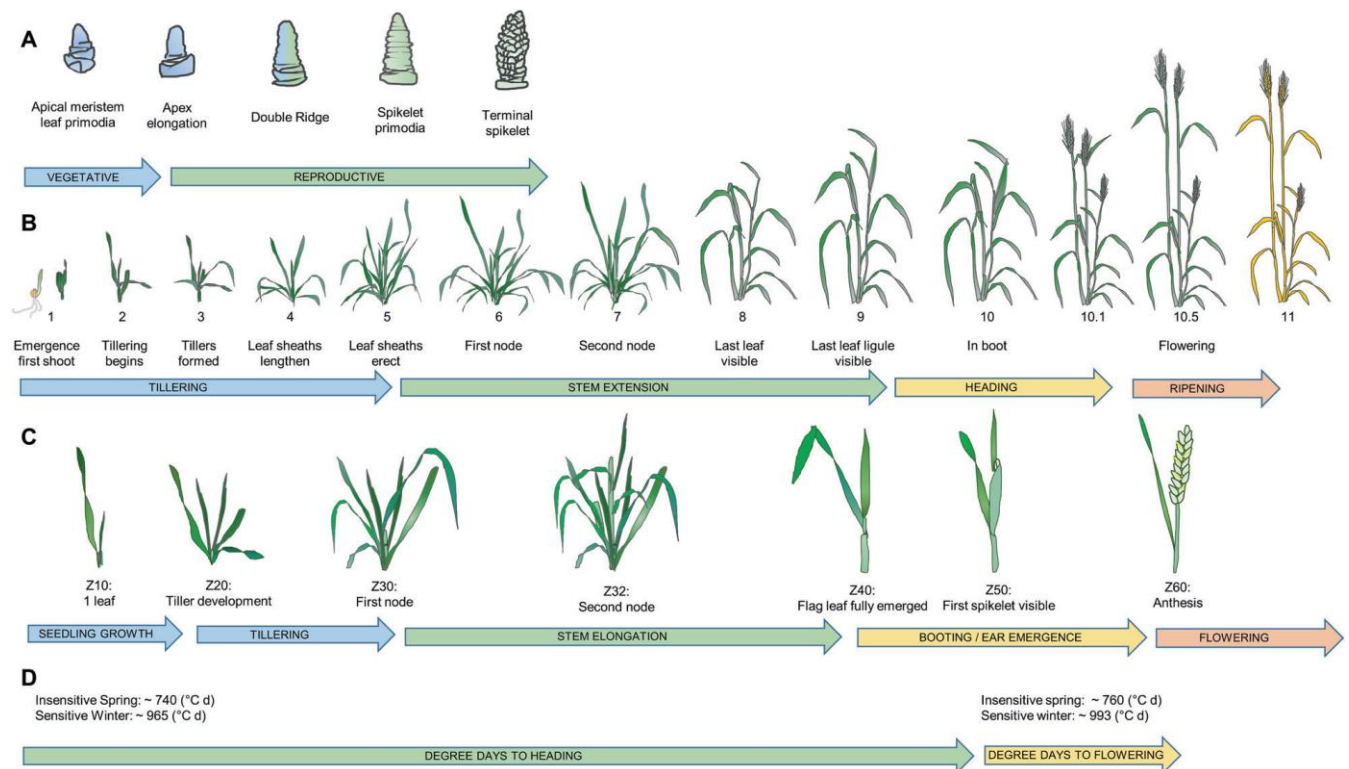


Figure 5: Différentes échelles de développements de la plante de blé (Hyles et al., 2020)

Une autre échelle (Figure 5A.) développée par (Haun, 1973) , par cette échelle on quantifie l'émergence progressive des feuilles sur la tige principale du blé, qui peut ensuite être utilisée pour déterminer le taux d'émergence des feuilles, également connu sous le nom de phyllochron. En outre, (Zadoks et al. 1974) ont proposé une échelle complète, en décrivant le cycle de vie du blé de la germination à la maturation dans un format décimal à deux chiffres « L'échelle de Zadoks » comprend 100 stades décrivant le développement du plant de blé (Figure. 4C).

Une autre échelle développée pour plusieurs espèces entre 1979 et 1988 appelé **BBCH** (*Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie*) dérivée de celle de zadoks (Meier *et al.*, 2009; Meier, 2018); cette échelle, dans le cas les céréales, a été proposée par (LANCASHIRE *et al.*, 1991) ; il s'agit aussi d'une échelle décimale qui permet de codifier les stades phénologiques des plantes cultivées ; les **stades principaux** sont décrits sur une échelle qui va de 0 à 9, celle-ci est complétée par les **stades secondaires** qui s'échelonnent de 0 à 9 à l'intérieur d'un stade principal. On obtient ainsi un code à deux chiffres composés par le stade principal et le stade secondaire. Compte tenu de la diversité des espèces, certains stades peuvent être inversés voire absents

La figure 6 montre une autre vision d'une échelle de développement et de croissance (AHDB, 2021)² basée sur l'échelle (BBCH) en divisant la dynamique du cycle biologique en 3 actions : **Fondation** (GS00-GS30), **Construction** (GS31-GS60) et **Production** (GS61-GS92). Cette échelle sera le repère qu'on va adopter dans la description du développement et croissance du grain ci-après.

La synchronisation de chaque phase de développement avec des conditions saisonnières optimales est nécessaire pour optimiser la production de biomasse et le rendement. Par exemple, le nombre de grains et donc le rendement du blé sont largement déterminés par les taux de croissance pendant la période critique qui s'étend de l'émergence de l'avant-dernière feuille jusqu'au début du remplissage des grains (Dreccer *et al.*, 2018). Sur le plan agronomique, il est donc primordial d'aligner cette phase sensible en l'occurrence avec de probables conditions saisonnières les plus propices à la croissance du blé (température, rayonnement et disponibilité en eau).

Le calendrier des phases de développement influence également la tolérance au stress abiotique comme la résistance à l'hiver.

² (AHDB): Agriculture and Horticulture Development Board (ahdb.org.uk)

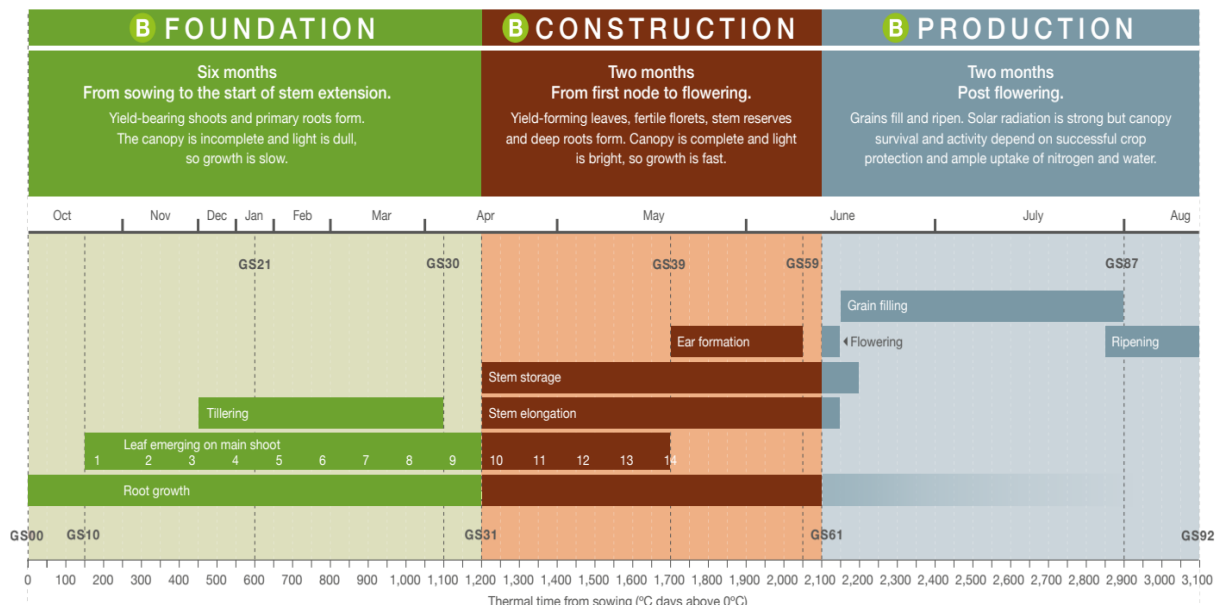


Figure 6: Les stades de développement selon (ADBH, 2021)

Les conditions saisonnières et les facteurs régionaux, y compris l'humidité disponible, la température, la latitude et la durée du jour, influencent tous la durée des phases de développement (Hyles *et al.*, 2020). Cette dépendance du développement phasique des cultures vis-à-vis de l'environnement de croissance, représente une forte interaction génotype-environnement et agit pour synchroniser le cycle de vie avec les conditions externes.

IV. Caryopse, le grain du blé

IV.1. Aspects généraux

Le caryopse des graminées, en termes botaniques, désigne un fruit à graine unique (Bewley *et al.*, 2013) qui comprend l'embryon et l'endosperme³ (Vallade, 2004) entourés des téguments et du péricarpe (la couche interne de l'ovaire). L'embryon se compose du scutellum et de l'axe embryonnaire (Figure 7). Lors de la fécondation, le pollen, gamète mâle, apporte l'information contenue dans le noyau, alors que l'ovule, gamète femelle, fournit, en plus de son noyau, son cytoplasme ; on attribue donc aux ADN mitochondriaux et chloroplastiques la responsabilité de l'hérédité de type cytoplasmique (gnis-pédagogie, 2018).

Le caryopse est donc un dérivé des produits du gamétophyte femelle fécondé et des tissus maternels environnants. Le gamétophyte se compose du nucelle et du sac embryonnaire et est entouré de deux couches protectrices distinctes de tissu maternel, les téguments interne et externe.

³ : on retient la terminologie anglaise « endosperme » pour décrire cette partie (l'albumen amylosé) du grain selon (Vallade, 2004).

En dehors des téguments se trouve la paroi du carpelle, qui deviendra plus tard le péricarpe. (Drea *et al.*, 2005).

Les caractéristiques morphologiques principales (Figure 7 et 8) sont la forme plus ou moins ovale du grain et le sillon longitudinal marqué, ou pli, sur sa face ventrale s'étend presque jusqu'au centre du grain, L'embryon est situé sur la face dorsale du grain (Surget and Barron, 2005). Une cicatrice subsiste dans la zone d'attache à l'extrémité basale du grain, tandis qu'à l'extrémité apicale une touffe de poils constitue la brosse. La forme du grain varie selon le type et le cultivar.

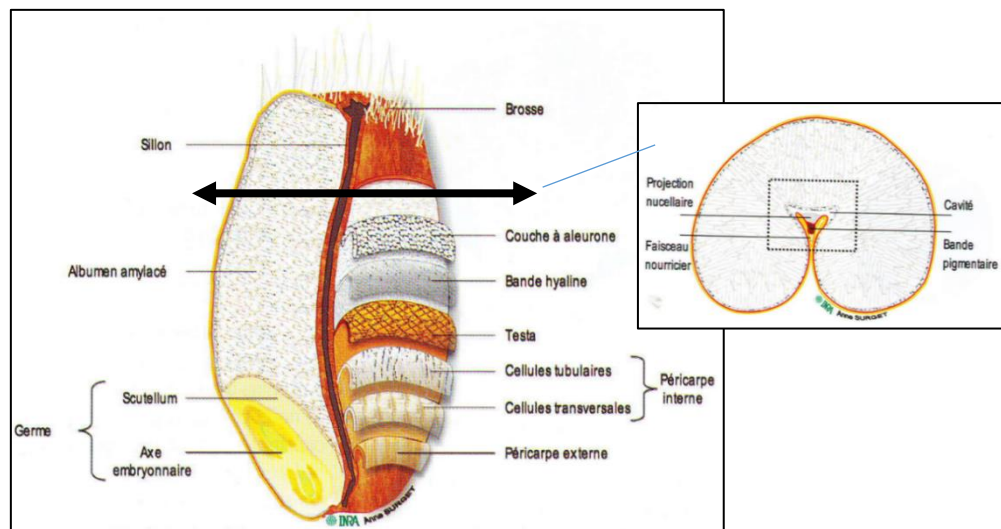


Figure 7 : structure d'un grain de blé mur (a : coupe longitudinale, b : coupe transversale) (Surget and Barron, 2005)

Le diagramme suivant illustre la constitution du grain de blé

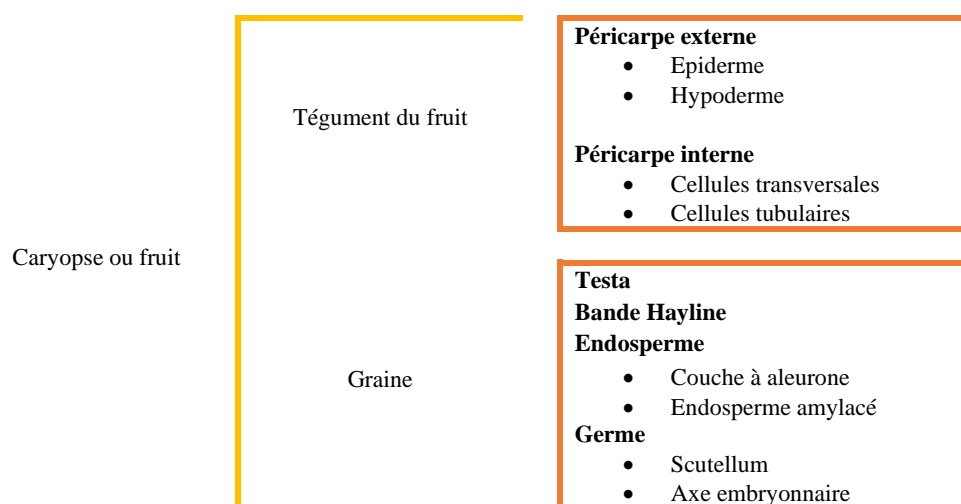


Figure 8: Diagramme de constitution du grain de Blé (Shewry *et al.*, 2009)

IV.2. Le développement et la croissance du caryopse

IV.2.1 Stades de Fondation et de Construction

Selon Kong, Guo and Sun (2015), le développement des caryopses est un processus génétiquement programmé. Il est lié à la sénescence des organes végétatifs et à la remobilisation des produits de dégradation vers les organes reproducteurs (puits). L'initiation et la progression de la sénescence des organes végétatifs sont sous le contrôle d'un ensemble de signaux environnementaux (tels que les stress biotiques et abiotiques, l'obscurité et la disponibilité des nutriments) et de facteurs endogènes (y compris le vieillissement, plusieurs hormones et la disponibilité du sucre).

Le développement du caryopse fait partie intégrante du développement **l'inflorescence** du blé, dont l'unité de base qui le compose est l'épillet (Li *et al.*, 2019). elle s'étale sur l'échelle de **Feekes (FGS 4-FGS 11), Zadoks (Z30-Z92) et l'Echelle BBCH (GS30-GS92)** ; dans certain cas comme l'évènement pivot dans le cycle de vie du blé qui est la transition de l'apex de la pousse du développement végétatif au développement reproducteur une échelle de développements quantitative est suivie comme celle de (Waddington, Cartwright And Wall, 1983), cette dernière commence par « la transition de l'apex » et se poursuivant par la morphogenèse de la fleur et du pistil (gynécée) jusqu'à la pollinisation.

Dans la jeune plante, l'apex de la pousse s'allonge, suivi de la formation de deux crêtes sur les côtés de l'apex de la pousse, là où auparavant seules des crêtes simples se formaient. Ceux-ci peuvent être visualisés au microscope; lorsque la plante atteint le stade de double crête, la transition végétative à reproductrice est terminée (Li *et al.*, 2018).

La figure 9 montre les différents stades de développement d'une étude récente établie par Zheng *et al.*(2016) pour décrire des stades de développement de la fleur de blé et des épillets (l'unité de base de l'inflorescence des graminées) selon l'échelle de WADDINGTON, CARTWRIGHT and WALL (1983).

Le panneau en haut sur la figure 9, représente la dynamique du développement en trois stades :

- **La différenciation** (de la première apparition du primordium de la fleur [W3,5] à la fin de l'initiation de la fleur [W7]),
- **La dégénération** (c'est-à-dire le stade tétrade du développement de la fleur, de W7 à W9).
- **L'Abortion** (de la maturité du pollen [W9] à la formation du grain).

Dans partie inférieure de la figure 9 :

- En premier lieu, c'est le développement des épillets à l'intérieur de l'épi.
- Deuxième position, c'est le développement des ébauches de fleurons au sein des épillets.
- Troisièmement, un fleuron atteignant avec succès le stade fertile final à la floraison (W10).

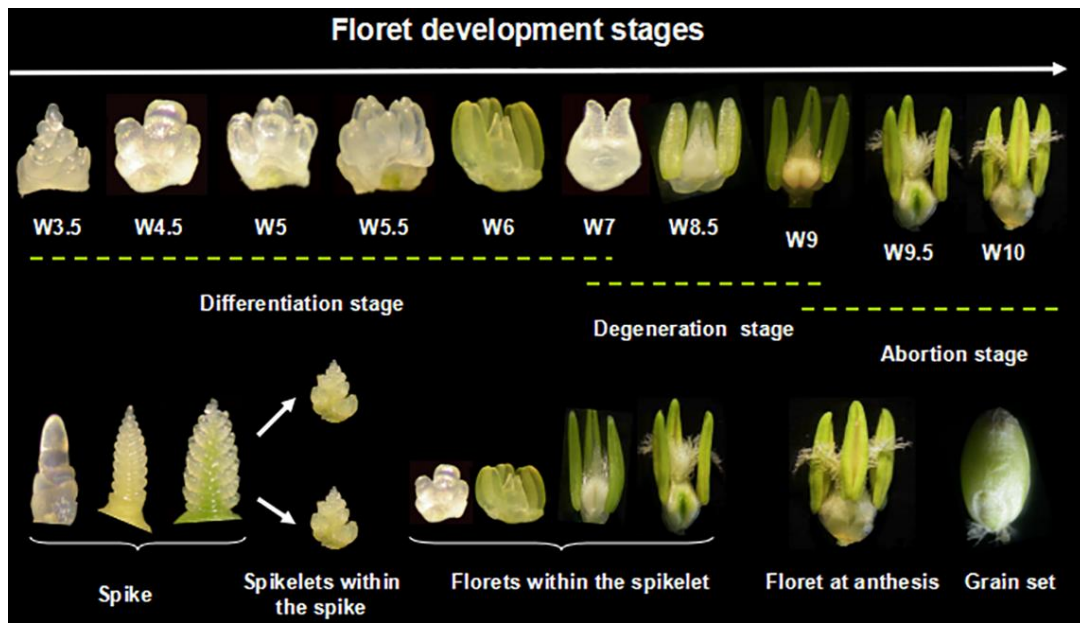


Figure 9: Les stade de développement du fleuron (W : weeks = semaines) (Zheng et al., 2016).

IV.2.2 Fin de la constriction et début de la Production

La floraison ne débute que quelques jours (2-4) après l'émergence de l'inflorescence. Dans la plupart des cas, il est possible d'identifier la floraison par la sortie des anthères. Quelques anthères sur la plante représentent le début de la floraison **BBCH : GS60**, alors qu'à **BBCH : GS65** toute l'inflorescence doit contenir des anthères (Figure 10) (Meier, 2018).



A : le carpelle et ses étamines associées, constituées d'anthères jaunes et de filaments extensibles. **B** : Vue ventrale de grains entiers à 24 et 48 heures après la fécondation. **C** : coupe transversale médiane d'un grain entier à 2 jours après la fécondation. Le sillon ventral est déjà

Figure 10: Aspect clés de la floraison du blé (WheatBP, 2020)

La croissance et le développement du caryopse, après la fécondation, font partie du background théorique le plus important de l'étude. La durée de développement et de croissance du grain de blé après fécondation s'étale jusqu'à 40 jours et durant cette période des changements internes et externes s'effectuent dans le grain. Elle se divise en 03 principaux stades du **BBCH : GS7** au **BBCH : GS9**. La phase développement du grain de blé est une étape importance dans la formation géométrique des grains ainsi que leurs caractéristiques physiques. Il faut noter que le stade **BBCH : GS75** est le stade ou le grain atteint sa taille finale (Xie, Mayes and Sparkes, 2015).

La figure 11 montre neuf jours de développement pendant la floraison, sur cette figure on peut constater que le grain épouse la forme du palea (glumelle).

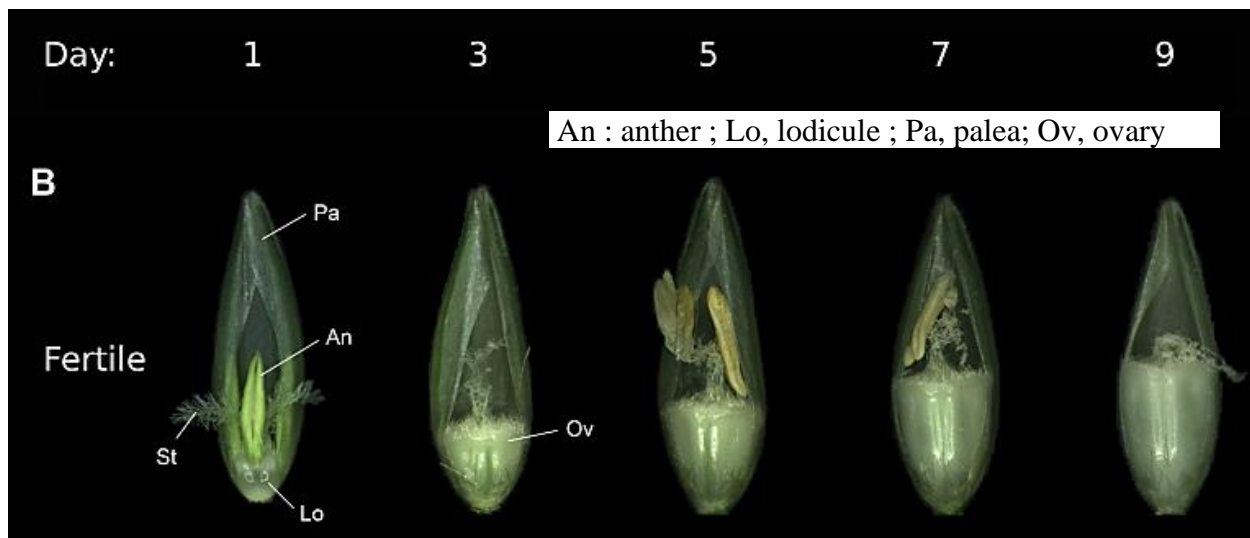


Figure 11: Un fleuron pendant la période de floraison et les débuts de développement des grains (Okada et al., 2018)

Ce qui suit plus bas est une adaptation à partir du site web du projet WheatBP et les travaux dans la littérature EORBTAA (2011); Xie, Mayes and Sparkes (2015); Meier (2018) ; on montre les différents stades de développement du grain post- fécondation avec des illustrations des moments clés dans cette phase du cycle biologique du blé.

I.1.1.A Développement du grain (Stade BBCH GS7) :

- **GS71 : stade aqueux**, les premiers grains ont atteint la moitié de leurs tailles finales
- **GS73 : début du stade laiteux**
- **GS75 : Stade mi-laiteux**, les grains ont atteint leurs tailles finales mais toujours verts
- **GS77 : fin du stade laiteux**, le liquide prendra une couleur laiteuse et apparaîtra également plus épais (figure 14).

Durant les stades secondaires le grain prend la forme et l'aspect illustrés dans les figures 12 ,13 et 14. Les cellules à paroi épaisse du sillon ventral deviennent "l'aleurone du sillon modifié", ces cellules sont responsables de l'absorption des assimilâtes pendant la période de remplissage des grains. Les cellules à aleurone de la face dorsale du grain se comportent différemment.

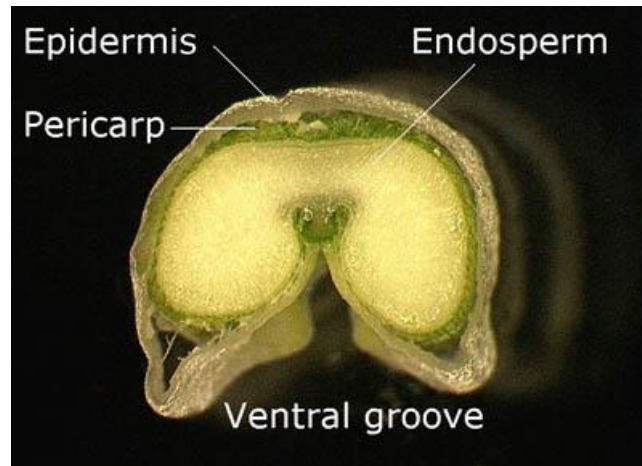


Figure 12: Coupe transversale du grain frais au début du remplissage du grain

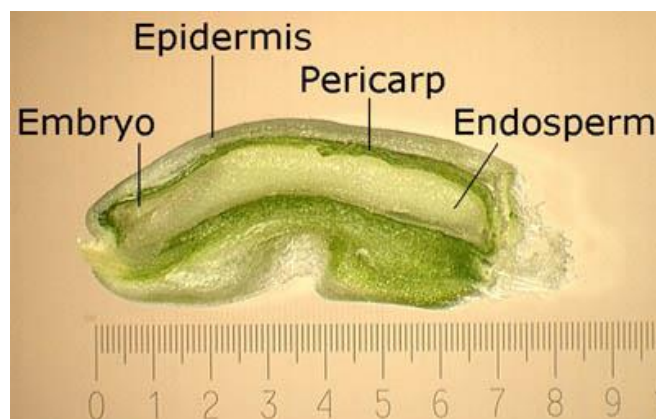


Figure 13: Coupe longitudinale du grain frais à travers le sillon ventral.



Figure 14: grain de blé à la fin du stade laiteux

L' épiderme maternel clair externe recouvre lâchement le péricarpe et laisse pénétrer la lumière. Au fur et à mesure que l' endosperme se remplit d'amidon, les couches externes seront étirées et écrasées.

I.1.1.B Maturation des grains (Stade BBCH GS8)

- **GS83** début du stade pâteux
- **GS85** stade pâteux mou (**soft dough stage**) : contenu du grain est tendre mais sec, une empreinte faite avec l'ongle est réversible.

Le péricarpe maternel change de couleur, du vert au jaune, l'embryon a environ la moitié de sa taille finale. (Figures 15-16)

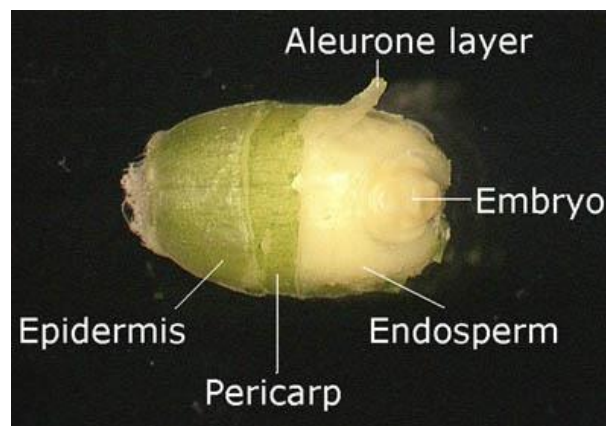


Figure 15: les différentes enveloppes du grain au stade GS85



Figure 16: grain de blé à la fin du stade pâteux

GS87 : stade pâteux dur (**Hard Dough stage**), le contenu du grain est dur, une empreinte faite avec l'ongle est irréversible.

GS89 : maturation complète (**Hard grain stage**), le caryopse est dur et difficile à couper en deux avec l'ongle. (Figure 17)



Figure 17: grain de blé au stade GS89

L'épi de blé entier change maintenant de couleur. Les glumes, lemmes (glumelle et paléole, qui protègent le grain, continuent de se dessécher. Le fleuron 1 (florete 1) a été découpé et ses éléments constitutifs sont séparés (Figure 18).

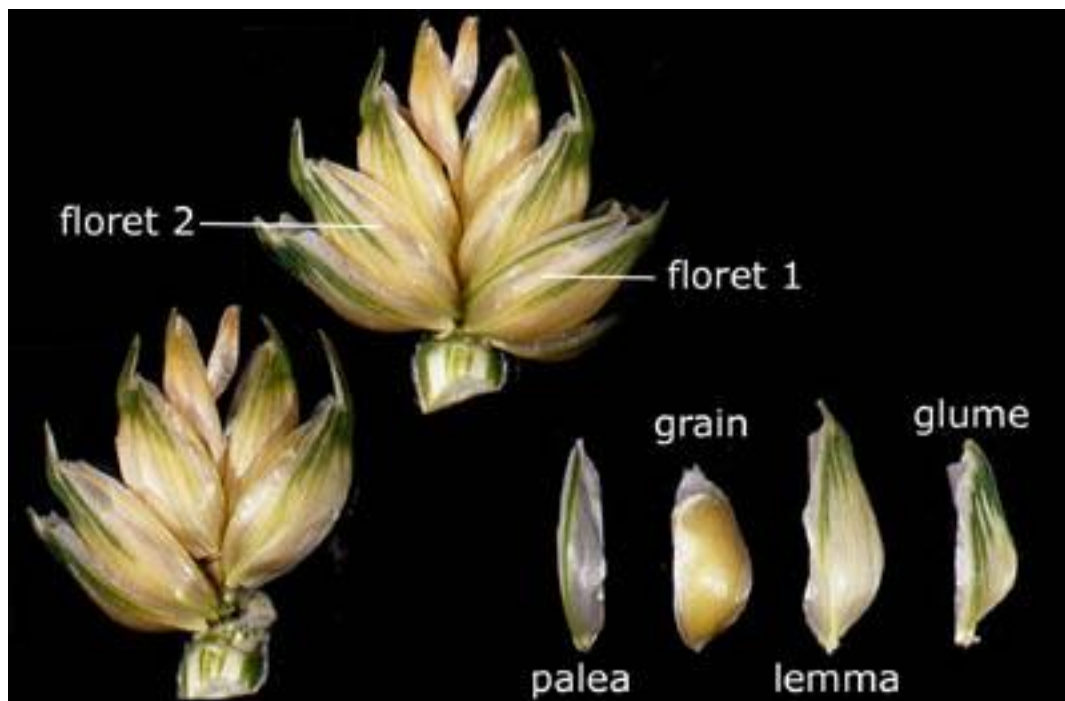


Figure 18 : Un épillet individuel 30 jours après la pollinisation

I.1.1.C La sénescence (Stade BBCH GS9)

- **GS92** sur-maturité, caryopse dur, ne peut pas être marqué à l'ongle (Figure 19).
- **GS93** des grains se détachent.
- **GS97** plante meurt et s'affaisse
- **GS99** produit poste récolte



Figure 19 : La sur-maturité des grains de blé dur

V. Le grain de blé, un ensemble de traits

A la fin du cycle de développement et croissance, en tant que matière biologique, les grains de blé vont avoir des caractéristiques physiologique, biophysique et biométrique diverses (Bechtel *et al.*, 2009; Stone and Morell, 2009). Parmi ces caractéristiques il y a ceux qui donnent une description visuelle de l'apparence du grain (Turnbull and Rahman, 2002) comme les caractéristiques physiques visibles (couleur, texture et/ou vitrosité), la morphologie (la taille, la forme et l'uniformité) et la partie invisible à l'œil humain (Bao *et al.*, 2019) cette dernière est relative à l'aspect microscopique et spectrale (bande spectrale, composition chimique, ...) qui reflètent les aspects internes liés aux propriétés physicochimiques dont les lois de la physique du solide sont applicables (ex. Comportements mécaniques, le poids...) (Kumar *et al.*, 2016).

D'un autre point de vue, la fin du cycle biologique marque le début des phases post-récolte, sous les différents aspects opérationnels du reste de la chaîne du grain, tel que le traitement, le stockage et la transformation. Ainsi que le management de la qualité, ce dernier fera l'objet d'un détail dans les prochaines sections.

V.1. Couleur

Au cours du développement du caryopse, les pigments s'accumulent dans le péricarpe interne, qui devient plus tard une partie du tégument mûr. Selon Bechtel *et al.*, (2009) les caryopses de blé mûr varient donc du chamois clair ou jaune au brun rouge selon l'absence ou la présence de pigmentation rouge dans le péricarpe interne. Dans un cultivar de sélection pure, la couleur ne varie pas, et ainsi les blés peuvent être systématiquement classés comme rouges ou blancs. La

couleur ambrée de certains blés durs est le résultat, non de la pigmentation du tégument, mais de pigments dans l'endosperme visibles à travers les téguments clairs des fruits.

La couleur du grain de blé dépend de l'espèce (Ponce-García *et al.*, 2017), elle est contrôlée par trois *Loci* génétiques distincts, et donc la profondeur de la couleur peut varier entre les cultivars rouges (Freed *et al.* 1976). Occasionnellement, notamment chez certains blés amidonniers éthiopiens, une pigmentation violette dans le tégument a été trouvée.

D'autres facteurs peuvent affecter la couleur naturelle des grains tels que l'infestation de moisissures dans le champ ou pendant le stockage, les dommages causés par la chaleur ou le gel et d'autres causés par les phytopathogènes (Ponce-García *et al.*, 2017). Shahin *et al.*, (2014) ont rapporté que la croissance du mildiou sur les grains de blé réduit la qualité du grain en raison de la décoloration grise caractéristique qui a un impact négatif sur la couleur des farines raffinées.

V.2. Texture

La texture (Turnbull and Rahman, 2002) appelée aussi vitrosité est une propriété optique définie par deux états possibles de l'endosperme (aspect vitreux ou farineux) (Rooney *et al.*, 1983). Comme la pigmentation, il s'agit d'une caractéristique variétale stable, mais elle est également sujette à un degré de variation selon les conditions de croissance ou les conditions climatiques au cours de la maturation (Symes 1961, 1965). La texture de l'endosperme affecte la perception de la couleur du grain (Turnbull and Rahman, 2002).

Selon Delwiche (2010), la texture du blé est déterminée par un mélange final de composants cellulaires pendant le séchage et une éventuelle liaison de composants cellulaires spécifiques à la surface des granules d'amidon.

L'endosperme mou a un aspect crayeux résultant de la diffusion de la lumière au niveau des nombreuses interfaces air-amidon et air-protéines qui existent dans la matrice discontinue. L'absence d'espaces vide dans l'endosperme des blés durs donne au tissu continu un aspect vitreux (Dexter *et al.*, 1989). L'endosperme tendre et crayeux augmente la pâleur des blés blancs et diminue la couleur des blés rouges ; l'inverse est vrai pour les blés durs et vitreux. Il n'est pas rare que l'endosperme contienne à la fois des régions vitreuses et farineuses; ces grains mitadinés se produisent parmi les variétés rouges et blanches (Turnbull and Rahman, 2002).

V.3. Les traits morphologique et les descripteurs géométriques

Les propriétés morphologiques font partie de la description de l'aspect visuel du grain. Parmi ces propriétés on a la taille et la forme des grains de blé. Comme la plupart des caractères d'intérêt biologique et d'importance agronomique et technologique, elles sont devenues des objectifs de sélection importants (Dholakia *et al.*, 2003). Il est important de noter que l'influence principale n'est pas la forme et la taille en soi, mais le degré de variation de ces attributs au sein d'un échantillon (Bechtel *et al.*, 2009).

Au niveau morphologique et génétique, les caractères de forme et de taille du grain montrent une association considérable entre eux. Selon Kumar *et al.* (2016) les traits morphologiques du grain (taille, forme et uniformité) sont des éléments importants des programmes de sélection, car ils déterminent la valeur marchande en raison de leur association avec le rendement à la mouture (c'est-à-dire la qualité et le rendement de la farine). Les grains sont considérés comme des sphères ou des ellipses en raison de leurs formes irrégulières (Gürsoy et Güzel, 2010). La forme du grain de blé peut être décrite comme ronde plus ou moins ovale (approche sphéroïde).

Morphologiquement, le grain de blé présente une rainure sur le côté ventral (le sillon) s'étendant sur toute la longueur du grain et plus profonde au milieu ; cependant, une variation se produit dans l'épaisseur, la grande taille et la largeur du grain. La forme de la rainure est une caractéristique de certaines espèces et cultivars (Evers et Millar, 2002).

Les dimensions axiales, citées plus bas, pour les caryopses bien remplies varient, mais les limites de 4 à 10 mm de long et de 2,5 à 4,5 mm de large sont peu susceptibles d'être dépassées et ne peuvent être atteintes que rarement. La taille des grains sur le même épi varie, ceux du centre étant les plus gros. Sur chaque épillet, les deux grains les plus proches de la base sont les plus lourds, le second en partant du bas étant légèrement plus gros que celui en dessous (Bremner et Rawson 1978 in (Rosentr and Evers, 2018)). Cela reflète le fait que la position du grain détermine l'apport de nutriments au grain en développement (Calderini et Ortiz-Monasterio 2003). Le sol et les conditions météorologiques affectent également la taille et la forme. Le flétrissement dû au gel, aux agents pathogènes ou à d'autres causes peut réduire considérablement la largeur, mais la longueur est rarement affectée (Simmons et Meredith 1979 in (Edwards, 2010)).

Dans un grain de blé, trois dimensions principales sont couramment mesurées (figure 20) : la longueur (L), la largeur (W) et l'épaisseur (T), qui sont généralement déterminées à l'aide d'un micromètre ou d'un pied à coulisse et exprimées en millimètres (Ponce-García *et al.*, 2017). Les

grains de céréales, étant des structures asymétriques, ne peuvent pas être décrits de manière adéquate par des sections dans un seul plan (Bechtel *et al.*, 2009).

Les dimensions des grains de blé dans un épi variaient considérablement et les taux de développement et les dimensions des grains étaient différents (Yinian *et al.*, 2008).

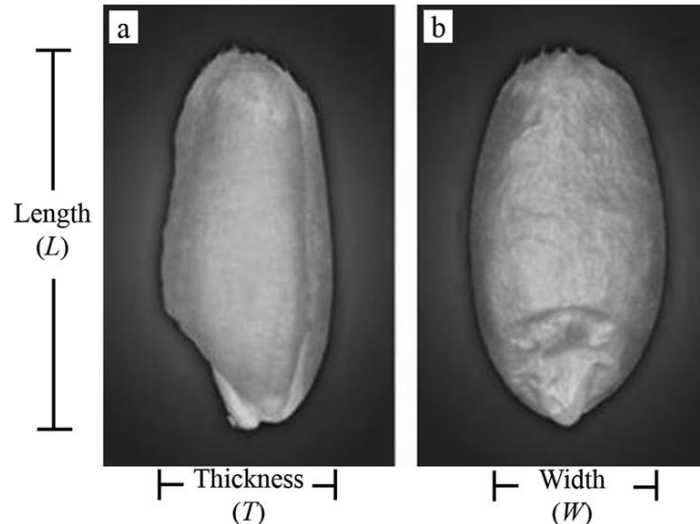


Figure 20: Dimensions axiales dans un grain de blé (Ponce-García *et al.*, 2017)

Dans les variétés élités modernes, bien que le W ait augmenté, le L a diminué, par rapport aux variétés anciennes, en raison de la préférence pour des graines plus uniformes. Cependant, une plus grande variabilité dans la taille et la forme des grains par rapport aux variétés modernes est présente chez le blé hexaploïde primitif (Gegas *et al.*, 2010).

VI. Conclusion

L'acquisition des différentes caractéristiques de chaque grain passe par un cycle biologique qui comprend plusieurs stades de développement ; ce résultat est l'expression génétique spécifique à chaque variété.

La morphologie, la texture, la couleurs et d'autres caractéristiques décrivant le grain ainsi obtenu constituent les paramètres et les signaux qui facilitent la reconnaissance de l'identité du grain.

Le caractère biologique représente à la fois l'origine et l'aboutissement des interactions omnidirectionnelles, cyclique, de l'entité biologique « génome de blé » avec l'environnement, les pratiques agronomiques et les moyens et solutions technologiques déployés; il est **l'origine** parce qu'il s'agit avant tout du matériels génétique présent dans chaque grain de blé ; il est aussi, ce caractère biologique, l'aboutissement, car durant le cycle biologique l'expression génétique et

phénotypique donnent naissance à de nouveaux individus, qui sont dotés d'une identité individuelle et ils partagent aussi en tant que population une identité commune, cette dernière amène à les distinguer en groupes homogènes comme des espèces ou des variétés distinctes ou bien en catégories de qualité.

Chapitre 2 : Opération de l'identification des grains

I. Introduction

Depuis des millénaires, la caractérisation et l'identification des plantes a été une pratique importante de l'Homme cueilleur avant et après qu'il soit devenu agriculteur. Le besoin de faire la distinction entre différentes espèces et variétés de diverses cultures et plantes en général s'est accru en raison de la découverte et de l'identification rapide d'un plus grand nombre d'espèces ainsi que de la sélection et de l'introduction de nouvelles variétés et cultivars (Magnin-Gonze, 2009). Cette pratique est basée sur la perception humaine, impliquant l'observation et l'enregistrement de caractères morphologiques ou de descripteurs divers (Korir *et al.*, 2013).

Dans le cas des grains, le principe de cette pratique consistait à caractériser et à distinguer les grains de valeur et d'intérêt, puis de les reconnaître ainsi que leurs plantes (Magnin-Gonze, 2009); en général l'objectif est soit de les classer qualitativement (en classes, en grades ou en catégories) et/ou bien taxonomiquement (espèce, variétés) (McDonald, 1993; MacDonald, 1994; Shewry, 1996; Wrigley, Batey and Miskelly, 2017). Actuellement, cette activité est capitale pour les sélectionneurs dans le cadre de leurs missions de concevoir de nouvelles variétés ; cette tâche conception nécessite des recherches approfondies de phénotypage et pouvant aller jusqu'à l'échelle moléculaire (Rahaman *et al.*, 2015; Yan *et al.*, 2017) . En effet, le concept est de définir une *identité unique authentifiable* et un *état évaluable* (quantifiable et qualifiable) et de *parvenir à les identifier, à les reconnaître et à les évaluer* (parfois individuellement) à tout moment dans la chaîne des grains. (Shewry, 1996; Wrigley, 2017a).

Comprendre l'opération de l'identification des grains dans tous ses aspects est important pour notre étude. Pour rappel, le contexte de départ est l'évaluation de la pureté d'un lot de semence, précisément l'identification variétale basée sur le grain.

A travers ce chapitre, on cherche à donner quelques définitions des concepts liés à l'opération et les différents contextes d'exécutions. On a essayé d'adopter un nouveau concept pour pouvoir relier l'actuel concept de l'identification variétal à celui qu'on veut proposer dans cette thèse l'intégration de la vision artificielle (Computer Vision et Machine Learning).

II. L'identité et l'état des grains, les enjeux de la qualité dans la chaîne des grains

L'évaluation de la qualité du grain dépend des demandes de l'utilisateur de grain. Ces demandes sont susceptibles d'impliquer différentes combinaisons de caractéristiques en fonction du but pour lequel le grain est acheté et utilisé. D'après Dixon (2007) une évaluation efficace de la qualité du grain fournit des connaissances essentielles sur les défauts et les avantages du grain concerné, depuis la sélection, la culture et la gestion de la culture, jusqu'à la récolte, la ségrégation, le stockage et le transport, jusqu'à la transformation et consommation finale. Aidant ainsi les acteurs concernés à prendre les bonnes décisions sur la meilleure façon de gérer les enjeux de chaque étape du processus global dans l'ordre évaluation-gestion-traitement.

La chaîne de valeur ajoutée pour les céréales (Figure 21) a été surnommée « du champ à l'assiette » (Dixon, 2007).

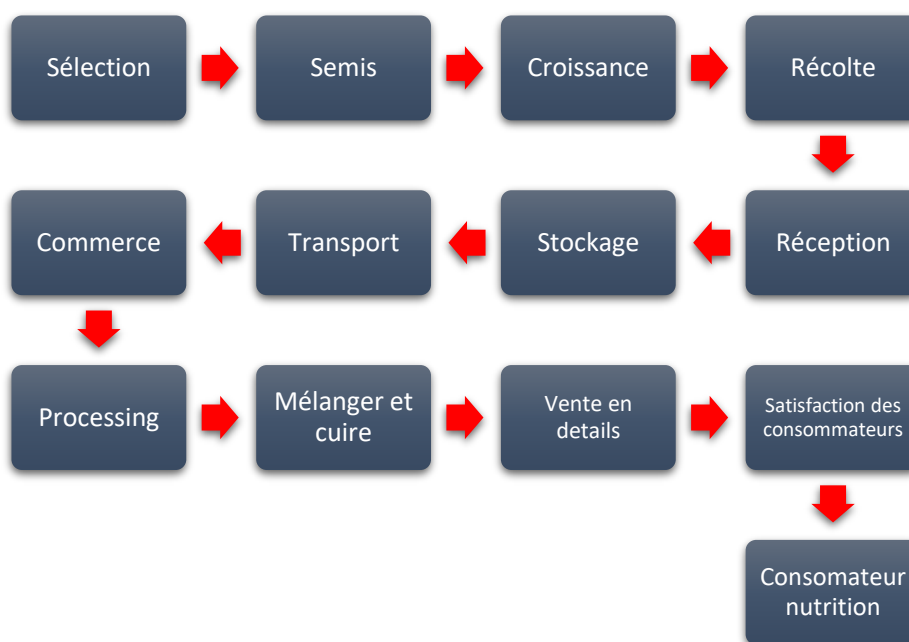


Figure 21: principales étapes de la chaîne des grains

Une gestion efficace de la qualité du grain repose sur des informations analytiques tout au long de la chaîne des grains, fournissant une base solide pour assurer la satisfaction du client. Les principaux aspects de la qualité nécessitant l'évaluation sont mentionnés dans le tableau 1, l'identification de la variété figure parmi les actions clés dans la chaîne des grains et les informations importantes de traçabilité.

Tableau 1: Les principaux aspects de la qualité nécessitant une évaluation à différents stades de la chaîne de grain (Wrigley, 2017).

Stades de production	Aspects demandant évaluation	L'approche analytique
Sélection	Correspondance entre le type de qualité et l'utilisation envisagée	Méthodes de dépistage à petite échelle. Marqueurs de gènes et de protéines.
Production de semences pures	Authenticité de la variété Faculté germinative	Identification de la variété basée sur la composition des protéines ou de l'ADN Test de germination
A la récolte - à la ferme et à la réception du grain	Densité en vrac Teneur en humidité Teneur en protéines (qualité) Contamination et défauts Identification des variétés	Test chondromètre Tamiser, peser et inspecter NIR Spectroscopie
Stockage et transport	Contamination par les insectes Teneur en humidité Mycotoxines	Inspection visuelle NIR Spectroscopie Immuno-essais
Traitement	Contaminants et défauts Teneur en humidité et en protéines	Inspection visuelle NIR
Produits alimentaires	Food safety Apparence, goût	Immuno-essais Réaction client
Production d'aliments pour animaux	Mycotoxines, pesticides Valeur énergétique Nutriments essentiels	NIR Spectroscopie Immuno-essais Chromatographie

On s'intéresse davantage aux trois contextes de la chaîne des grains, la sélection, la production (semence et grains de consommation) et la réception des grains (plus accentué sur la semence) ; dans ces maillons de la chaîne des grains l'authenticité et l'identification des grains sont primordiaux.

L'évaluation des valeurs agronomiques d'un lot de grains, ainsi que la valeur économique (Ponce-García *et al.*, 2017) passe essentiellement par l'évaluation de la qualité de ce lot.

La reconnaissance de l'identité des grains est une opération **taxonomique** de caractérisation et d'identification à la base des descripteurs à pouvoir discriminant ; elle suit la chaîne du grain comme étant une **échelle contextuelle**. L'objectif opérationnel change à chaque niveau ; pour le sélectionneur (l'améliorateur) c'est une question de trouver une nouvelle variété et à de la doter des caractéristiques distinctes. Afin qu'il puisse l'enregistrer au catalogue officielle (UPOV) comme unique (une nouvelle identité) l'obteneur de la variété doit définir une nouvelle identité

variétale. Dans ce contexte, la caractérisation et l'identification est **une échelle taxonomique** (espèces et variété) ; l'identification s'effectue par des moyens **phénotypique** ou **moléculaire** et selon une **échelle temporelle** (au cycle biologique, au post-récolte). (UPOV, 1990 ; OCDE, 2012 ; ISTA, 2018).

Par ailleurs, et **en Post-récolte**, on a deux applications d'ordre contextuel de l'identité et l'état des grains. La première concerne l'identification variétale des grains issus du circuit de production pour la consommation (Shewry, 1996) et la deuxième concerne l'identification variétale des grains de **semence** issu d'un **système semencier** (Posner, 2009; Elias *et al.*, 2012). La deuxième application est primordiale pour l'opération de contrôle de qualité et certification de semence, surtout au moment de la réception des grains de semences. Dans ce dernier cas on s'intéresse à l'évaluation **des puretés** (pureté spécifique, pureté variétale, pureté physique) d'un lot de semence, qui est basée sur la reconnaissance de l'identité et l'évaluation de la proportion des grains et autres objets présents dans un échantillon représentatif d'une récolte (Meyer and Wiersema, 2016; ISTA, 2018a; OCDE, 2019).

II.1. La sélection et la reproduction (Breeding)

Le sélectionneur est responsable du développement de nouvelles variétés qui répondent aux besoins des céréaliculteurs, impliquant trois objectifs majeurs, à savoir : augmenter le rendement en grains, surmonter les obstacles des stress biotiques et abiotiques et améliorer la qualité de transformation et d'utilisation finale du grain (Wrigley, 2017a; Hyles *et al.*, 2020). L'obligation du sélectionneur est de délivrer une semence pure et distinct.

Le développement de nouvelles variétés, en Algérie est réalisé principalement et exclusivement par l'ITGC, qui répondent aux besoins des multiplicateur, ces derniers ont les exigences agronomiques spécifiques qui vont optimiser les rendements en grains, ainsi que les exigences en qualité de grain qui répondent aux règles de qualité du processus de contrôle et de certification, ce dernier est assuré exclusivement par le CNCC.

Par définition, une variété d'une plante cultivée diffère des autres variétés de la même espèce par une ou plusieurs caractéristiques spécifiques. Des caractéristiques telles que la maturité, la résistance à la verse, la résistance aux maladies, la hauteur de la plante et la qualité commerciale distinguent les variétés les unes des autres. Plus récemment, les progrès de la biologie moléculaire ont conduit à la diffusion de nouvelles variétés qui peuvent différer d'aussi peu qu'un seul gène pour un caractère spécifique tel que la tolérance aux herbicides ou la résistance aux insectes.

Vue que Les opportunités du sélectionneur de répondre à ces exigences sont principalement limitées au potentiel génétique (Kong, Guo and Sun, 2015). Selon Wrigley (2017), les modalités d'expression de ce potentiel dépendent des aléas des conditions de croissance et des pratiques de gestion agricole. Il y a donc une interaction ($G \times E$) entre génotype (G, variété) et milieu de développement et croissance (E). Néanmoins, le sélectionneur a une opportunité significative d'intégrer une tolérance génétique à diverses menaces de perte de qualité, telles que la résistance à la germination prématurée due à la pluie à la récolte.

Lorsque de nouvelles variétés sont développées par les sélectionneurs de plantes, une quantité limitée de semences est augmentée à des quantités suffisantes pour répondre aux besoins plus importants des producteurs. Au fur et à mesure que cette graine est augmentée, elle doit être surveillée pour s'assurer que la pureté génétique de la graine de sélection n'est pas compromise.

Les résultats d'évaluation de la qualité seront essentiels au stade où des lignes spécifiques seront finalement approuvées pour la diffusion. À ce stade, la gestion de l'enregistrement impliquera un examen DHS, à savoir l'évaluation que la nouvelle variété potentielle est distincte, uniforme et stable, répondant ainsi aux directives de l'Union internationale pour la protection des obtentions végétales (UPOV, 1990).

II.2. La multiplication et la production des grains

Après la sélection et l'enregistrement de la nouvelle variété au catalogue national, deux catégories de producteurs de grain vont intervenir : les producteurs de semences (appelés aussi les multiplicateurs) qui représentent le programme de multiplication ; et les céréaliculteurs qui représentent le programme de consommation.

L'évaluation principale de la qualité dans programme de multiplication, porte sur l'assurance que les semences commercialisées, semées et réceptionnées sont la bonne variété et qu'elle est exempte de semences de mauvaises herbes et de maladies transmises par les semences.

Des évaluations détaillées de la pureté et de l'identité des semences doivent être fournies à l'agriculteur comme assurance des aspects de qualité au moment du semis.

Lorsque la responsabilité de la gestion est transférée à l'agriculteur, la gestion de la qualité du grain à la ferme est interactive avec le rendement en grain, cette interactivité est représentés par la combinaison du volume de grain et de la valeur marchande (MacDonald, 1994). Ainsi, la gestion (M, management) de la culture du grain par l'agriculteur s'ajoute à l'interaction entre génotype (G, variété) et environnement de croissance (E) pour fournir la combinaison de trois facteurs : ($G \times E \times M$). (Nuttall *et al.*, 2017).

II.3. La Réception des grains

L'évaluation de la qualité à la réception va déterminer le mode de ségrégation de la récolte. Ce moment présente une opportunité majeure de gérer la qualité du grain, à condition que des informations analytiques adéquates soient disponibles rapidement (Sapirstein and Kohler, 1999; Venora, Grillo and Saccone, 2009). Concernant le programme de semence, c'est une question d'authentification de la variété en premier lieu et l'évaluation de la qualité dans le cadre des phases des tests de semences (Seeds Testing), à savoir essentiellement l'analyse de la pureté du lot de semence réceptionné (Elias *et al.*, 2012; ISTA, 2018b).

Dans de nombreuses régions céréalières, les grains de céréales récoltées sont livrées directement à une structure de stockage régionale, appelée « élévateur de campagne » en Amérique du Nord ou « silo » en Australie (Wrigley, 2017a) et les CCLS en Algérie. Selon (Copeland and McDonald, 2001), les grains telle qu'ils viennent du champ ne sont presque jamais purs. La récolte réceptionnée contient généralement des grandes quantités de déchets, de feuilles, de segments de mauvaises herbes, d'autres graines de cultures et d'insectes, d'où la nécessité de traitement de la récolte. Selon le même auteur, le processus d'élimination de ces matières indésirables dans le cas d'un lot de semences, ainsi que l'amélioration globale de la qualité des semences, sont connus sous le nom de conditionnement des semences.

Selon Wiregly (2017), la priorité est plus souvent donnée pour réceptionner et d'entreposer la récolte avant que les conditions favorables à la récolte changent, plutôt que de passer un temps précieux à vérifier que la qualité du grain est correctement évaluée, pour permettre le stockage séparé de grains de différents types de qualité (MacDonald, 1994; Delwiche, 2010). Au-delà de cette étape, il est nécessaire de préserver la valeur du grain agrégé qui a été évalué comme étant d'une qualité spécifique, en le stockant et en le transportant avec son type de qualité définie (Wrigley, 2017a).

Enfin, lorsque l'agriculteur livre le grain au site de réception, il doit y avoir un moyen pour déterminer rapidement si le grain est réellement la variété déclarée, pour empêcher le mélange de différentes variétés ou classes de qualité, et pour la perception finale des redevances. Ce besoin est reconnu depuis longtemps et les scientifiques ont développé des méthodes pour identifier avec précision les différentes variétés.

Dans les cultures où la qualité est un différenciateur de valeur et où il existe une variation significative de la qualité du produit final entre les différents cultivars, il est nécessaire d'identifier avec précision le génotype d'une manière rapide et efficace. Ce besoin est important à toutes les

étapes de la chaîne de valeur et commence dans le processus de sélection où le sélectionneur doit être capable de distinguer les nouvelles variétés de celles déjà disponibles. Dans de nombreux pays le sélectionneur doit être capable de démontrer la distinction, l'homogénéité et la stabilité (DHS) (Cooke, 1995 ; Cooke et Wrigley, 2004).

III. Test de semence

La semence a été et pour toujours le moyen unique et essentiel pour la régénération et la propagation des blés, sans cette semence il n'y aurait et il n'y auras probablement pas d'agriculture ni d'aliments.

Par définition l'analyse des semences est l'art et la science d'évaluer la qualité des semences à des fins agricoles et économique. Bien qu'initialement développé pour évaluer la qualité de la plantation des semences de grandes cultures et de légumes, il est également utile pour déterminer la qualité des semences de toutes les espèces (Powell, 2009; Elias *et al.*, 2012).

Partout dans le monde, les agriculteurs et les producteurs de semences ont des exigences claires vis-à-vis des semences qu'ils sèment. Premièrement, ils veulent que l'espèce et la variété correspondent à ce qu'ils croient avoir acheté. Deuxièmement, ils veulent que ces semences viables, soient pures et saines, ce qui permettra l'établissement uniforme et réussi d'une culture (Powell, 2009). La réalisation de ces exigences est facilitée par les méthodologies d'analyse de la qualité des semences qui sont développées et normalisées par des organisme local et international comme l'International Seed Testing Association et d'autres.

Des procédures d'essai conventionnelles ont été développées et normalisées pour déterminer de nombreux attributs de qualité du grain requis tout au long de la chaîne céréalière qualité (MacDonald, 1994; Delwiche, 2010; ISTA, 2018c). La plupart de ces analyses font appel à des méthodologies conventionnelles qui nécessitent des équipements de laboratoire et des niveaux d'expertise importants de la part de l'opérateur (CCLS, CNCC). Le développement de ces méthodes standard est un processus laborieux, impliquant une collaboration entre diverses combinaisons de scientifiques d'un éventail de laboratoires impliqués dans le type d'analyses évaluées (Wrigley, 2016 a,b).

Le test des semences est un processus essentiel (Copeland et McDonald, 2001 ; Powell, 2009) qui fournit les informations de qualité nécessaires sur la pureté, la viabilité, la germination et la teneur en graines de mauvaises herbes nuisibles, pour l'étiquetage des semences à vendre (Elias et al., 2012). Un point à souligner d'emblée est que tous les tests sont effectués sur des échantillons

tirés du lot de semences, qui est la population qui sera semée. Les méthodes d'échantillonnage des semences sont décrites dans les règles de l'ISTA pour l'essai des semences (ISTA, 2016, 2018d), avec plus d'informations et de détails dans le manuel de l'ISTA sur l'échantillonnage des semences (ISTA, 2021).

III.1. Test de pureté

Avec la germination, les tests de pureté sont parmi les tests les plus anciens et les plus courants effectués par les analystes de semences (Elias *et al.*, 2012). Le test de pureté des semences est effectué impérativement pour déterminer le niveau de pureté physique et variétale d'un lot de semences, en effectuant une séparation détaillée et précise sur un petit échantillon de travail représentatif. L'intégrité génétique du cultivar de blé d'origine peut être modifiée par un mélange mécanique et un étiquetage incorrect (Copeland et McDonald, 2001).

La procédure consiste à séparer l'échantillon en quatre composants (graines pures, graines d'autres cultures, graines de mauvaises herbes et matière inerte) afin que la composition en pourcentage en poids de chacun puisse être déterminée.

La semence pure est la portion de l'échantillon de travail représentée par l'espèce cultivée pour laquelle le lot est testé (Figure 22).

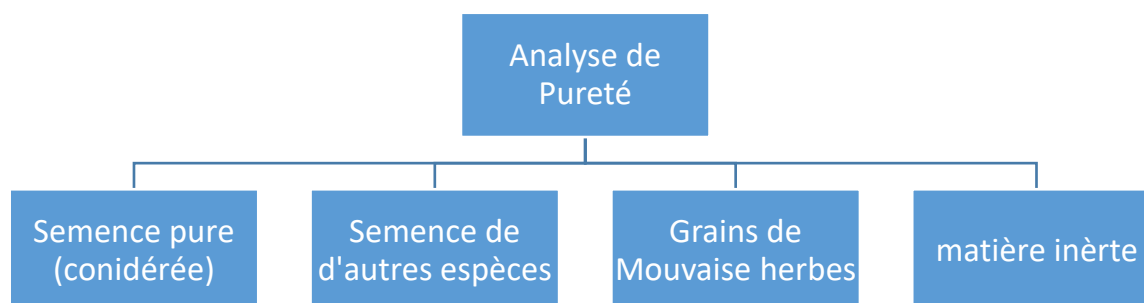


Figure 22: Les composants de l'échantillon dans l'analyse de pureté conformément aux règles de l'AOSA

Selon (Laabassi *et al.*, 2021), le test de pureté est une opération qui est basée essentiellement sur la reconnaissance de l'identité du grains (*Grain Identity Recognition (GIR)*), ce test est réalisé dans le contexte du programme national, selon une approche de classification taxonomique et phénotypique.

D'une manière générale, le **testeur** de semences effectue plusieurs niveaux de classification (Laabassi *et al.*, 2021) :

- Une classification au niveau objet (IO/OSLC) pour le test physique (Pureté Analytique).
- Une classification au niveau d'espèces (SLC) pour le test de pureté spécifique.
- Une classification au niveau variétal (VLC) pour la pureté variétale.

En s'appuyant sur :

- La Perception Humaine.
- La Perception machine.

Par deux moyens d'analyse pour la SLC et la VLC :

- L'analyse des grains (rapide et opérationnel).
- L'analyse de la plante durant sans cycle biologique (pour la nouvelle récolte sera un test *posteriori* lent).

En utilisant deux principales méthodes d'analyses :

- Une analyse phénotypique incluant :
 - Visuelle non destructive et directe à la base des caractéristiques visuelles (ISTA, 2018 ; Meyer et Wiersema, 2016 ; OCDE, 2019).
 - Une analyse visuelle destructive : analyse de comportement : mécanique, à la germination, réaction aux réactifs et produits chimiques (pesticides).
- Analyse moléculaire incluant :
 - Une analyse non destructive à base spectrales et proche infrarouge.
 - Une analyse destructive à base de protéines.
 - Une analyse destructive à base d'ADN.

III.2. Pureté analytique

L'analyse de la pureté analytique de la graine examine dans quelle mesure un échantillon de graine est contaminé par d'autres graines (mauvaises herbes et autres cultures) et d'autres plantes et matières inertes (ISTA, 2015).

Les méthodes d'analyse de pureté impliquent l'observation d'échantillons de graines à l'aide de lentilles et de microscopes et la séparation des graines en différentes portions. Les lentilles à main ou les microscopes binoculaires aident à l'identification et à la séparation des petites unités de

graines et des fractions de grains ; des tamis peuvent être utilisés pour séparer les déchets, la terre, les petits morceaux de graines et autres petites particules ;

Le résultat du test de pureté analytique révèle le pourcentage en poids de graines pures présentes dans un échantillon, les autres graines (qui sont identifiées) et la matière inerte.

III.3. Pureté génétique ou Test de variété

Selon Elias et al.(2012), une variété d'une plante cultivée diffère des autres variétés de la même espèce par une ou plusieurs caractéristiques spécifiques. Selon Les tests de variétés comportent deux aspects. La première est de s'assurer qu'un échantillon en effet est l'espèce (SLC) ou la variété (VLC) requise et la seconde de s'assurer de la pureté de la variété, c'est-à-dire que la variété n'est pas contaminée par les semences d'autres variétés (VLC).

Selon Smith and Register (1998) les tests de pureté génétique sont importants pour que :

1. La protection de la propriété intellectuelle par le biais de la protection des variétés végétales ou des brevets d'utilité puisse être obtenue puis maintenue par la suite,
2. Les variétés puissent être créées avec une apparence uniforme et des performances agronomiques qui répondent aux exigences des agriculteurs, des transformateurs et des consommateurs,
3. Des variétés avec des identités génétiques stables peuvent être créées afin que les performances des plantes puissent être aussi prévisibles que possible compte tenu des fluctuations environnementales imprévisibles.
4. Les sélectionneurs peuvent caractériser complètement et plus précisément et mesurer la diversité génétique, afin que les ressources génétiques puissent être évaluées de manière approfondie en termes de performance des plantes et utilisées efficacement pour la création de variétés améliorées.

Sur le plan opérationnel, Cooke (1998) a défini cinq raisons principales pour soumettre des grains à des tests variétaux :

1. Identification au sens strict - Quelle est la variété ?
2. Confirmation/vérification de l'identité - Cet échantillon est-il la variété « X » ?
3. Test de distinction - Cette variété est-elle différente de toutes les autres ?
4. Pureté génétique - Cet échantillon contient-il plus d'une variété (et, si oui, à quel niveau) ?
5. Caractérisation génétique - Quelle est la description de la variété ?

Les méthodes traditionnelles d'analyse des variétés et largement utilisée, comprennent les méthodes phénotypiques morphologiques. Dans de telles méthodes, les caractéristiques des graines peuvent être comparées ; par exemple, les différences de couleur des graines peuvent révéler que les variétés ont été mélangées. Alternativement, les caractéristiques des plantules peuvent être observées en laboratoire ou sur le terrain, ou d'autres caractéristiques des plantes ou des fruits peuvent être observées sur le terrain. Des caractéristiques telles que la maturité, la résistance à la verse, la résistance aux maladies, la hauteur de la plante et la qualité commerciale distinguent les variétés les unes des autres.

Des méthodes plus modernes d'analyse variétale peuvent faire appel à toute une gamme de techniques biochimiques et moléculaires. Les méthodes biochimiques comprennent l'analyse des réserves protéiques des graines par électrophorèse. Les méthodes moléculaires comprennent l'utilisation de marqueurs moléculaires. Ces méthodes impliquent l'extraction de l'ADN et la réaction en chaîne par polymérase (PCR) dans laquelle l'ADN sélectionné est multiplié. Une autre approche est l'utilisation de microsatellites et de séquences répétées simples (SSR).

IV. Conclusion

La distinction entre les variétés de *Triticum* spp sur la base des caractéristiques visuelles de leur caryopse est difficile, en raison du niveau élevé de similitude morphologique des graines entre ces espèces (Chiara Delogu, 2013). L'identification des variétés nécessite une expertise et une expérience de haut niveau, car les caractéristiques utiles telles que la taille, la forme et la couleur, sont également affectées par les conditions de croissance, (Wrigley et Batey, 2010 ; Howitt et Diane, 2017 ; Patrício et Rieder, 2018). Bien que l'identification précise des grains individuels, ou même des cultivars individuels, ne puisse être assurée uniquement par la couleur et la morphologie globale, ces caractéristiques fournissent une base utile pour le tri préliminaire.

Une gamme de méthodes a été utilisée pour l'identification variétale (Korir et al., 2013 ; Howitt et Diane, 2017), allant d'une identification visuelle humaine, des techniques physico-chimiques par ex. la reconnaissance des motifs (J, Smith et Alberico, 1991) et l'analyse avancée de l'ADN (Nader, Brendel et Schubbert, 2016).

L'analyse visuelle des grains présente un grand avantage en termes de temps par rapport à l'analyse de la plante entière qui nécessite plus de temps et dépend du stade du cycle biologique.

En effet, les traits phénotypiques des plants de blé sont perdus aux stades post-récolte et le caryopse fait partie du reste disponible pour reconnaître la variété de blé. (Wrigley et Batey, 2010).

Le principe de l'identification est basé sur les mécanismes de la perception humaine des attributs contenus dans les grains qui permettent de développer une identité et une image mentale du grain de blé, l'opération de l'identification est donc une problématique de classification et de détermination des caractéristiques les plus discriminantes pour les différents niveaux de classification et aussi différents cas d'études, ces concepts sont développés dans le chapitre suivant.

Chapitre 3 : Identité Biodigitale du grain

I. Introduction

De nouveaux défis émergent dans l'agriculture en termes de durabilité (Kuhlman and Farrington, 2010), de précision (Patrício and Rieder, 2018), d'intelligence (Wolfert *et al.*, 2017) et de décision (Barry *et al.*, 2017). Un réel besoin se fait sentir pour de nouveaux concepts, systèmes et méthodes innovantes afin d'assurer un flux de travail à haut débit, précis, productif et efficace. Dans ce contexte, des solutions technologiques et Agroéquipements intelligentes, des normes, des protocoles et des règles de pratiques doivent être reconçues dans un écosystème multidisciplinaire, incluant les technologies des systèmes d'information et des outils d'aide à la décision (O'Grady and O'Hare, 2017).

L'intégration et le développement de la Computer Vision (CV), du Machine Learning (ML) et de l'intelligence artificielle (IA) dans les activités agronomiques vise à améliorer l'efficacité et les conditions de travail ; ceci par l'automatisation de toutes les opérations agronomiques basées sur la perception et la décision humaines et tout en concevant un environnement d'interaction par de nouvelles méthodes ergonomiques, intelligentes, plus efficaces, plus précises et traçable. Ainsi on va de capturer, exploiter et explorer une quantité massive de données (Bigdata) générées à partir de ces activités agronomiques (Wolfert et al. 2017).

Dans le cas de la chaîne des grains tel que nous l'avons décrit dans les chapitres précédents, l'intérêt critique de la classification automatique des grains de blé est principalement lié aux concepts de l'évaluation de la qualité, du contrôle et de la traçabilité du produit céréalier et aussi de la réduction de l'effort et de l'erreur humaine à tous les niveaux de la chaîne des grains. Par conséquent une classification intelligente et automatique permet d'augmenter le traitement de travail, l'objectivité et la précision.

Dans ce contexte, les techniques de Computer Vision et de Machine Learning, qui permettent d'analyser, de caractériser et de classer les images, sont d'utiles solutions pour l'identification des variétés de blé et la continuité de l'évolution sous la contrainte des nouvelles orientations des solutions technologiques et industrielles. Nous avons considéré que la vision Artificielle comme étant la solution technologique et qu'elle est l'implication de la Computer vision et le Machine Learning dans le cadre de l'intelligence artificielle.

A rappeler que nous avons proposé pour entreprendre le problème de l'identification variétale, l'utilisation d'un processus de classification d'images par l'intégration de la solution technologique « vision artificielle ».

L'approche doit garder les critères suivants :

- La perception artificielle qui peut remplacer la perception humaine.
- La préservation des avantages du VGA en utilisant le grain entier
- La capacité à effectuer une VLC avec précision.

Les approches automatiques pourraient améliorer considérablement les conditions de travail des tests de pureté, ainsi que d'autres avantages liés à sa mise en œuvre.

Ce chapitre introduit globalement les concepts clés de la solution technologique proposée pour le problème de l'identification variétale des grains de blé. A ce stade l'objectif est la construction intégrative d'un background théorique et de compromis entre la Computer Vision et le machine Learning d'une part et d'autre par ce qui a été avancé dans les deux sections précédentes, à savoir la nature biologique de l'objet « grain de blé », la nature agronomique de l'opération de l'identification en son mode perception humaine (test de pureté) ; cette opération est en fait la reconnaissance de l'identité des grains de blé en niveau variétale.

II. La Perception Machine (machine perception)

La perception machine est le domaine de recherche sur la possibilité de construire des systèmes qui peuvent être dotés de la capacité de la perception humaine (Nevatia, 1982). La perception machine concerne principalement *la reconnaissance d'objets, la détection d'objets et la résolution de problèmes de navigation et de positionnement*. La recherche en perception machine est largement liée à d'autres domaines comme : *la vision artificielle, la reconnaissance des formes, la robotique et la compréhension d'images*. Les machines qui perçoivent leur environnement et effectuent les tâches requises ont une utilité évidente pour divers domaines d'application tels que l'assemblage et l'inspection industriels, l'exploration spatiale planétaire ou le dépistage médical automatisé par rayons X (Les and Les, 2020).

III. L'intelligence artificielle

Les nouvelles technologies sont conçues pour rendre un processus spécifique plus facile, plus précis, plus rapide et moins cher. Elles permettent aussi d'effectuer des tâches ou de créer des objets qui étaient auparavant irréalisables (Anantrasirichai and Bull, 2021). Un des cas le plus marquant et le plus avancé de cette ère, sur le plan technique et scientifique, est l'intelligence artificielles (IA). Cette dernière a pour but d'imiter la perception humaine à travers un ensemble de codes, de techniques, d'algorithmes et de données ; cet ensemble permet à un système informatique de développer et d'imiter un comportement humain et donc de prendre des décisions similaires (Wolfert *et al.*, 2017; Liakos *et al.*, 2018; Lu and Young, 2020).

(Pantazi *et al.* (2020) ont défini l'IA comme étant un domaine de l'ingénierie qui tente de reproduire (reverse engineering) les fonctions du cerveau humain in silico. Ils précisent que le terme « intelligence » est suivi de définitions typiques, notamment :

1. Capacité à comprendre.
2. L'acquisition d'informations.
3. Capacité innée à interpréter l'environnement.
4. La capacité de comprendre les événements, les faits ou la conscience de la situation.
5. La capacité de résoudre des problèmes par la planification et le raisonnement, la dérivation d'abstractions, la capture d'idées, la compréhension du langage et le comportement adaptatif.

Les notions théoriques de la Computer Vision et le Machine Learning et leurs implémentations, pivotent donc autour du concept de l'intelligence artificielle.

IV. Computer vision

La CV comprend les méthodes et les techniques à travers lesquelles des systèmes de vision artificielle peuvent être construits et utilisés de manière raisonnable dans des applications pratiques (Patrício et Rieder, 2018). Ce domaine de l'informatique comprend les logiciels, le matériel et les techniques d'imagerie nécessaires à ces méthodes (Davies, 2005).

La technologie de CV, selon Chen, Lin and Zhao (2021), est une description explicite et significative d'une image d'un objet physique en lumière visible. L'évaluation des caractéristiques d'apparence est une méthode d'évaluation couramment utilisée dans le domaine de la qualité des produits, y compris l'apparence, la taille, l'évaluation de la couleur, l'analyse des caractéristiques de texture, l'analyse des défauts et des dommages et l'analyse de la composition des produits alimentaires et agricoles.

Chen et al. (2021) et Sun (2016) ont adapté aussi une description structurée de la CV appliquée dans le domaine de l'alimentation, ils stipulent que les principales tâches de la technologie de CV sont : l'acquisition des images en utilisant des capteurs d'images physiques et des logiciels (acquisition de l'image) et du matériel informatique adapté, ensuite d'analyser et de traiter les images (traitement et analyse d'image) avec un objectif ou un cas d'étude bien défini.

Davies (2012) et Zaitoun & Aqel (2015) ont défini fondamentalement la CV en trois niveau (figure 23) :

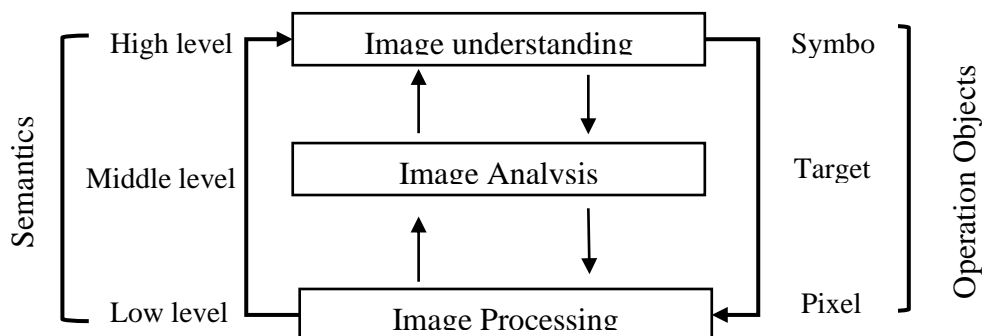
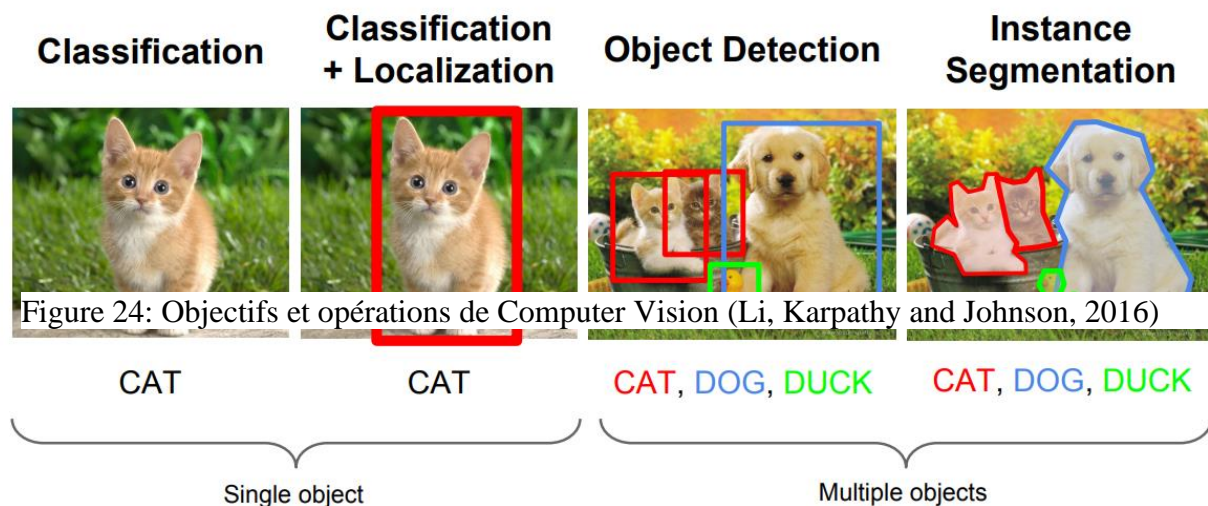


Figure 23: Les trois niveau de vision par ordinateur (Zaitoun and Aqel, 2015)

Elle comprend :

- Vision de bas niveau : englobant les opérations d'acquisition et de traitement.
- Vision de niveau intermédiaire : c'est l'analyse d'image et l'extraction des caractéristiques.
- Vision de haut niveau : représente la compréhension du contenu de l'image

La figure 24 montre selon Li, Karpathy and Johnson (2016) les différentes opérations effectuées et les objectifs à atteindre en Computer Vision.



IV.1. Vision de bas niveau (Low level vision)

Elle englobe les opérations d'acquisition et les opérations de traitement de l'image (Davies, 2009; Patrício and Rieder, 2018); selon Jähne (2005) un système d'acquisition et de traitement d'images se compose généralement de quatre composants essentiels :

1. Un système d'acquisition d'images. Dans le cas le plus simple, il peut s'agir d'une caméra CCD (*charge coupled device*), d'un scanner à plat ou d'un enregistreur vidéo.
2. Un dispositif connu sous le nom de carte d'acquisition d'images (frame grabber) pour convertir le signal électrique (un signal vidéo analogique) du système d'acquisition d'images en une image numérique qui peut être stockée.
3. Un ordinateur personnel ou une station de travail (workstation) qui fournit la puissance de traitement.
4. Un logiciel de traitement d'images qui fournit les outils pour manipuler et analyser les images.

IV.1.1 Acquisition d'images :

L'acquisition d'images est la première étape du traitement d'images numériques (Jähne, 1993). L'analyse quantitative de toute image nécessite une bonne compréhension du processus de la formation de l'image. Seule une connaissance approfondie de toutes les étapes de l'acquisition d'images permet d'interpréter correctement le contenu d'une image.

Les étapes nécessaires pour rendre un objet du monde tridimensionnel en une image numérique stockable dans la mémoire d'un ordinateur sont les suivantes selon Jähne, B. (1993):

Devenir visible : Un objet devient visible par l'interaction avec la lumière ou, généralement, un rayonnement électromagnétique. Les quatre types d'interaction de base sont la réflexion, la réfraction, l'absorption et la diffusion. Ces effets dépendent des propriétés optiques du matériau à partir duquel l'objet est fabriqué et de sa structure de surface. La lumière collectée par un système de caméra est déterminée par ces propriétés optiques ainsi que par l'éclairage, c'est-à-dire la position et la nature des sources de rayonnement.

Projection : Un système optique collecte les rayons lumineux réfléchis par les objets et projette le monde tridimensionnel sur un plan image bidimensionnel.

Numérisation (Digitization) : L'image en continue (continuous image) sur le plan d'image doit être convertie en points image sur une grille discrète. De plus, l'intensité à chaque point doit être représentée par un nombre déterminé approprié de *valeurs de gris* (Quantization).

Au moment de l'acquisition d'images, les signaux électroniques d'un capteur sont transformés vers une représentation numérique ou digitale ; deux types de capteur sont utilisés, les capteurs à *balayage en zone* qui génèrent une image à chaque cycle d'exposition et les capteurs à *balayage linéaire* qui ne capturent qu'une seule ligne de pixels à la fois. Pour l'acquisition d'images en deux dimensions comme pour le scanner, il est nécessaire de déplacer l'objet à capturer à l'aide d'un convoyeur ou de déplacer la caméra le long d'un objet fixe. (Patrício and Rieder, 2018).

La qualité de l'image acquise par un système de CV est directement affectée par l'éclairage utilisé lors de la phase d'acquisition. L'utilisation d'un éclairage adéquat augmentera les performances et la fiabilité du système, réduisant la complexité du software utilisé dans l'étape de traitement (Hornberg, 2017).

IV.1.2 Traitement d'image

Il implique des tâches pour manipuler des images numériques dans le but d'améliorer leur qualité, de réduire le bruit ou de corriger des problèmes d'éclairage. Ces opérations produisent une nouvelle image et sont utilisées pour améliorer la qualité de l'image ou pour modifier la position de l'objet d'intérêt par des transformations géométriques (Hornberg, 2017).

De plus, l'analyse d'image fait référence au processus de différenciation des régions d'intérêt des autres régions pour extraire des informations.

IV.2. Vision de niveau intermédiaire

Quant aux processus de niveau intermédiaire, ils impliquent des opérations de segmentation (partitionnement des images en régions), de description et de classification des objets présents dans l'image (Jähne, 1993, 2005; Davies, 2009; E. R. Davies, 2012) :

IV.2.1 La segmentation d'image

La segmentation permet de définir un ensemble de contours ou de régions distinct. L'extraction des attributs qui caractérisent ces régions est nécessaire pour évaluer une série de caractéristiques de la région d'intérêt. Par exemple, les paramètres d'ellipse permettent de déterminer l'orientation et la taille de la région. En identifiant donc quel ensemble de pixels de l'image appartient à quel objet ou région d'intérêt (Radhakrishna and Junsong, 2020).

La boîte englobante (The bounding box) permet de calculer la hauteur et la largeur de la région d'intérêt (Schaeffel, 2017). Des filtres tels que *Sobel*, *Laplacien* et *Laplacien de Gaussien* peuvent être utilisés pour déterminer les régions où des changements brusques d'intensité se produisant dans l'image. Le filtre *Gabor* est sélectif pour les éléments de l'image dans une certaine plage d'orientation et de fréquence. Les filtres de type *Haar* sont constitués de régions rectangulaires adjacentes qui sont équilibrées où la valeur moyenne du filtre est nulle et invariante aux changements d'éclairage. (Katiyar and Arun, 2014; Patrício and Rieder, 2018).

Le but de la détection des bords est de produire une image binaire où les valeurs non nulles indiquent la présence d'un bord dans l'image. Les détecteurs peuvent éventuellement et également renvoyer d'autres informations telles que l'échelle et l'orientation par rapport à un bord. Quelques exemples de détecteurs de bord sont *Canny*, *Harris* et *SIFT*. Le détecteur *SIFT* (Scale Invariant Feature Transform) est une méthode d'identification des points d'intérêt. Il associe des

informations d'échelle et d'orientation pour chaque point résultant du processus de détection (Prince, 2012 *in* Patrício and Rieder, 2018).

IV.2.2 L'extraction des attributs (Feature Extraction)

L'efficacité et la robustesse d'un système de vision sont souvent largement déterminées par la qualité des attributs de l'image dont il dispose. Selon Dollár et al.(2007) une bonne caractéristique (attribut) doit être : (1) informative, (2) invariante au bruit ou à un ensemble donné de transformations, et (3) rapide à calculer. En outre, dans certains paramètres (4), une faible densité (sparsity) de la réponse des attributs, soit entre les images, soit au sein d'une seule image, est souhaitée.

Dans la technologie de traitement d'image, qu'elle soit binaire, en couleur ou en niveaux de gris, le traitement d'image peut être réalisé en extrayant des attributs, afin d'effectuer l'identification, la classification, le diagnostic, le clustering, la reconnaissance et la détection. La méthode d'extraction d'attributs (Feature Extraction) est utilisée pour obtenir autant d'informations que possible sur l'image. (Mutlag *et al.*, 2020).

Selon Kumar and Bhatia (2014) un bon ensemble d'attributs contient des informations discriminantes d'un objet visé. Il doit être le plus robuste possible afin d'éviter de générer des codes de caractéristiques différents pour les objets d'une même classe. Les mêmes auteurs classent les attributs en deux catégories :

1. *Les attributs locaux*, qui sont généralement géométriques (par exemple, parties concaves/convexes, nombre d'extrémités, branches, articulations, etc.).
2. *Les attributs globaux*, qui sont généralement topologiques (connectivité, profils de projection, nombre de trous, etc.) ou statistiques (moments invariants, etc.)

L'extraction de caractéristiques est un problème de réduction de dimensionnalité dans le but de trouver des projections significatives des vecteurs de données d'origine (Jia, Kuo and Crawford, 2013). Pour cela, de nombreuses méthodes sont utilisées pour extraire les attributs, qui peuvent dépendre des caractéristiques géométriques, statistiques, des caractéristiques de texture et de couleur de l'objet en question. La figure 25 montre les caractéristiques et les méthodes d'extraction affiliées (Mutlag *et al.*, 2020).

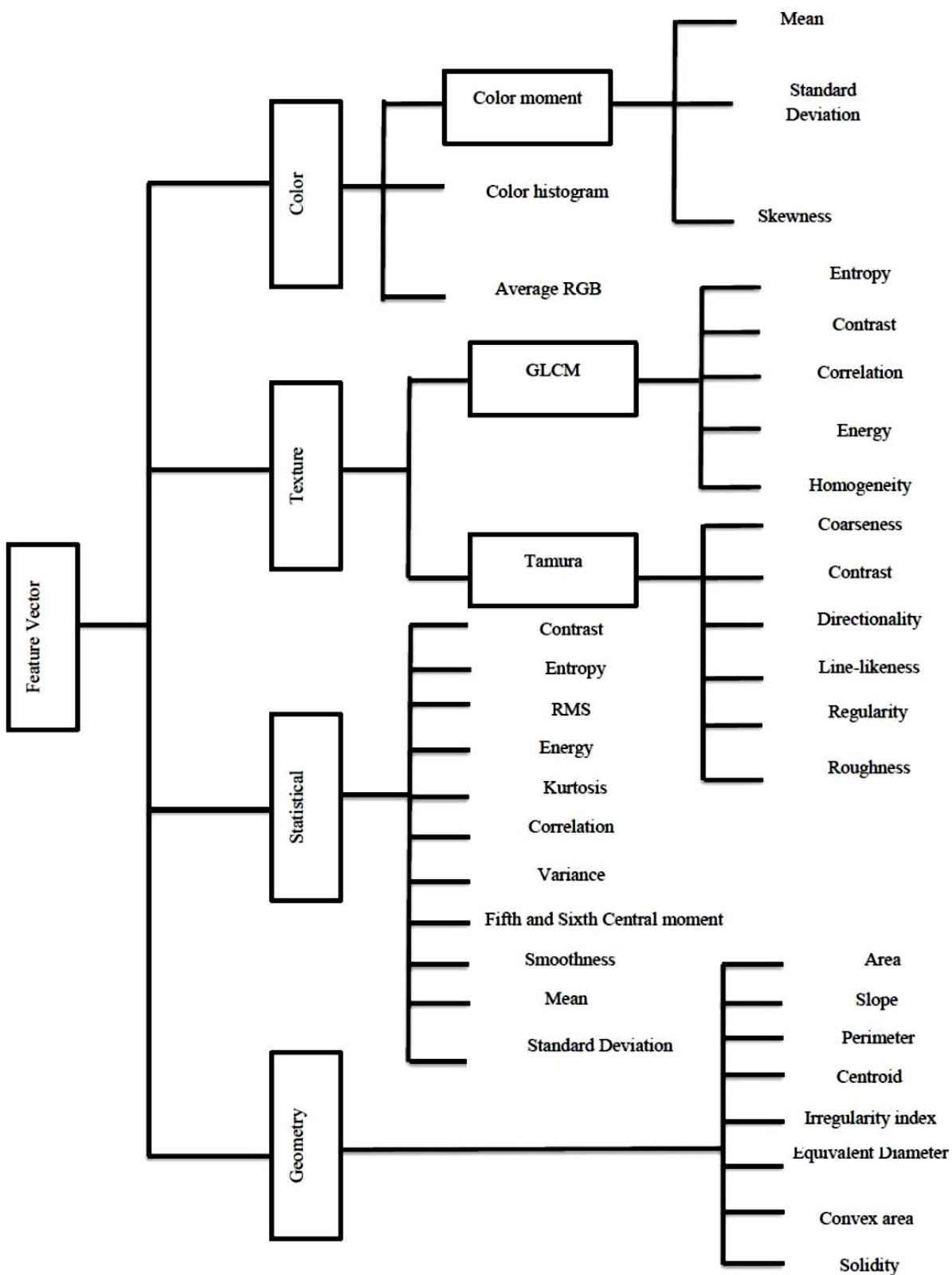


Figure 25: Méthodes d'extraction des Attributs (Mutlag *et al.*, 2020)

Les caractéristiques morphologiques telles que la rondeur, l'allongement, la compacité, etc., sont largement utilisées dans les opérations automatiques de classement, de tri, de détection et de contrôle de qualité des produits dans l'industrie alimentaire (Jayas, Paliwal and Visen, 2000).

Les mesures les plus courantes pouvant être effectuées sur les caractéristiques des images pour la classification des grains de céréales peuvent être regroupées en trois classes : morphologie, couleur et texture (Zapotoczny, Zielinska and Nita, 2008)

Selon Du and Sun (2004) trois caractéristiques couramment utilisées pour la mesure de la taille d'un objet peuvent être trouvées pour l'évaluation de la qualité des aliments : la surface, le périmètre, la longueur et la largeur. La mesure la plus élémentaire pour la taille est la superficie. Le périmètre d'un objet est particulièrement utile pour faire la distinction entre des objets aux formes simples et complexes. Les mesures de surface et de périmètre sont facilement calculées lors de l'extraction d'un objet à partir d'une image segmentée. La longueur et la largeur d'un objet peuvent également être utilisées pour mesurer la taille d'un objet (Du and Sun, 2004).

De plus, la forme est l'une des mesures d'objet les plus courantes pour l'évaluation de la qualité des aliments. Par rapport à d'autres caractéristiques, telles que la couleur et la texture, la forme est plus facile à mesurer à l'aide de techniques de traitement d'image. Les caractéristiques de forme peuvent être mesurées indépendamment et également en combinant des mesures de taille (Zapotoczny, Zielinska and Nita, 2008).

IV.3. Vision de haut niveau (reconnaissance d'objet)

Les processus de haut niveau impliquent la reconnaissance et la classification des régions d'intérêt qui sont généralement effectuées par des classificateurs statistiques ou des réseaux de neurones.

La classification en effet répond à des questions comme, *est ce qu'il y a l'objet A dans l'image ?* Une tâche légèrement complexe serait la détection parce qu'elle doit répondre à la question de savoir *où se trouve l'objet A dans l'image*. La reconnaissance visuelle implique la conception d'algorithmes capables de classer les images et les vidéos, de détecter et de localiser les objets, d'estimer les attributs sémantiques et géométriques, et de classer les activités humaines.

La vision de haut niveau implique un outil incontournable qui est le ML pour réaliser les tâches de la reconnaissance par CV.

V. Machine Learning

Le Machine Learning (figure 26) est un domaine de l'informatique issue de l'étude de la reconnaissance des formes et théorie de l'apprentissage informatique en intelligence artificielle. Les procédures fonctionnent par construction d'un modèle à partir d'exemples d'entrées afin de faire des prédictions ou des choix basés sur les données plutôt que de suivre un ensemble d'instructions de programme statiques (Selvam et Babu, 2015).

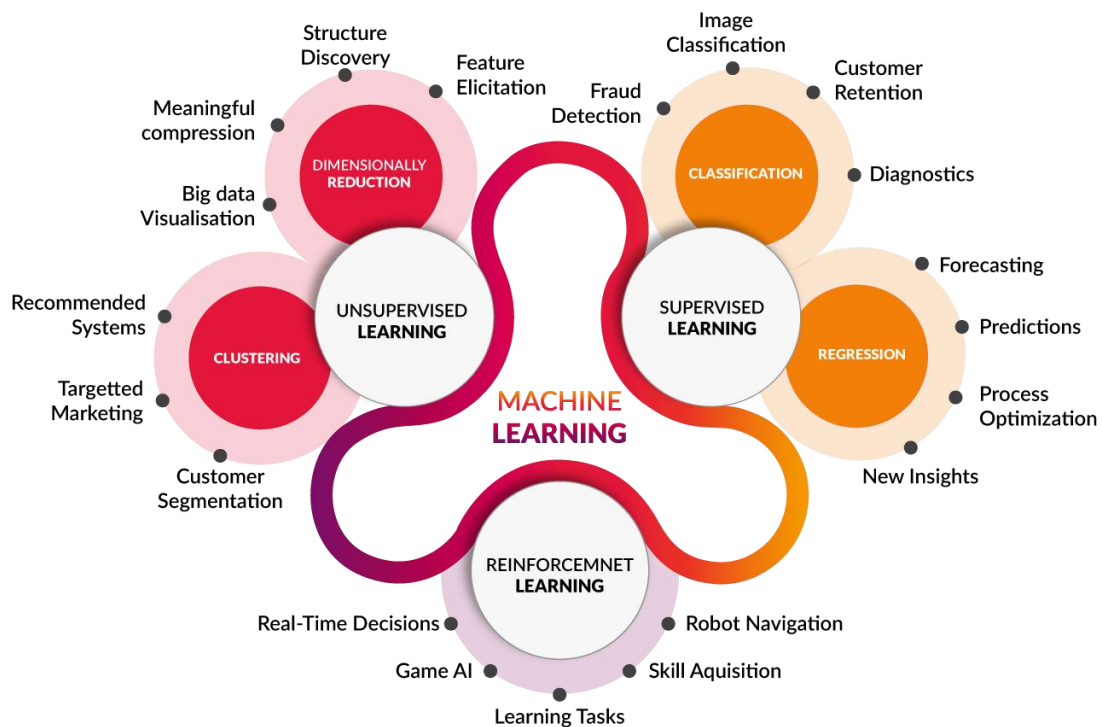


Figure 26: les domaines de Machine Learning (Krzyk, 2018)

V.1. Apprentissage non supervisé

Les algorithmes d'apprentissage non supervisé comparent les entrées les unes aux autres pour identifier les similitudes afin de les classer en catégories (Patrício et Rieder, 2018). Selon (Alom *et al.*, 2019) les systèmes d'apprentissage non supervisé sont ceux qui peuvent se passer de la présence d'étiquettes de données. Dans ce cas, l'agent apprend la représentation interne ou les caractéristiques importantes pour découvrir des relations ou une structure inconnue dans les données d'entrée. Souvent, le regroupement, la réduction de la dimensionnalité et les techniques génératives sont considérés comme des approches d'apprentissage non supervisé. Les principales méthodes d'apprentissage automatique non supervisé sont l'extraction des règles d'association, l'exploration séquentielle de motifs et le clustering (Tzanis *et al.*, 2006). Plusieurs modèles de la

famille de Deep Learning sont utilisés pour le regroupement et la réduction de la dimensionnalité non linéaire, notamment les auto-encodeurs (AE), les machines Boltzmann restreintes (RBM) .

V.2. Apprentissage supervisé

Selon Alom *et al.*(2019) l'apprentissage supervisé est une technique d'apprentissage qui utilise des données étiquetées. Dans le cas des approches DL supervisées, l'environnement a un ensemble d'entrées et de sorties correspondantes. L'apprentissage supervisé incarne donc le concept « *apprendre à partir d'exemples* ». Dans l'apprentissage supervisé, le système cognitif doit apprendre un concept ou une fonction qui est en fait la description d'un modèle (Tzanis *et al.*, 2006).

L'apprentissage supervisé est reconnu par les deux techniques suivantes : la classification et la régression. La classification concerne la construction de modèles de prédiction pour des fonctions à intervalle discret, tandis que la régression concerne la construction de modèles de prédiction pour des fonctions à intervalle continue.

Il existe différentes approches d'apprentissage supervisé pour l'apprentissage en profondeur, notamment les réseaux de neurones profonds (DNN), les réseaux de neurones convolutifs (CNN), les réseaux de neurones récurrents (RNN), y compris la mémoire à long court terme (LSTM) et les unités récurrentes fermées (GRU). Ces réseaux seront décrits en détail dans les sections respectives.

V.3. Réseaux de neurones artificiels (ANN)

Les réseaux de neurones artificiels (ANN) sont des algorithmes de calcul parallèles, qui consistent en de nombreuses unités de traitement liées. Ces unités de traitement ont une structure très simple et sont capables de convertir un signal d'entrée en un signal de sortie avec une fonction d'activation en correspondance avec les neurones biologiques. Le traitement parallèle d'un grand nombre d'entrées simultanément dans une structure hiérarchique permet de simuler plusieurs fonctions complexes à partir de données d'entrée.

Le terme réseaux de neurones artificiels (ANN) a été établi plus tôt que le terme d'intelligence artificielle (IA). Les premières tentatives d'utilisation des RNA ont pris la forme de modèles de neurones simplifiés et remontent au début des années 1940 (McCulloch & Pitts, 1943). Le concept d'apprentissage ANN (Bergenti, Giezes, & Zambonelli, 2004, Wiener, 1948) et plus précisément,

l'idée de développer différents types de systèmes informatiques avec une capacité d'intelligence comparable à celle des organismes biologiques commence vers 1950 (Von Neumann, 1958).

Le terme « artificiel » est utilisé pour indiquer l'effort de simulation de la structure des réseaux de neurones biologiques. Cette tendance est justifiée par la capacité particulière des réseaux de neurones biologiques à traiter de grandes quantités de données et à tirer des conclusions dans un laps de temps. Les avantages de l'application de ces réseaux de neurones biologiques les rendent idéaux pour les architectures de calcul prédictif et de traitement des données. Les caractéristiques particulières liées à la structure des réseaux de neurones sont leur capacité à apprendre, à stocker des informations, à généraliser à de nouvelles données et à regrouper des concepts. (Alom et al., 2019).

VI. La représentation machine

Selon Remagnino et al. (2017) la reconnaissance des espèces appartient au domaine de la taxonomie biologique et la « reconnaissance » doit être comprise selon deux modes distincts ; le premier sens consiste à discerner les unités et à tenter leur délimitation c'est ce qu'on appelle la « segmentation » en reconnaissance de formes informatisée, et la « classification » ou le « regroupement » en taxonomie biologique. Le second sens est d'identifier ou de "déterminer" un organisme comme appartenant à une espèce préexistante, c'est-à-dire la "classification" au sens des informaticiens. La figure 27 illustre selon (Wäldchen and Mäder, 2018) la description donnée d'une fleur par un Botaniste et sa représentation comparée à celle d'une machine.

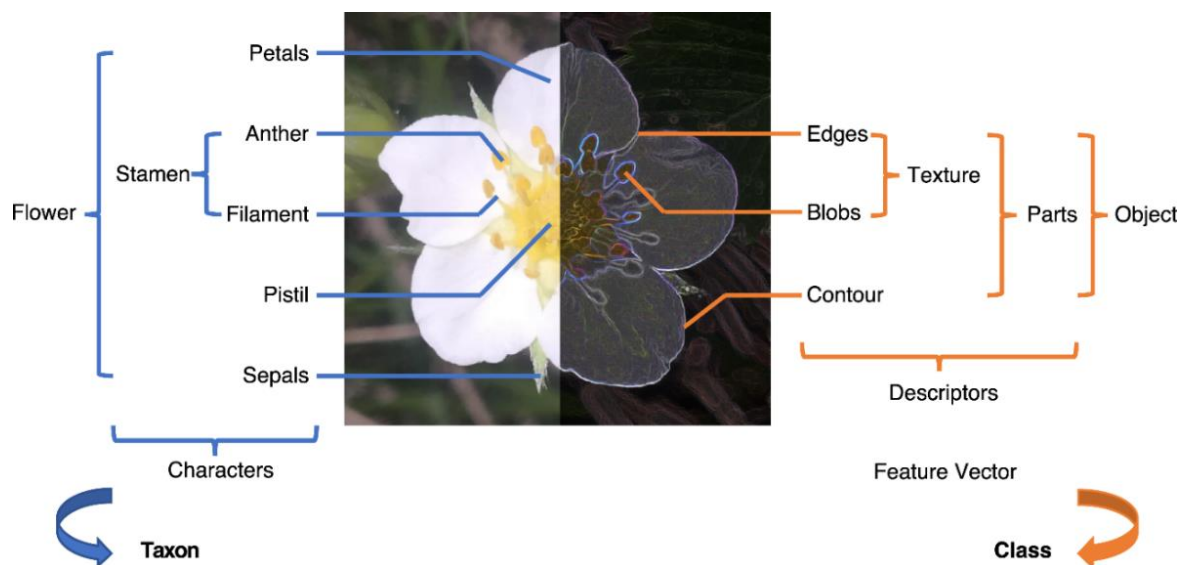


Figure 27: Description d'une fleur par un botanistes vs celle de la computer vision

La figure 28 illustre selon (Sladojevic *et al.*, 2016) la caractérisation d'une feuille de vigne dans des couches d'un réseau de neurones Deep Learning.

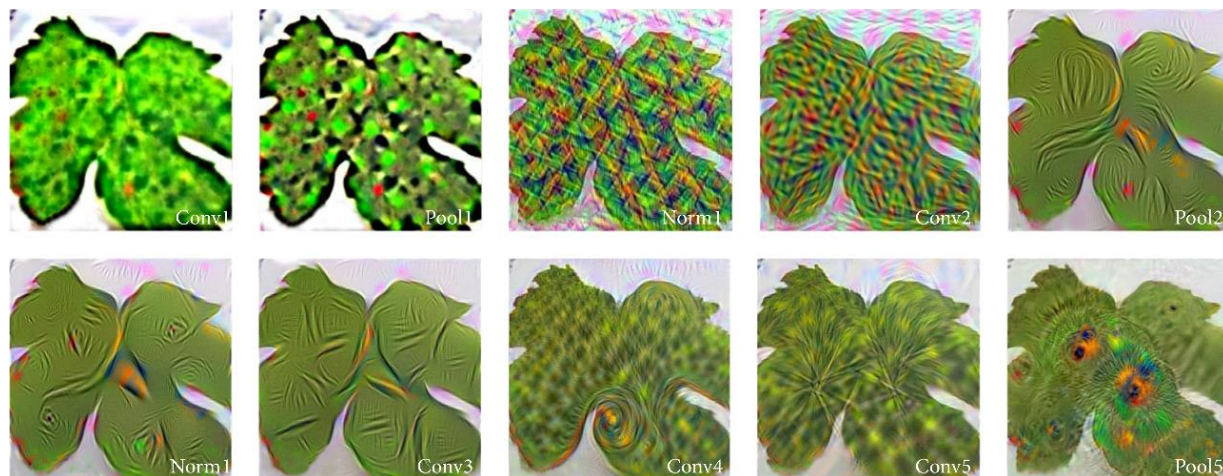


Figure 28: Les attributs et la formation des nouvelles images en Deep Learning (Sladojevic *et al.*, 2016)

VII. Travaux en relation avec l'étude

Les solutions CV et ML (E. R. Davies, 2012; Makridakis, Spiliotis and Assimakopoulos, 2018) sont mises en place dans les pratiques agricoles et l'agro-industrie (Dutta Gupta and Ibaraki, 2014; Vithu and Moses, 2016) ce qui a donné et donne une impulsion à la recherche agronomique avancée sur céréales et les autres cultures (Davies, 2009; Pridmore, 2015; Patrício and Rieder, 2018; Tripodi *et al.*, 2018). Le ML est une approche intrinsèquement multidisciplinaire de l'analyse de données qui s'inspire et emprunte fortement à la théorie des probabilités, aux statistiques, à la théorie de la décision, à la visualisation et à l'optimisation (Singh *et al.*, 2016).

Wäldchen and Mäder,(2018) ont réalisé une revue systématique de la littérature, dans le but d'une analyse approfondie et d'une comparaison des études primaires sur les approches CV pour l'identification des espèces végétales. La classification d'images est l'une des techniques de ML les plus utilisées permettant d'améliorer les opérations agricoles et la transformation agro-industrielle des aliments pour les tâches d'identification et de reconnaissance d'objets (Du and Sun, 2006; Patrício and Rieder, 2018).

En référence aux définitions de (E. R. Davies, 2012) et à la réflexion conceptuelle des systèmes ML (LeCun, Bengio and Hinton, 2015), deux approches de ML sont principalement utilisées pour

la classification d'images : l'approche traditionnelle ou les modèles Shallow classiques (Brahimi, Boukhalifa and Moussaoui, 2017; Alom *et al.*, 2019) et les modèles Deep Learning récemment ré-émérgés .

Une différence clé entre le ML traditionnel et le DL réside dans la manière dont les attributs sont extraits. Les approches ML traditionnelles utilisent des fonctionnalités d'ingénierie artisanales en appliquant plusieurs algorithmes d'extraction d'attributs, puis on applique les algorithmes d'apprentissage.

De plus, d'autres approches de renforcement sont souvent utilisées lorsque plusieurs algorithmes d'apprentissage sont appliqués aux caractéristiques d'une seule tâche ou d'un ensemble de données et qu'une décision est prise en fonction des multiples résultats des différents algorithmes.

D'autre part, dans le cas de DL, les attributs sont appris automatiquement et sont représentées hiérarchiquement à plusieurs niveaux. C'est le point fort de DL par rapport aux approches traditionnelles d'apprentissage automatique.

Comme illustré dans la figure 29, la mise en œuvre des CV et ML pour les espèces végétales de blé dépendait de la considération suivante (scénario), soit le végétale entier, soit un organe spécifique comme dans le cas de notre étude.

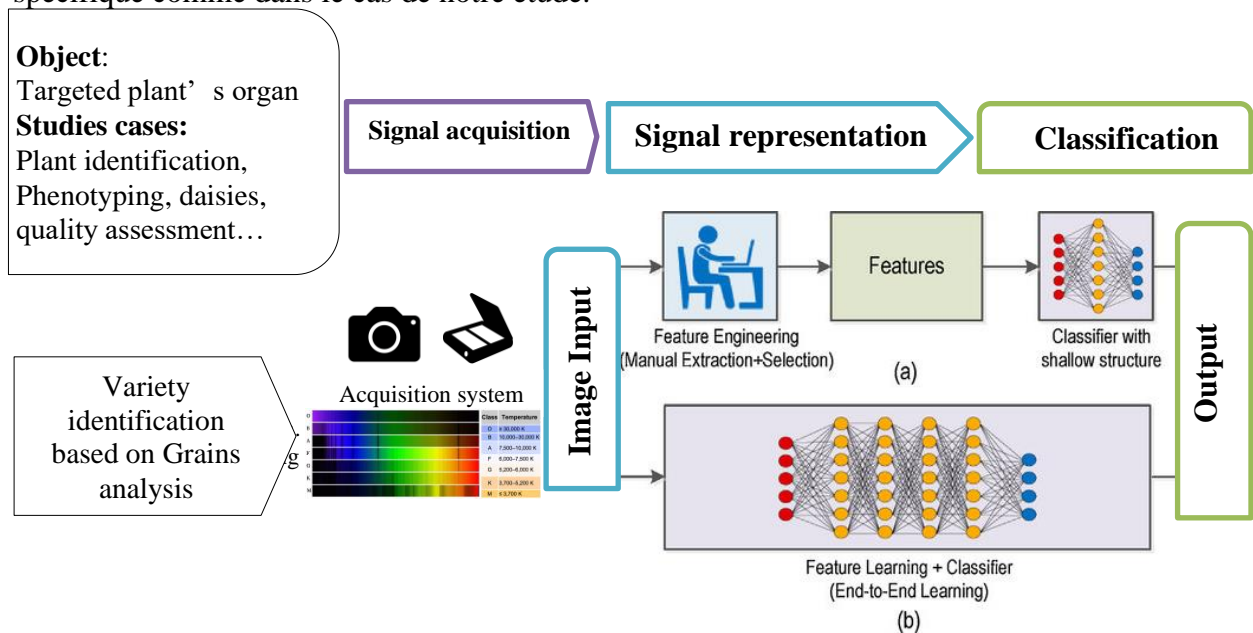


Figure 29: Shallow et Deep Learning approches et l'applications pour l'identification des plantes.

Rahaman *et al.*,(2015) and Barmeier and Schmidhalter,(2016) ont travaillé sur le phénotypage à haut débit et l'analyse scientifique des données phénotypiques des cultures de blé (figure 30).

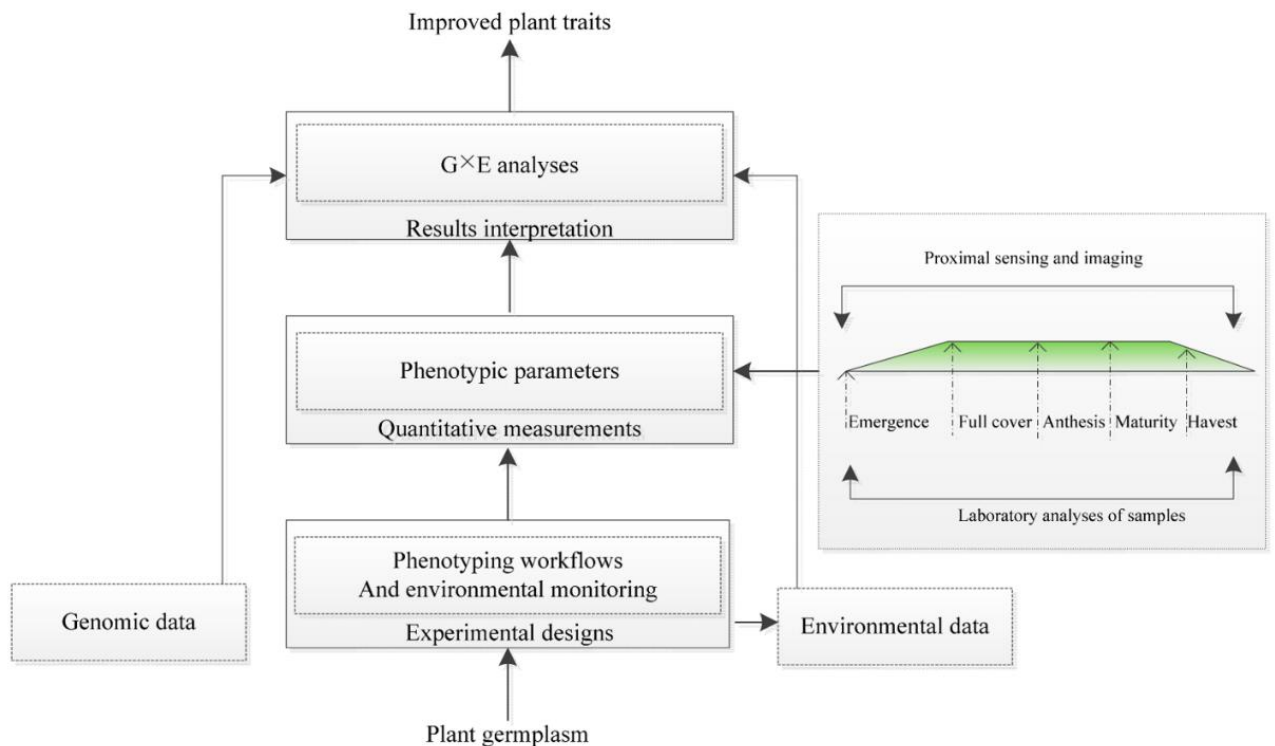


Figure 30: Exemple d'un schéma de phénotypage des plantes (Li, Zhang and Huang, 2014)

Des chercheurs comme (Francl and Panigrahi, 1997; Bauer, Korč and Förstner, 2011; Serranti, Cesare and Bonifazi, 2013) ont étudié les capacités du Machine Learning pour le monitoring de la santé des plantes et la protection des cultures, Brahim et al., 2019 ont effectué un travail de recherche pour l'identification des maladies des plantes en se basant sur le Deep Learning.

Dans ce qui suit nous allons traiter les applications des deux approches de classification pour les deux niveaux SLC et VLC.

VII.1. Shallow approches pour la classification des grains de blé

Patrício & Rieder (2018) et Vithu & Moses (2016) ont effectué une synthèse sur les approches Shallow exploitées dans plusieurs domaines liés à l'agriculture et à l'alimentation. Le défi du GIR par classification automatique a été investigué par plusieurs chercheurs depuis les années 80' notamment avec les travaux de Zayas et al. (1985), Neuman et al. (1987), Symon et al. (1988) et Zayas et al. (1989), ces auteurs ont défini les attributs couramment extraits des images numériques de grains de blé ; Majumdar et Jayas (2000) ont reconsidéré pour leur expérimentation de

classification d'espèces de blé (SLC), trois groupes de caractéristiques : caractéristiques morphologiques (MF) comprenant la taille, la forme et des facteurs de forme, les caractéristiques de couleur (CF) et les caractéristiques de texture (TF). Guevara-Hernandez and Gomez-Gil, (2011) ont utilisé l'analyse discriminante (DA) et des algorithmes des K-plus proches voisins (K-NN) pour effectuer un SLC (objet blé et non-blé). Les deux auteurs ont atteint une précision de 99% lors de l'utilisation de fonctionnalités combinées (MF.CF.TF.). Une analyse discriminante des moindres carrés partiels a été appliquée par (Vermeulen *et al.*, 2018) pour un SLC entre les grains de blé dur et de blé tendre. Les résultats ont montré que la classification avec des données mixtes de (MF) et spectrales NIR peut archiver une précision de 99%.

D'autres caractéristiques ont été explorées, tel que dans les travaux de Williams, Munkvold et Sorrells (2013), qui visent à phénotyper la forme du grain chez le blé hexaploïde (*Triticum aestivum* L.), ils ont utilisé les descripteurs *elliptiques de Fourier* (Godefroy *et al.*, 2012) pour capturer les variations de forme telles que la rugosité, l'inclinaison asymétrique et d'autres aspects bidimensionnels non englobés par des axes ou des distinctions dans la zone globale de l'objet.

Les caractéristiques invariantes à l'échelle dense (DSIFT) sont aussi un autre type de caractéristiques qui a été utilisé par Olgun *et al.* (2016) pour la classification du blé. Dans ce travail, des caractéristiques DSIFT ont été extraites d'images couleur d'échantillons de grains de blé en vrac. Les auteurs ont utilisé un classificateur de machine à vecteurs de support (SVM).

(Choudhary, Paliwal and Jayas, 2008) ont mené une SLC avec des données mixtes en utilisant comme caractéristiques : longueur d'ondes, MF, CF et TF. La combinaison de toutes les caractéristiques a donné la meilleure classification en utilisant le classificateur discriminant linéaire avec une précision de classification de 99,4 % pour le blé CWRS, suivi de 99,3 %, 98,6 %, 98,5 % et 89,4 % pour le seigle, l'orge, l'avoine et le blé CWAD.

Concernant le cas de VLC, le défi est plus difficile en raison de la haute similitude entre les classes. (Zapotoczny, 2011a, 2011b, 2014) ont mené plusieurs recherches sur la discrimination des grains de blé de 11 variétés couvrant trois années de culture successives et d'une région de culture, sur les deux études de l'année 2011, ils ont effectué une VLC à grain unique, la base de données compte 552 grains pour chaque variété. Dans la première étude, 20 TF ont été sélectionnés comme entrée pour un classificateur de réseaux neuronaux artificiels (ANN), la précision avait atteint 100 %. Dans la deuxième étude, l'auteur a utilisé 54 variables géométriques MF, réduites à 20 à l'aide des méthodes *Class Ranker* et *Class Rankers-Search*, après avoir utilisé une analyse progressive

par étapes et un classificateur *Meta MultiClass*. Le modèle statistique proposé classait les variétés à 90-100 %.

Zapotoczny (2014) a effectué une autre expérience VLC sur des échantillons en vrac, en utilisant 49 variables TF et une analyse multidimensionnelle, le classificateur a archivé une précision comprise entre 98 et 100 %. Dans les études précédentes, l'auteur Zapotoczny a utilisé des méthodes destructives dans des conditions contrôlées pour la préparation et le traitement des échantillons de grains ainsi que pour les étapes d'acquisition d'images. L'ensemble de données peut ne pas refléter les aspects de la vérité terrain et la diversité.

Cependant, l'approche n'est pas adaptée pour une classification à un seul grain, car l'analyse de pureté doit déterminer l'identité de chaque grain considéré au test. Le même problème avec l'étude de (Choudhary *et al.*, 2009), les auteurs ont effectué une VLC de huit classes de blé en utilisant des caractéristiques d'ondes à partir d'images hyperspectrales proche infrarouge d'échantillons en vrac pour un système d'acquisition sophistiqué. La précision de classification moyenne la plus élevée de huit classes était de 99,1% lorsque les 90 principales caractéristiques d'ondes ont été utilisées pour la classification dans un classificateur discriminant linéaire.

Dans une autre étude, Olgun *et al.* (2016), ont utilisé un classificateur de machine à vecteur de support (SVM) avec des caractéristiques invariantes à l'échelle dense (DSIFT) comme données d'entrée extraites d'un ensemble d'images de grains en vrac. Le classificateur a atteint un taux de précision global de 88,33 %.

Les approches de classification d'images des réseaux de neurones artificiels (ANN) ont été explorées dans plusieurs travaux de recherche. Khoshroo and Arefi, (2014) ont procédé à un VLC utilisant neuf MF pour évaluer les performances des réseaux. En utilisant un petit ensemble de données, les résultats ont montré une précision de classification globale de 85,72 %. (Paliwal *et al.*, 2003) effectuent un SLC pour l'identification des grains de céréales et des impuretés en utilisant 230 caractéristiques combinées (MF.CF.TF.), les précisions étaient d'environ 90 %.

L'utilisation de l'approche de classification multi-classificateur était une tentative d'amélioration de la précision, Ebrahimi *et al.* (2014) ont combiné deux classificateurs, Imperialist Competitive Algorithm (ICA) et Artificial Neural Networks (ANN). Le taux de classification total de cette approche pour les SLC (grains de blé vs graines autres que le blé) était de 96,25 %.

Un autre travail utilisant cette technique de multi-classificateur a été mené par (Visen *et al.*, 2002). Les auteurs indiquent dans ce travail et sur la base des recherches antérieures de (Paliwal, Visen and Jayas, 2001) que l'utilisation d'un réseau de neurones peut ne pas aboutir à une précision de classification élevée si les objets à classer présentent des différences très subtiles de morphologie, de couleur et caractéristiques de texture. Pour surmonter ce problème, on peut ajouter une combinaison de plusieurs réseaux de neurones qui se « spécialisent » dans l'identification d'un type particulier d'objet, Visen and Jayas (2001) ont obtenu, pour un SLC, une précision comprise entre 96 et 99 %.

Les performances et la fiabilité du modèle shallow ne sont pas seulement corrélées à la technique d'extraction des caractéristiques et à la force discriminante des caractéristiques extraites, mais également à la qualité de l'ensemble de données, y compris la diversité, la taille et la proportion de chaque classe. La vérité terrain doit être reflétée comme la condition de croissance et l'effet agronomique, par ex. techniques de culture. Un grand ensemble de données à besoin d'un classificateur et d'un cadre (framework) adaptés, ce qui impliquait l'utilisation d'une machine plus puissante et de nouvelles approches informatiques basées sur une unité de traitement graphique (GPU) pour CV et ML. Une CV et un ML tout-en-un (all in one) étaient devenus possibles avec l'exploration du Deep Learning (DL) et du GPU qui permet le parallèle lors de l'enseignement de grands ensembles de données d'images.

VII.2. The Deep Learning approaches

Le Deep Learning (DL) est une classe des techniques de la technologie Machine Learning, composée de plusieurs couches de traitement qui permettent l'apprentissage de la représentation de l'abstraction de données à plusieurs niveaux (Lee *et al.*, 2017). Il est aussi un ensemble de méthodes universelles (LeCun, Bengio and Hinton, 2015) qui ont été récemment appliquées avec succès et largement dans de nombreux domaines et tâches (Alom *et al.*, 2019). L'essentiel de DL est sa capacité à créer et à extrapoler de nouvelles attributs à partir de représentations brutes de données d'entrée sans avoir à dire explicitement quels attributs utilisés et comment les extraire (Lee *et al.*, 2017). Les réseaux de neurones convolutifs (CNN) sont l'une des meilleures architectures DL utilisées pour reconnaître, détecter et récupérer le contenu d'une image.

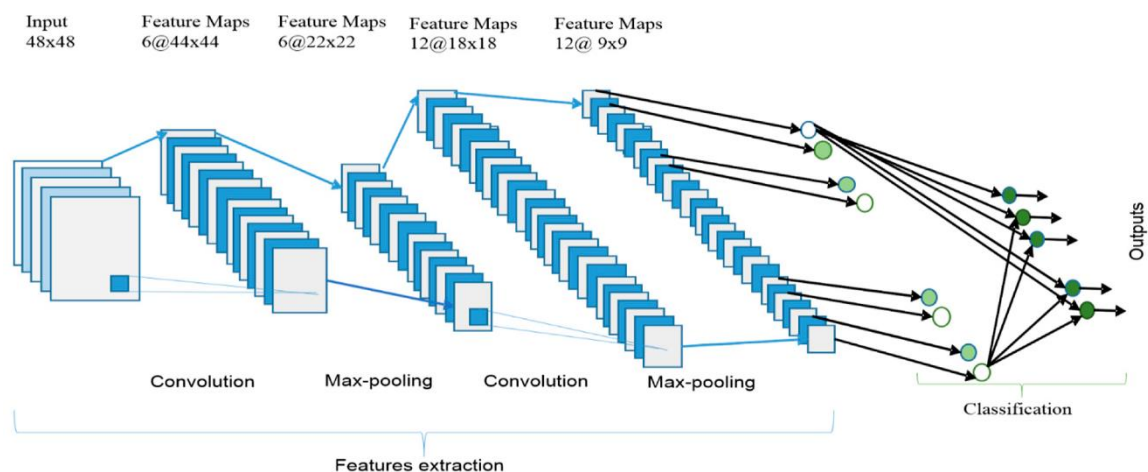
Kamilaris and Prenafeta-Boldú, (2018) ont présenté une enquête récente sur les techniques d'apprentissage Deep Learning à l'agriculture avec un examen détaillé de 40 articles de recherche pertinents.

Selon Alom et al., (2019) l'architecture globale d'un réseau neuronal convolutif (CNN) comprend (Figure 31) une couche d'entrée (input layer), plusieurs couches alternées de convolution et de mise en commun maximale (max pooling), une couche entièrement connectée (fully connected layer) et une couche de classification.

Cependant, plusieurs auteurs ont utilisé des modèles Deep Learning pour la classification des plantes. L'état de l'art montre deux tendances. La première est liée au phénotypage à haut débit et à l'identification des plantes comme l'étude de (Lee *et al.*, 2017) et (Ubbens and Stavness, 2017). La seconde tendance est l'étude de surveillance de la santé des plantes et de détection des maladies (Ferentinos, 2018).

Thenmozhi and Srinivasulu Reddy (2019) ont proposé un modèle CNN efficace pour classer les espèces d'insectes dans des grains de blé à l'aide de trois ensembles de données publiques sur les insectes. Le modèle proposé a été évalué et comparé aux architectures optimisées d'AlexNet, ResNet et GoogLeNet, la précision de classification est de 96,75 %, 97,47 % et 95,97 % respectivement aux architectures citées.

En se référant au point de vue de l'apprentissage par transfert (transfer learning), les modèles entraînés sur un grand ensemble de données dans un processus d'apprentissage par transfert sont



capables de classer avec succès des images dans un nouveau domaine (LeCun, Bengio and Hinton,

Figure 31: L'architecture globale du réseau neuronal convolutif (CNN). (Alom et al., 2019)

2015).

Too et al. (2019) ont proposé dans une récente étude d'identification de plantes, une étude comparative des méthodes de réglage fin pour un modèle d'apprentissage en profondeur. Le modèle DenseNet avait atteint une précision de test de classification de 99,75 %.

Uzal et al. (2018) ont introduit une méthode de CV qui prédit le nombre de grains dans les gousses de soja. À cette fin, les auteurs ont utilisé un CNN standard avec un total de 12 couches convolutives, plus une couche softmax de sortie, définissant une profondeur de 13 couches. Leur CNN a été comparé à un modèle de classification SVM. La méthode SVM atteint une précision de 90,2 % pour la validation et de 50,4 % pour le test. La méthode CNN sans *augmentation des données* produit une précision de 93,6% pour la validation et de 82,7% pour le test.

Concernant les grains de blé, Shen *et al.*, (2018) ont testé la possibilité de détecter des insectes dans les grains stockés en utilisant le Deep Learning. Ils ont utilisé un Fast R-CNN pour extraire les zones susceptibles de contenir les insectes dans les images et classer les insectes dans ces zones. Un réseau Inception amélioré a été développé pour extraire des cartes de caractéristiques. Les résultats ont montré que la méthode développée pouvait détecter et identifier les insectes dans des conditions de stockage de grain, et la moyenne de précision moyenne (MAP *Mean Averg Precision*) atteinte était de 88 %.

VII.3. Identification variétale par le Deep Learning

Un récent travail de classification de niveau variétale en utilisant des CNN a été réalisé sur des variétés d'orge par (Kozłowski, Górecki and Szczypiński, 2019). Les auteurs ont utilisé neuf implémentations CNN : cinq qui ont été personnalisées et une architecture CNN entièrement construite de 9 couches, et avec les deux principaux modes d'apprentissage par transfert (réglage fin et extracteur d'attribut) à l'aide de deux architectures CNN, AlexNet et ResNet18. L'auteur a obtenu une précision supérieure à 93% pour son CNN entièrement construit (64 filtres de taille 3x3) et son modèle ResNet18 affiné.

L'auteur avait utilisé un ensemble de données équilibré de 60 753 images à grain unique représentant six variétés. Les auteurs ont effectué une étape de pré-traitement suivant l'approche présentée dans (Szczypiński and Zapotoczny, 2012). Les orientations dorso-ventrale et antéropostérieure du grain étaient aléatoires. Des opérations intéressantes ont été réalisées : la correction de l'orientation antéropostérieure (même direction à tout grain), la suppression de l'arrière-plan et le redimensionnement de l'image tout grain à la même taille 80 x 170 pixels.

Cela pourrait pousser le CNN à concentrer l'apprentissage sur la région des pixels du grain, mais l'utilisation du même appareil d'acquisition d'image pourrait également être un problème *d'attributs discriminatoires* qui peut affecter la qualité de l'apprentissage.

L'opération de suppression de l'arrière-plan peut améliorer la précision mais en question de segmentation les informations morphologiques pourraient être affectées. En outre, la correction de l'orientation antéropostérieure dans une direction peut affecter la classification, la rotation de l'image est l'une des techniques d'augmentation des données pouvant améliorer la précision de la (Shorten and Khoshgoftaar, 2019).

De plus, le problème de l'orientation dorso-ventrale a été étudié par (Dolata and Reiner, 2018). Les auteurs ont supposé que la nature du processus d'imagerie est imprévisible et que chaque côté du grain a un impact sur la classification, la même affirmation a été confirmée tôt par (Szczyński, Klepaczko and Zapotoczny, 2015). L'étude a montré que l'analyse du côté ridées du grain de l'orge a un impact significatif sur les résultats de la classification des grains. Outre le fait que le blé et l'orge sont deux genres différents, la face ventrale du grain de blé se caractérise par la présence d'un pli (sillon ventral) et d'autres caractéristiques qui sont plus diverses par rapport à la face dorsale comme les poils de la brosse, la pointe de la radicule (Wrigley, 2017b).

Le classificateur CNN exécute les mêmes tâches pour les deux niveaux de classification la SLC et la VLC. Les performances sont liées à l'architecture choisie et aux hyperparamètres des CNN et à la qualité des ensembles de données d'images, y compris la taille et la diversité.

La vision humaine est une vision tridimensionnelle, il pourrait être intéressant d'utiliser des images 3D ou un système d'acquisition plus sophistiqué pour améliorer la précision, mais une technique idéale à faible coût doit être conçue. Le scanner à plat a été largement utilisé par plusieurs auteurs car il offre une facilité d'utilisation sans un étalonnage complexe et une stabilité d'image (aucune mise au point n'est nécessaire). L'utilisation de plusieurs modèles d'appareils peut augmenter la diversité des jeux de données d'images.

VIII. Conclusion :

Nous avons cherché dans ce chapitre à construire et à traiter le concept de l'identité biodigitale du grain du blé. en passant en revue les grandes lignes, les définitions et les approches de l'implémentation de la CV et le ML ; en général il s'agit d'un mimétisme de la perception humaine par une perception machine grâce à la technique de classification d'images ; cette dernière est exécutée pendant chaque opération de reconnaissance d'objet en particulier dans notre cas l'opération de la reconnaissance de l'identité des grains de blé (GIR) à un niveau variétale (VLC) ; la classification VLC permet d'atteindre divers finalités tel que le test de pureté qu'on avait traité dans le chapitre 2 .

La GIR suit les deux approches de classification d'images Shallows et Deep learning ; nous avons passé en revue les méthodologies suivies par les divers auteurs qui ont réalisé des travaux similaires ou en relation ; de cette revue nous avons consolidé la partie du background théorique de notre démarche pour réaliser la VLC. L'analyse des différents travaux scientifiques en relation direct et indirect avec notre recherche nous a permis de repérer les enjeux et les limites de chaque approche de classification.

CONCLUSION de la partie bibliographie :

Le background théorique est indispensable pour la construction d'une plateforme de connaissances nécessaire à la conception et/ou à l'intégration de nouvelles technologies dans un domaine qui n'était pas nécessairement à l'origine de leurs existences ; l'objectif opérationnel de cette partie bibliographique est de regrouper ces connaissances et les mettre en interaction de manière explicite pour un objectif d'éclaircissement et d'accompagnement de la résolution du problème posé par la reconnaissance de l'identité des grains de blé .

La plateforme de connaissances regroupe les définitions et les notions pour la compréhension et l'analyse du caractère biologique, en particulier : les moments clés de la formation et la différenciation des grains, l'explication de l'origine de la formation et la nature de chaque caractère (attribut), les contraintes qui peuvent affecter positivement ou négativement ce processus de formation et enfin la dotation du grain d'une identité biologique définissable en taxon.

Notre vision a été élargie par l'analyse et la redéfinition de l'opération de l'identification du grain dans le contexte de la chaîne des grains et précisément dans le cadre de la démarche qualité, ici nous avons défini la qualité des grains par deux composantes principales à savoir *l'identité* et *l'état des grains*.

En suite la solution technologique basée sur la CV et ML a été traitée selon le même principe ; on a revu les grandes lignes théoriques et pratiques en relations avec les démarches de classification d'images (shallow et Deep Learning) de manière méthodologique et on a retenu les principales modalités pour atteindre l'objectif de notre travail de recherche.

A la lumière de toutes les informations et contraintes énoncées, une nouvelle approche ergonomique est suggérée pour la GIR en utilisant le processus de classification d'image par l'intégration de la solution CV et ML.

Ainsi, l'approche devrait avoir les critères suivants : une perception artificielle qui peut se substituer à la perception humaine, la préservation des avantages pratiques de la VGA en utilisant l'image de la face ventrale du grain et ensuite être capable d'effectuer une VLC avec précision. Les approches automatiques pourraient améliorer considérablement les conditions de travail des tests de pureté, ainsi que d'autres avantages liés à sa mise en œuvre, tel qu'un cadre d'analyse d'en réseau.

Partie Bibliographique

Conclusion de la partie bibliographique

Au niveau de connaissances actuelles un GIR des grains de blé par VLC basée sur la mise en œuvre de la classification d'images n'a pas encore été étudiée profondément. De plus, une VLC ou une SLC utilisant plusieurs modèles de scanners à plat et dans des conditions de terrain réel, en tant que simulation du réseau et du cadre des testeurs, n'a pas encore été étudiée.

PARTIE EXPERIMENTALE

Introduction Expérimentale

De l'analyse de la revue bibliographique récente nous avons constaté l'existence d'au moins deux approches principales de classification d'images pour l'identification des grains, un choix et une continuité Chronologique, en fonctions des objectifs fixés et l'application du progrès technique. les deux approches de la classification d'image sont l'approche « *Shallow Classification* » (Yin *et al.*, 2014) et l'approche « *Deep Learning Classification* » (Laabassi *et al.*, 2021). Durant ce travail de recherche, nous avons réalisé deux principales expériences en appliquant les deux approches de la classification. Dans chaque approche nous avons investigué la faisabilité des deux niveaux de classification à savoir le niveau espèce SLC et le niveau variété VLC.

La figure 32 montre le diagramme du flux de travail de nos expérimentations. En amont de ce flux de travail, nous avons placé la collecte des données (*Data collection*), cette opération est considérée comme une étape commune des deux approches car nous avons utilisé les mêmes échantillons et donc la même procédure de collecte. Cette étape inclue deux principales actions : La collecte et la préparation des échantillons des grains de blé (*Grain collection*), La collecte des images de ces grains (*Images dataset collection*).

Ensuite dans la conduite de la première expérimentation « *Shallow Classification* », une séquence de taches où sont exécutées la préparation, l'analyse par *Computer Vision* (extraction des caractéristiques) et le Machine Learning. Enfin, se connecte la conduite de la deuxième expérimentation « *Deep Learning Classification* », en exploitant les images des grains

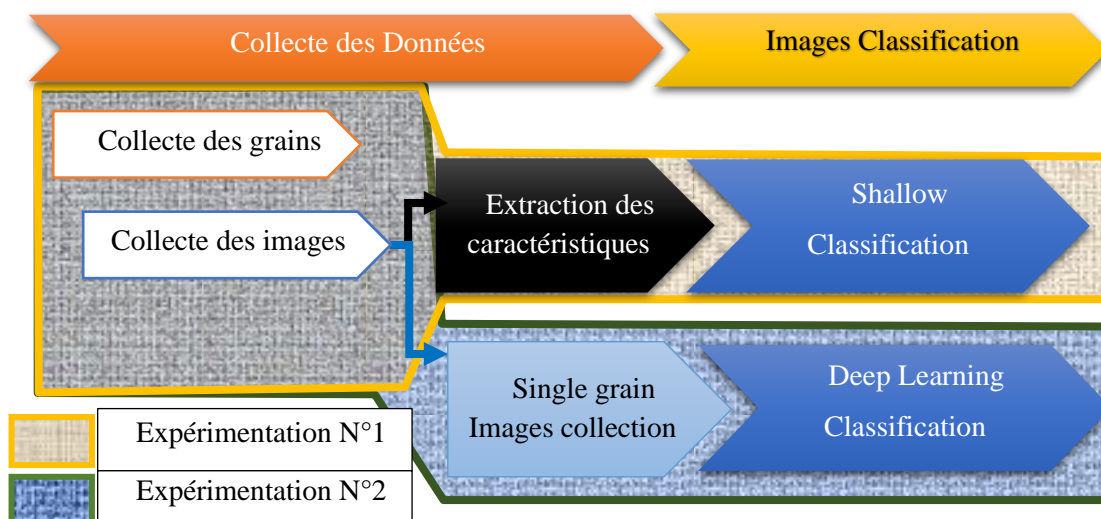


Figure 32: Flux du travail d'expérimentation

individuels , il s'agit de la séquence de tâches spécifiques à l'implémentation du Deep Learning en utilisant la technique de transfert de savoir « **Transfert Learning** » (Laabassi *et al.*, 2021). Nous allons rapporter simultanément le matériel et la démarche suivie ainsi que les résultats pour chaque expérimentation. Et d'autres spécifications et informations que chaque approche peut fournir (une comparaison entre les résultats).

Cette partie comprend 03 chapitres :

Chapitre 1 : La collecte des données

Ce chapitre est consacré à la phase de collecte des données et la construction de la base d'apprentissage, la démarche et la stratégie suivie. Ce chapitre décrit en finalité l'approche de collecte des données à suivre dans un système d'information sur la qualité des grains.

Chapitre 2 : l'expérimentation N°1 Shallow approche

Ce chapitre est l'implémentation de la Computer Vision classique basée sur l'extractions et la classification par caractéristiques morphologiques. Dans ce chapitre nous avons eu des résultats et des conclusions sur la faisabilité et les limites de cette technique qui nous ont conduit à exploiter une autre méthode basée sur le Deep Learning.

Chapitre 3 : l'expérimentation N° 2 Deep Learning approche :

Ce chapitre est l'implémentation de la classification d'images par des CNNs, cette technique nous a donné des résultats confirmant la faisabilité de l'identification des grains à l'échelle variétale.

Le travail de recherché propose :

Une approche et des modalités pour l'utilisation de la vision artificielle pour une classification de niveau variétale ; ceci permettra de réaliser l'identification des variétés de blé à base de leurs grains et par conséquent elle va ouvrir des horizons pour l'utilisation de cette solution à tous les niveaux de la chaine de grains (contrôle, inspection, certification, phénotypage...).

Chapitre 1 : la collecte des données « the Data collection »

I. Introduction

La collecte des données d'apprentissage est une opération primordiale dans la construction du modèle de classification machine (Coppens *et al.*, 2017). À rappeler que la reconnaissance automatique des grains de blés est réalisée grâce à la technique de classification d'images. L'importance de la collecte des données est aussi inscrite en vue de perfectionner en continue les modèles d'apprentissage élaborés. Ceci en introduisant au fur et à mesure de nouvelles données, les mises à jour peuvent contenir de nouvelles informations sur l'interactions plant-environnement ou de nouvelles caractéristiques permettant de mieux reconnaître automatiquement une variété de blé en assurant une bonne qualité d'apprentissage des modèles automatiques.

Le travail qui doit se faire est d'implémenter progressivement un système d'informations sur la base des données de bonnes qualités et diversifiées.

Afin de mieux explorer les concepts cités au profit de notre investigation, nous pouvons spécifier ces types de données, à savoir les :

- **Données biologiques** : représentent les échantillons des grains et leurs labels, en incluant les informations relatives au matériel végétal tel que le phénotype.
- **Données images** : représentent les images capturées ainsi que leurs métadonnées et labels
- **Données numériques** : concernent principalement les valeurs des caractéristiques extraites de l'abstraction des grains sur images, ou les résultats de l'analyse d'images (computer vision).

Ensuite, et pour la collecte, on s'est basé sur une stratégie de diversification des données d'apprentissage (Laabassi *et al.*, 2021), étudiée récemment par (Gong, Zhong and Hu, 2019) . Le but de cette stratégie est de créer une base de données représentative, qui reflète la réalité du terrain, ça permettra de réaliser un apprentissage robuste.

A cette finalité notre stratégie consiste à multiplier les sources de variabilité d'une part, et d'autre part c'est de réaliser l'expérimentation sous des conditions proches de celles imposées dans un travail réel et ambiant (**Under field conditions, grounds Truth**). Nous avons donc

considéré pour cela des sources de variations en relation avec le matériel végétal et en relation aussi avec le système d'acquisition, ceci comme suit :

1. **La variation temporelle** : les années de récolte.
2. **La variabilité spatiale** : les lieux de collecte. En collectant les échantillons de grains de plusieurs et différentes zones de cultivation.
3. **La variabilité agronomique** : facteurs agro-techniques de production. Elle provient du système de production (itinéraire technique) pratiqué par chaque agriculteur.
4. **Système d'acquisition** : provient en utilisant différents types de scanners à plats « flatbed desktop scanner ».

Il est possible de récupérer, par *la diversification*, des données brutes contenant les instants et les valeurs clés et déterminants, pour effectuer une discrimination proche parfaite. Étant donné que le phénotype, dans une partie, est le résultat de l'interaction du génome avec le milieu et les aléas de ce dernier, des moments critiques ou des combinaisons critiques durant le cycle biologique et la chaîne des grains peuvent donner naissance à une empreinte spécifique. Ceci permettra de prédire l'environnement dans lequel la plante a évolué, accidents physiologiques, échaudage... (c'est un niveau plus bas de classification : zones, année de cultivation, conditions de cultivations : l'empreinte du sol, de l'environnement ...itinéraire technique,) ceci semble de la science-fiction mais avec le progrès technologique rien n'est impossible c'est juste une question de temps et de savoir.

Nous nous limitons à l'objectif principal de notre recherche qui est de faire une classification d'images des grains de blé à un niveau variétal.

La stratégie de « diversification », transcrit une partie de notre vision, projection et valorisation du présent travail de recherche. Ceci par la construction d'un Framework de base pour un système d'information sur la qualité de semences et l'identification et la reconnaissance des espèces et variétés de blé. C'est une simulation des différentes interactions qui peuvent exister sur le terrain.

II. La collecte des grains

II.1. Le matériel végétal

Les grains de blé que nous avons utilisés dans nos expérimentations sont issus du programme national algérien de la multiplication des semences en classe *prébase et base* ; le programme est supervisé et suivi par des établissements étatiques (CNCC, ITGC, OAIC). Le programme implique les multiplicateurs comme étant les acteurs principaux de production de semence. Les échantillons collectés (Tableau 2) représentent les variétés les plus cultivées à savoir : SIMETO et VITRON de l'espèce blé dur *Triticum durum Desf.*, et les variétés : ARZ var et HD de l'espèce de blé tendre *Triticum aestivum*.

Tableau 2: identité du matériel végétal utilisé

Variété	Origine	Pedigree	Obtenteur	Demandeur	Date d'inscription
HD	Mexique	HD 1220/3*Kel/Mal CM	CIMMYT	ITGC	1998
ARZ	Mexique	Mayo54IL/TAC S	CIMMYT	ITGC	1998
SIMETO	Italie	Capeiti*valvona	IAOITALIO	ITGC	1998
VITRON	Espagne	*	*	ITGC	1998

A rappeler, qu'il existe une différence de qualité significative entre le produit issu du programme de multiplication et celui du programme de consommation, comme nous l'avons décrit dans la description de la chaîne du grain (chapitre 2). Les grains de semence sont de meilleure qualité grâce aux règles et directives de production imposées par (ISTA, UPOV, OCDE), ces derniers organisent la conduite du programme de multiplication, ce qui doit regrouper les conditions de culture optimales (optimum culture).

En ce qui concerne la quantité finale utilisée pour la classification *shallow* elle était de **80 600** grains représentant une **variation temporelle** de 03 années de récoltes 2017/2018/2019/ ; quant à la quantité utilisée pour la classification *Deep Learning* elle est de 60 368 grains de semence intacts couvrant la même période.

La variabilité spatiale de la base de données compte 48 emplacements de 12 wilayas (Tableau 3); nous avons supposé que les emplacements choisis sont répartis dans les principales zones bioclimatiques de l'Algérie, malgré que cela a été plutôt influencé par la disponibilité ; concernant **la variabilité agronomique**, nous avons obtenu les grains de 48 agriculteurs, dont on a présumé l'existence d'une différence corrélative entre les systèmes de production

(itinéraire technique) pratiqués par chaque agriculteur en fonction de son sol et disponibilité du matériels.

Tableau 3: informations relatives aux grains collectés

Espèce	Variété	Nombre de Zone		Matricule	Nombre d'agriculteur
Blé dur Hard Wheat (<i>Triticum durum Desf</i>)	SIMETO	9	9	2 ; 9 ;10 ;15 ;25 ;35 ; 42 ; 44 ;47	12
	VITRON	5		9 ; 25 ; 35 ; 42 ; 47	11
Blé tendre Baird Wheat (<i>Triticum aestivum</i>)	ARZ	6	11	2 ; 10 ; 22 ; 25 ; 26 ;44	11
	HD	8		9 ;15 ;20 ;22 ;25 ;35 ;42 ;44	11

Le précédent cultural et la rotation des cultures ne permettent pas d'avoir des échantillons de la même parcelle pour deux années successives, par contre on peut cumuler les années pour la même zone géographique.

La disponibilité nous a posé un sérieux problème, durant les années de la réalisation de cette thèse, la réglementation étant stricte et le cadre collaboratif n'avait pas encore un environnement favorable à un échange direct avec le milieu socio-professionnel .

II.2. Démarche et modalités de la collecte des grains

En tenant compte de la variabilité spatiale et agronomique, les échantillons des grains ont été prélevés selon deux approches (Figure 33). La première approche est la collecte directe au champ, où nous avons récolté directement et aléatoirement des épis de blé des 04 variétés, ensuite nous les avons battus manuellement pour obtenir les grains qui représente 50 % de l'ensemble de données. On vise par la première approche à préserver toutes les caractéristiques possibles des grains qui peuvent être détériorés par le battage et la manutention mécanique aux étapes suivant la récolte.

Concernant la deuxième approche, nous avons obtenu des échantillons de grain en vrac (qui représentent 50 % de l'ensemble de données) à partir des installations de collecte d'entreposage officielles régionales (CCLS), ainsi qu'à partir de l'ITGC et le CNCC. Ce dernier reçoit les échantillons pour le contrôle et la certification.

La plupart des variantes de forme et de taille des grains sont supposées être présentes dans chacun des échantillons collectés, car aucune étape de tamisage n'a été appliquée. Les

échantillons ont été emballés dans des sacs appropriés, étiquetés et entreposés dans des conditions ambiantes.

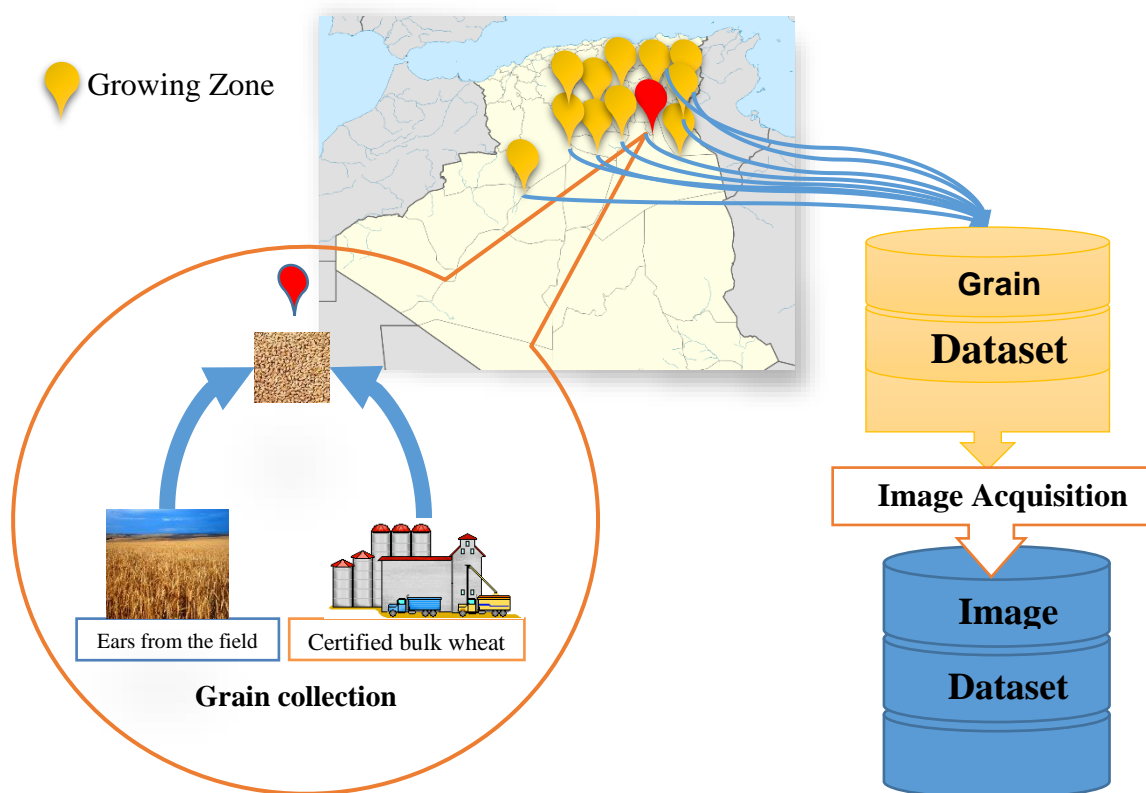


Figure 33: flux de travail de la collecte des Grains (Laabassi *et al.*, 2021)

II.3. Constat et prospective

La présente démarche de collecte de grain sera une des règles à intégrer dans le processus automatique d'identification. Nous pensons à normaliser la collecte pour cette finalité comme suit :

- La collecte d'apprentissage et de mise à jour se fera chaque année à partir des sites de cultures références officielles (variétés cultivées) ces sites témoins sont choisis dans chaque région et **Géo-référenciés (GPS)**.
- Pendant la campagne de contrôle on peut prélever des échantillons d'apprentissage des différents endroits où les grains peuvent avoir de nouvelles apparences ou de nouvelles informations.

- L'agriculteur ramène des échantillons en épis par exemple pour le contrôle et la certification. La base de données de l'apprentissage du modèle de classification va contenir des échantillons certifiés de la même zone ou les mêmes conditions de culture, ce qui permet de d'augmenter la précision.
- Au niveau des moissonneuses batteuses, à la réception, des silos.

Une empreinte numérique de la variété sera, et doit être, déclarée par l'obteneur de la variété et aussi exigé dans le protocole d'inscription au catalogue officielle ainsi que dans les procédures de contrôle et de certification des semences.

Plus loin, prédire la zone de production pour une meilleure traçabilité du produit, c'est le produit qui doit se définir lui-même, ce n'est pas par l'emballage qui le contient ou son étiquette.

III. La collecte des images

La collecte des images représente la deuxième étape la plus importante dans le flux du travail de la collecte des données et l'implémentation de la méthode d'identification automatique. Cette étape se résume dans l'opération de l'acquisition du signal représentant le grain, et la conversion (abstraction) des grains « objet » en données pixels, c'est-à-dire « représentation digitale de l'objet ».

Etant donné que le problème de l'apprentissage machine traité est relatif à la classification d'images, dans ce cas de figure on implique deux jeux de données, un pour l'apprentissage : entraînement et validation (**Training and validation data**) ou la construction du modèle et un autre qui servira comme un jeu de données de test (**Testing data**).

Nous avons donc créé deux bases de données images, pour exécuter les deux expérimentations de notre travail de recherche. Ces bases de données sont :

- **Une base de données de grandes images** : qu'on a appelé **principales images**, qui sont les images sources, directement exploitable dans l'expérimentation *Shallow approche* par l'application *Smart Grain* et aussi adaptée pour le script que nous avons développé pour la segmentation.
- **Une base de données d'images individuels de grains** : cette base est issue de la segmentation des grandes images en petites images de grains individuels, destinées pour l'apprentissage en *Deep Learning*

Nous avons monté et adapté un système d'acquisition et une méthodologie que nous décrivons ci-dessous.

III.1. Le système d'acquisition

La mission du système d'acquisition est de capter des données nécessaires pour l'abstraction des objets. Dans cette étude, les informations recherchées concernent les traits caractéristiques des grains sur une image (la forme et la taille des grains, leurs couleur et texture). Pour les extraire par la suite en données d'apprentissage et qu'on évaluera.

Vue les difficultés d'acquérir des capteurs plus sophistiqués tels que les cameras multispectrales ou des scanners 3D, ainsi que les préférences et les exigences en général qui tendent vers des systèmes basiques, disponibles et fonctionnels, le choix a été porté sur un système d'acquisition aussi basique que possible. Le système est composé à la base d'un scanner à plat (un scanner de bureau). Les plus disponibles sont les modèles équipés avec des capteurs de technologie CIS (Contact Imaging Sensor)⁴.

Le scanner à plat a déjà été utilisé par plusieurs auteurs (Patrício and Rieder, 2018) tel que nous l'avons mentionné dans la partie bibliographique . Ce système d'acquisition permet de capturer des images en 2D d'un seul côté du grain (face ventrale, tel que détaillé plus bas dans le protocole d'acquisition).

Les facteurs de choix sus cités ont été pris en considération dans notre étude pour mieux contextualiser la solution technologique que nous avons proposée.

Un autre facteur de choix aussi important, c'est l'utilisation de différents modèles de dispositif pour l'acquisition d'images visant à réduire l'uniformité éventuelle qui pourrait être fournie en raison de l'utilisation du même dispositif. Nous avons supposé que chaque scanner peut avoir sa propre empreinte et elle ne doit pas être une clé de détermination (classification en Deep Learning), nous avons utilisé selon la disponibilité plusieurs modèles. Tel que nous l'avons mentionné dans la stratégie de collecte en matière de variabilité, quatre scanners ont été utilisés comme suit :

- Trois (03) appareils pour acquisition des images de la base de données destinées à la construction des Modèles. (Image data set for training and validation step).

<https://www.tavco.net/wide-format-plotter-scanner-blog/bid/107329/compare-ccd-vs-cis-scanner-technologies>⁴

- Un appareil (01) pour l'acquisition des données images de test des modèles
- Un laptop pour l'acquisition et le pré-processing.
- 04 scanners de différentes marques : Epson, Brother, Canon et HP.

Cette solution va servir de plateforme pour la conception d'un système d'acquisition opérationnel et ergonomique et fonctionnant à haut débit.

III.2. Démarche de l'acquisition des images

Le protocole (figure 34) est réalisé en plusieurs étapes :

- 1) Le dépôt des grains d'une variété bien définie sur le plateau du scanner avec une distribution et un arrangement des grains étudiés ;
- 2) Mettre l'échelle sur la scène et
- 3) Mettre l'arrière-plan et en
- 4) Le scan.
- 5) L'annotation, doit être définie instantanément pour chaque image scannée.

Au total un nombre de 131 images principales des grains en 2D sont acquises avec les caractéristiques suivantes : Chaque image peut contenir entre 500 à 1013 grains. Toutes les images sont scannées avec un arrière-plan noir. La taille des images est et doit être identique (normalisée A4) ; la profondeur des images est de 24bits avec une résolution de 300 et 600 dpi (dots per inch), les images sont stockées en format JPEG et TIFF. En suite à partir de ces images principales nous avons obtenu par segmentation les images des grains individuels (single grain image).

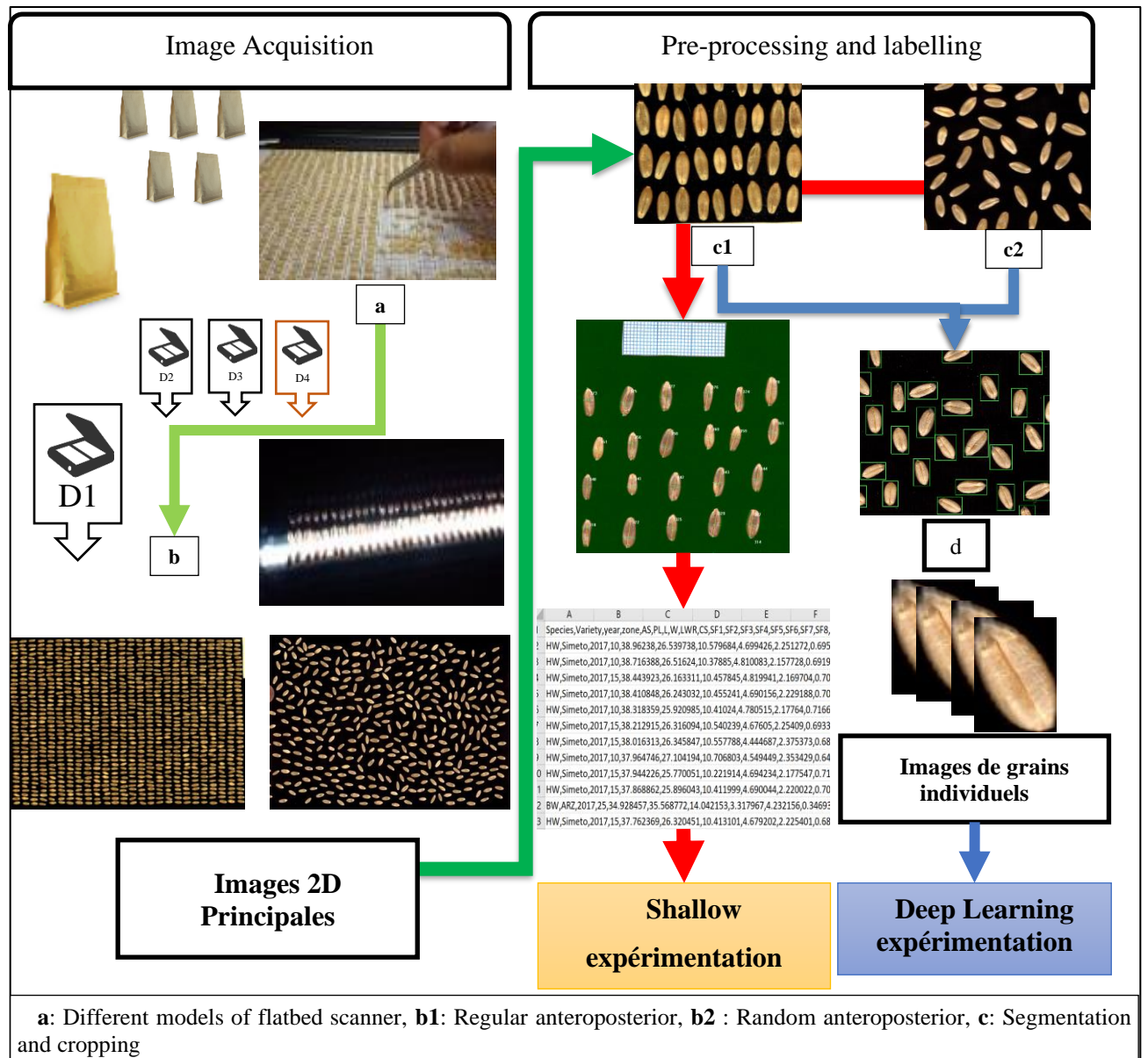


Figure 34: flux de la collecte des images des grains de blé

III.2.1 Distribution et arrangement des grains

Les modalités de dépôt des grains sur le plateau sont conditionnées par plusieurs facteurs liés à l'objectif principal de cette étude (VLC et SLC). Ainsi que, des conditions du contexte d'application, dans le cas des tests de pureté variétale et les analyses de pureté physique, ici la détermination et la reconnaissance n'est applicable et fonctionnelle que si l'identité de chaque grain est déterminée individuellement.

De ce fait, la distribution des grains est en fonction de la nature des grains, la face du grain à scanner et le contact entre les gains et la segmentation. On peut la décrire aussi comme étant un compromis entre la nature, la quantité et la qualité des informations à extraire du grain. Ceci se traduit par la recherche de la face effective à scanner et l'orientation du grain sur le plateau du scanner.

La figure 35 montre les cas et les situations de la distribution et l'arrangement des grains sur le plateau du scanner qu'on a observé et identifié :

1. Cas des grains séparés :
 - Situation N°1: La même face, même direction (orientation arrangé)
 - Situation N°2 : La même face, direction (orientation) aléatoire.
 - Situation N°3 : les faces aléatoires, direction (orientation) aléatoire.
2. Cas des grains en contact (en vrac) :
 - Situation N°4 : face aléatoire, direction (orientation) aléatoire, contact faible
 - Situation N°5 : face aléatoire, direction (orientation) aléatoire et un contact arrangé (compacte).

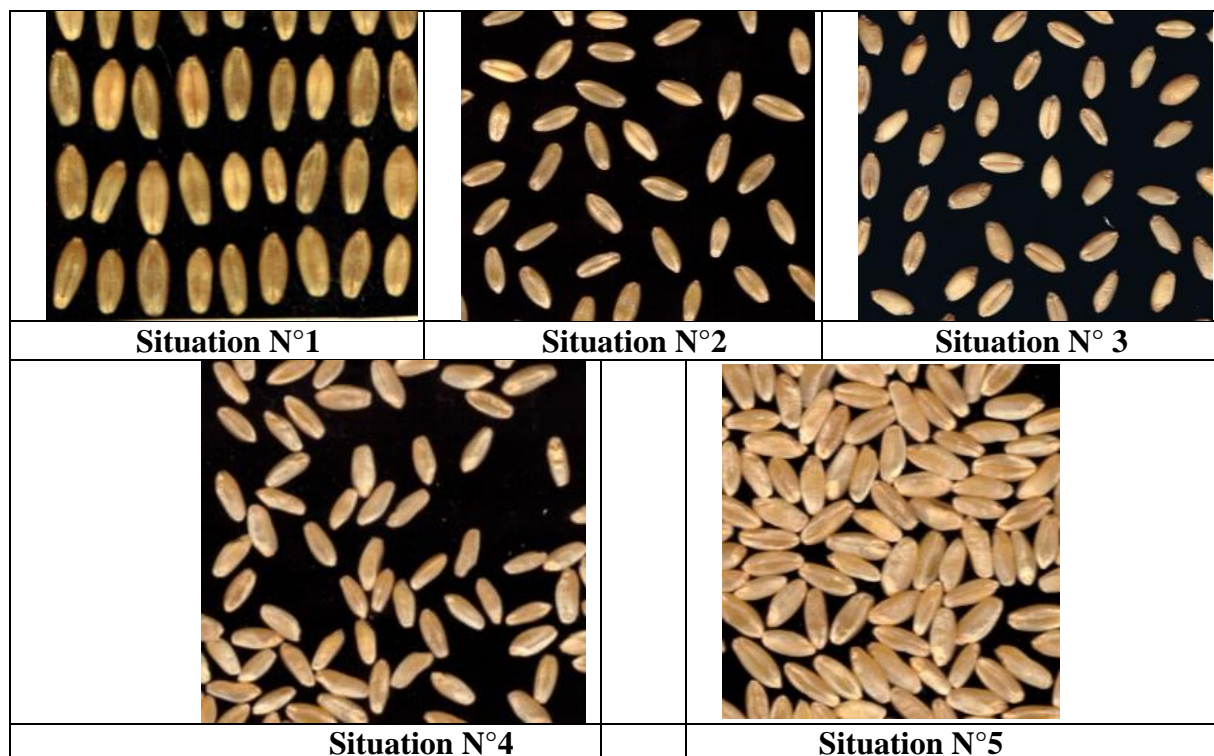


Figure 35 : différentes situations de dispersion, la quelle est la plus adéquate

La séparation entre les grains (une distance minimale) et la fixation de la face qui peut contenir le plus grand nombre d'informations visuelles aident à réaliser facilement la segmentation et l'extraction des caractéristiques (morphologique surtout) pour la classification des grains d'une manière homogène et pertinente.

La face ventrale et les situations de dispersions des grains « situation N°1 » et « situation N°2 » sont les choix les plus adéquats à notre étude (figures 35-36). La situation N° 1 permet un certain nombre d'avantages liés à la connaissance de la position exacte du grain sur l'image, comme dans une étude d'estimation du poids des grains (un master a été réalisation durant l'année 2020-2021) ainsi que d'autre cas d'études.

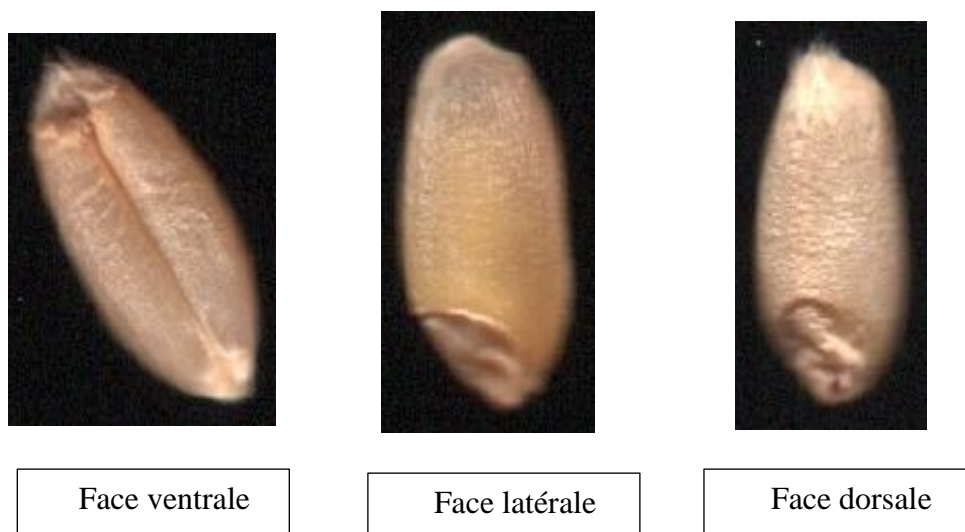


Figure 36: différentes faces de la graine de blé, laquelle faut-il choisir ?

A noter que la situation N°3 qui représente des grains dispersés aléatoirement (selon les faces), cette situation n'est pas adéquate pour extraire les caractéristiques morphologiques vu l'hétérogénéité des dimensions entre chaque face des grains. Quant aux situations N°4 et N°5, ces deux dernières reflètent un autre arrangement dont la finalité recherchée n'est pas l'authentification individuelle des grains mais de l'ensemble.

Une fois que les grains sont déposés, on procède aux étapes suivantes du protocole, il faut vérifier que les conditions de dispersions soient respectées pour éviter d'avoir des grains en contact entre eux ou sur d'autres faces, ce type de données peut réduire la pertinence de l'apprentissage. L'expérimentation a été réalisée sur un ensemble d'images acquises à partir de quatre scanners à plat.

III.2.2 Choix de l'arrière-plan

On a essayé plusieurs couleurs pour l'arrière-plan afin de faciliter les opérations de traitement et d'analyse des images par le Smart Grain, et aussi pendant l'étape de la segmentation pour obtenir les images des grains individuels. Puisque les couleurs des grains de blé sont plus claires, le fond noir était le plus adapté pour l'acquisition des images. Afin d'obtenir le fond noir, nous avons effectué l'opération de l'acquisition dans une chambre noire et puis récemment nous avons utilisé une couverture opaque (feuille de couleur noire) pour réduire les bruits provenant des sources de lumières externes.

III.2.3 Définition de l'échelle :

Mettre sur le plateau du scanner (la scène) l'échelle de dimensionnement (scale) est une obligation pour permettre de définir le rapport nombres de pixels et la valeur des grandeurs de base de millimètres [pix/mm] ; nous avons omis cette étape dans les premières tentatives de faire la caractérisation avant de s'apercevoir de sa grande importance, c'est une étape importante du protocole d'acquisition. Nous avons utilisé une feuille millimétrique d'une dimension de 10*30 mm (figure 37), on doit prévoir un repère normalisé jusqu'à 0.01mm. Pour garder la valeur de l'échelle identique pour toute les images, la taille du scan est normalisée en format A4 pour le scan et le fichier de sortie.

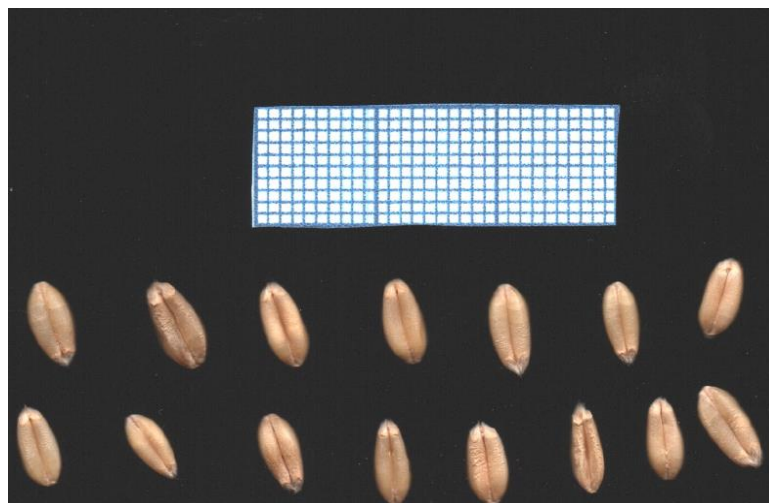


Figure 37: Une feuille millimétrique pour définir l'échelle sur les images

III.2.4 Paramétrage du scan et annotation des images :

Chaque scanner a sa propre application qui le gère et qui permet de fixer les valeurs de la résolution, la taille et d'autres paramètres. Nous avons utilisé une application qui nous a permis de prendre en charge tous les modèles de scanners utilisés. Cette application nous a permis aussi de structurer les dossiers et les images et les annoter (figure 38). Le chemin d'accès aux images prend la forme suivante. Chaque sous dossier représente un niveau de classification, la structuration des dossiers est importante pour l'implémentation des scripts de classifications et d'explorations des données en général.

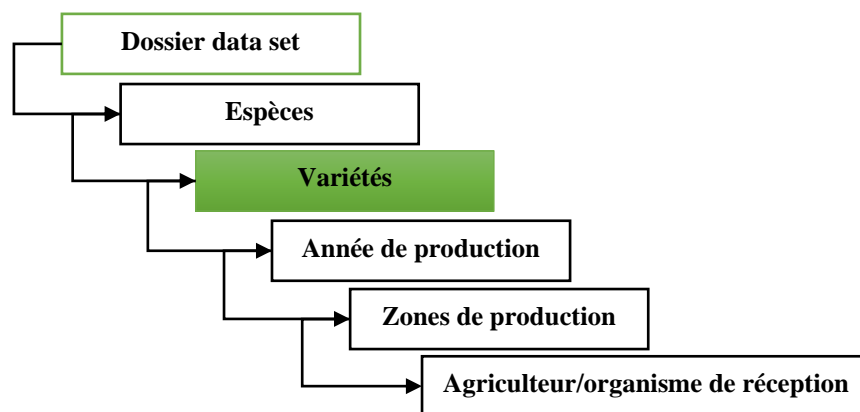


Figure 38: Organisation des dossiers de la base de données images

III.3. Obtention des images de grains individuels

Les images de grains individuels destinées à l'expérimentation N°2, sont obtenues à partir des images principales par une opération de segmentation. Cette opération consiste à localiser les grains sur l'image et faire la découpe « cropping » de chaque grain sans supprimer l'arrière-plan. Un script de segmentation écrit en langage Python et un autre script segmentation en utilisant Matlab (qui ont donné le même résultat).

Nous avons obtenu 60 368 images de grains individuels (Figure 39) enregistrées en format TIFF avec une profondeur de 24 bits et une résolution de 96ppp. Les dimensions des images ne sont pas identiques vu que naturellement les grains ne sont pas identiques et ces images n'ont pas été redimensionnées ou normalisées.

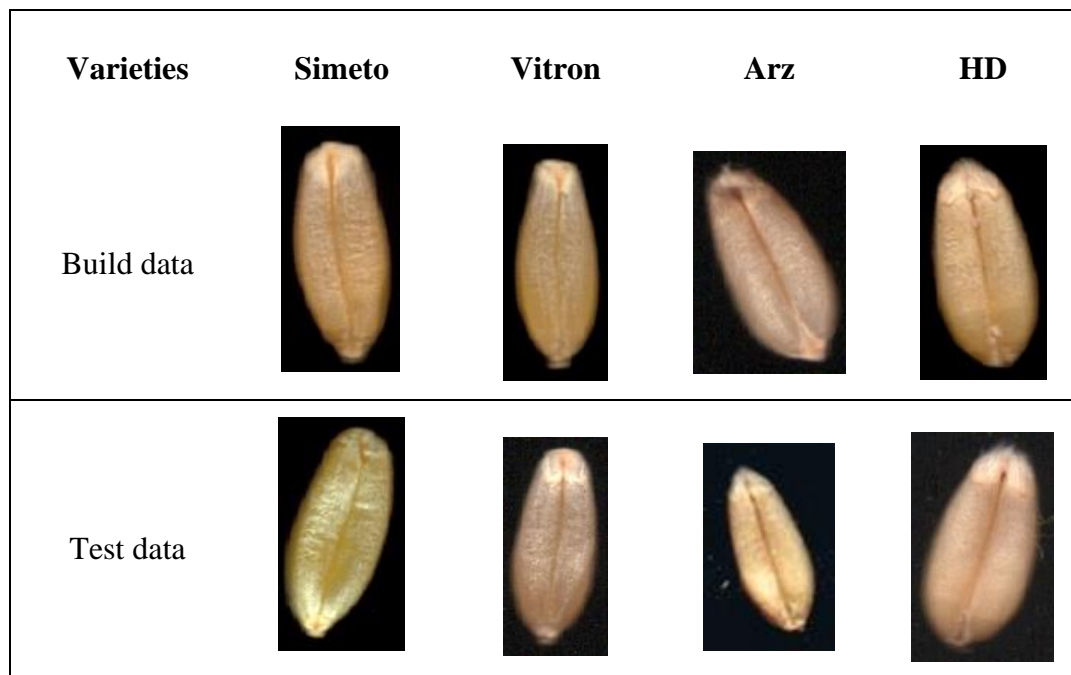


Figure 39: images des grains de chaque variété

IV. Conclusion

Les données sont la matière de base pour la construction des modèles de Machine Learning, la pertinence de ces derniers est liée étroitement aux attributs que contient chaque base de données et leurs capacités à définir un modèle d'une identité biodigitale.

Nous avons construit dans ce chapitre une démarche de collecte des données nécessaires à effectuer la GIR et par conséquent la VLC, incluant la variation spatiale temporelle et technologique et aussi les différents format et disposition de l'entité biologique afin de capter les attributs voulus et aussi les attributs inconnus ou non quantifiable par notre système de perception (humain).

Cette formulation trouvera sa démonstration dans les deux expérimentations que nous avons réalisées dans ce travail de recherche.

Chapitre 2 : Expérimentation N° 1

SHALLOW APROCHE

Matériels et méthodes

I. Introduction

L'expérimentation *Shallow* est l'application classique de la technique de classification d'images. Durant les années de notre investigation et de la réalisation de ce travail, cette méthode a été un passage inévitable dans l'implémentation des techniques de vision artificielle. C'est la résultante des premières tentatives à la recherche d'une méthode automatique et non destructive pour la caractérisation, l'identification et la reconnaissance des grains de blé à l'échelle variétale.

Comme nous l'avons développé dans la revue bibliographique, dans notre cas d'étude cette méthode permet de récupérer des informations détaillées sur l'aspect morphologique des grains dont la finalité est de représenter et de quantifier les caractéristiques liées à la taille et la forme en données numériques pour ainsi réaliser les tâches de ML.

Une précision par rapport à l'opération '*extraction des caractéristiques*', la couleur et la texture n'ont pas été investiguées, en raison de difficultés techniques liées à la définition de l'échelle colorimétrique et à l'utilisation des algorithmes d'extraction des textures. Cependant, nous avons estimé que ces deux caractéristiques sont inévitablement prises en considération dans la technique de Deep Learning (vu que l'extraction se fait automatiquement). Par souci de vocabulaire et pour plus de commodité dans cette expérimentation, nous avons utilisé le terme « **attribut** » à la place le terme « **caractéristique** ».

Toutefois, le principal objectif dans cette expérimentation N°1, est de vérifier la faisabilité d'une classification d'images basée sur les attributs morphologiques des grains ; ce qui permettra la reconnaissance de l'identité des grains (SLC et VLC).

D'autres objectifs opérationnels récurrents :

- Déterminer le classificateur qui affiche les meilleurs résultats.
- Déterminer quels sont les attributs morphologiques les plus pertinents et discriminants.
- Vérifier et réajuster la conduite de la collecte des données (feedback) et l'implémentation de la Computer Vision.
- Ecrire et améliorer un script de classification en Python.
- Autres.

Les données collectées dans cette première expérimentation nous ont permis de réaliser d'autres investigations connexes :

1. Une étude sur la variabilité morphologique des variétés de blé en Algérie pour établir une carte variétale.
2. Une partie de ces données a fait l'objet d'un essai de classification dans un mémoire de fin d'étude que nous avons encadré.
3. Une autre étude pour la prédiction du poids des grains basée sur la technique de régression.

II. La base de données de l'apprentissage :

L'expérimentation est fondée sur l'utilisation des caractéristiques morphologiques qui reflètent la taille et la forme des grains ; elles sont exploitées sous forme de (**données numériques**) extraites au préalable à partir des **131** images principales des grains de blé (**données images**).

Nous avons impliqué pour cette expérimentation, une base de données contenant des informations morphologiques de **80 600** grains, représentées par un ensemble de **16 attributs morphologiques**, dont **06 attributs** ont été extrait par Computer Vision ; ces 06 attributs ont été utilisés pour calculer les **10** autres attributs qu'on appelle *facteurs de formes*. (Voir détails plus bas).

III. Protocole expérimental SHALLOW

Le protocole (Figure 40) suit en continuité l'étape de la collecte des données (images principales) mentionnée dans le chapitre précédent (data collection). Deux principales opérations supplémentaires sont réalisées comme suit :

- **Computer Vision** :
 - Extraction des attributs morphologiques : Traitement et analyse des images principales ;
- **Apprentissage machine** (classification d'images) :
 - Le prétraitement des données de sortie numérique de la phase Computer Vision
 - L'analyse des données numériques (classification), cette dernière comprend des tâches élémentaires :
 - La construction du modèle de classification
 - La sélection des caractéristiques de classification performant
 - Le test du modèle de classification.
- **Analyse** et visualisation des résultats.

Nous avons effectué deux manipulations pour chaque niveau de classification :

- Manipulation A (**Mnp A**) : La classification par les données brutes (tous les attributs).
- Manipulation B (**Mnp B**) : La sélection (réduction) des caractéristiques de classification les plus discriminants.

La figure 40 montre les différentes étapes et le déroulement de l'expérimentation Shallow

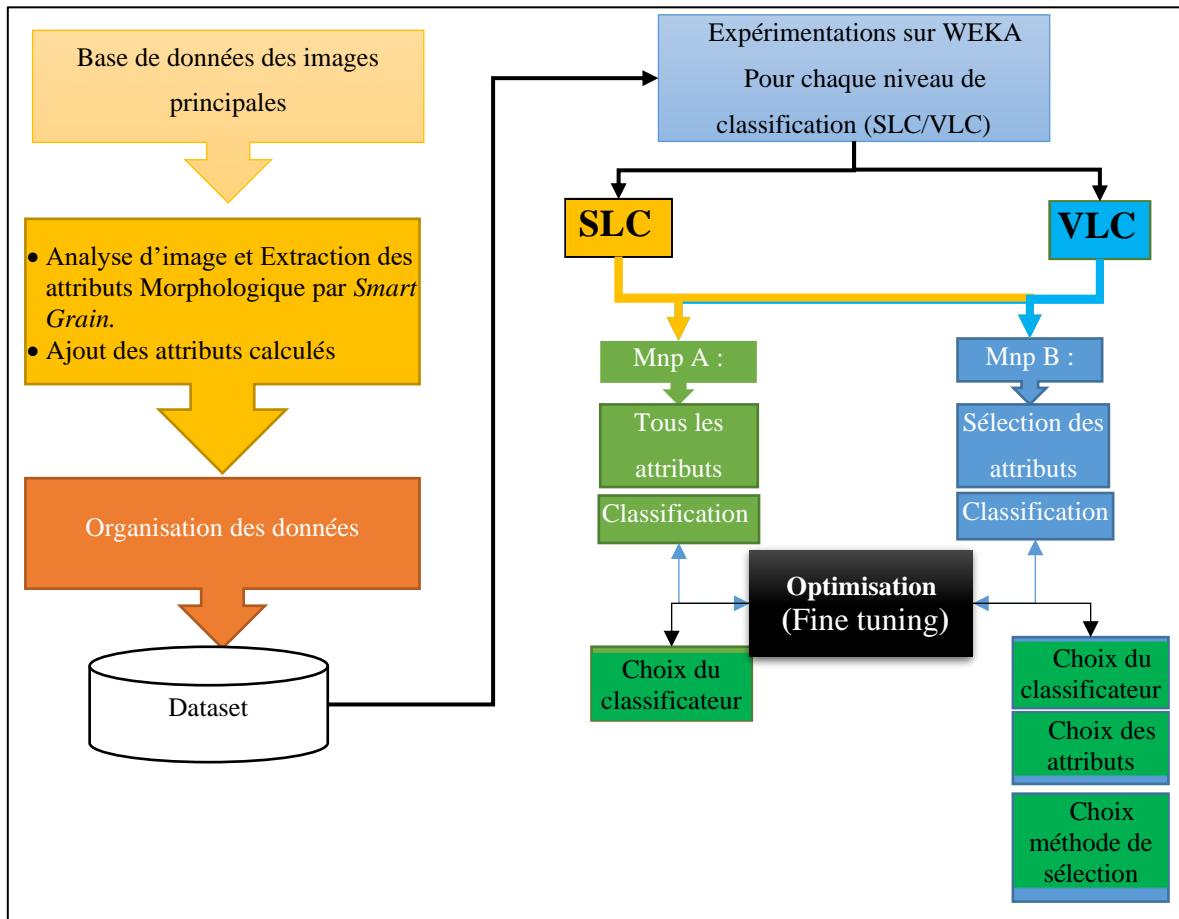


Figure 40: Protocole expérimentale SHALLOW

IV. L'extraction des Attributs morphologiques

L'extraction des données de la forme et de la taille de chaque grain a été réalisée par l'application de traitement et d'analyse d'images *Smart Grain*® développée par (Tanabata *et al.*, 2012) dans le cadre de leurs travaux de phénotypage des grains.

La figure 41 représente les différentes étapes du flux de travail de l'extraction des attributs morphologiques, elles sont comme suit : la lecture du dossier des images principales, le paramétrage de l'application incluant la définition de l'échelle (rapport pixels/ longueur en mm), le choix de la couleur de l'arrière-plan, l'ajustement de la précision et de la finition, le choix de la technique de segmentation, ensuite le lancement de l'analyse et la sauvegarde des données en fichiers CSV.

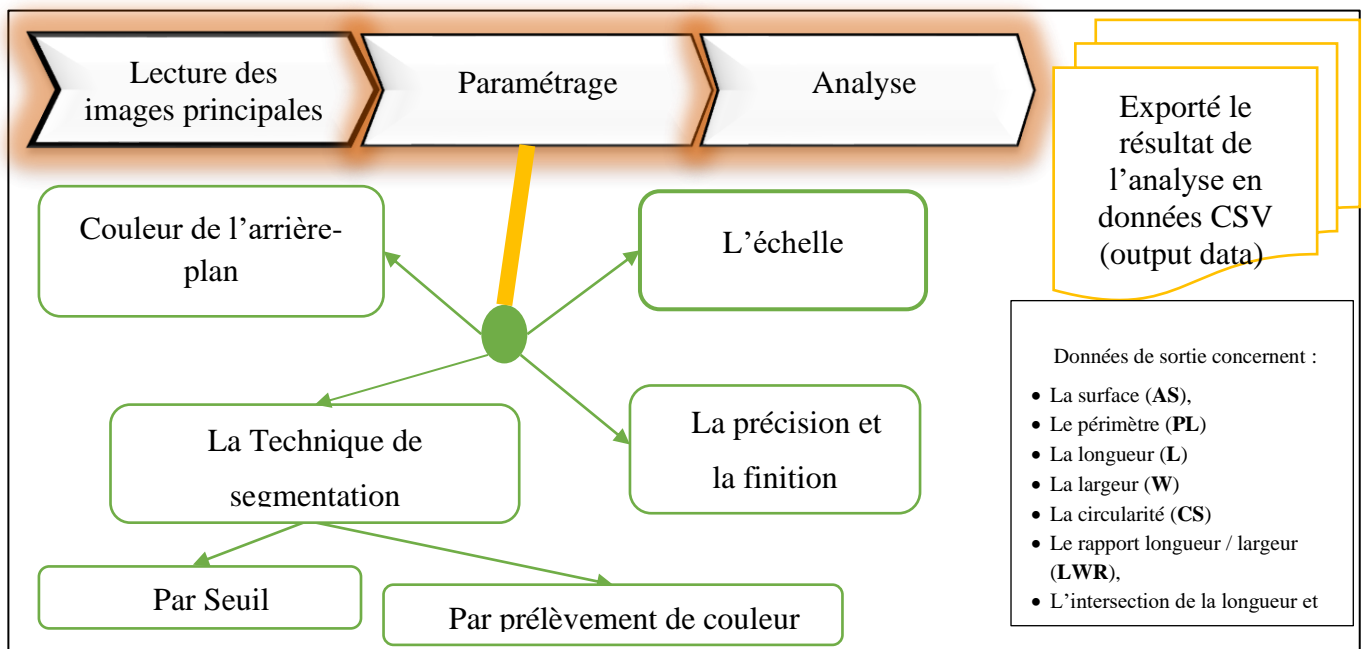


Figure 41: les différentes étapes du flux de travail de l'extraction des attributs morphologiques

L'application peut mesurer et calculer plusieurs paramètres : la surface (**AS**), le périmètre (**PL**), la longueur (**L**), la largeur (**W**), la circularité (**CS**), le rapport longueur / largeur (**LWR**), l'intersection de la longueur et de la largeur (**IS**), le centre de gravité (**CG**) et la distance entre IS et CG (**DS**). Nous avons complété la base de données d'entrées par des facteurs de formes **SF** (Shape Factor) que nous avons calculés à la base des formules proposées par différents auteurs (tableau 4).

Tableau 4: Attributs morphologiques calculés recensés de la bibliographie

Caractéristique	Formule de calcul	Références
Shape Factor1 (SF1)	$SF1=AS/L$	(Stephen J Symons and Fulcher 1988)
Shape Factor2 (SF2)	$SF2= (L*W)/AS$	
Shape Factor3 (SF3)	$SF3=L/AS$	
Shape Factor4 (SF4)	$SF4= AS/((L/2)2*\pi)$	
Shape Factor5 (SF5)	$SF5= AS/((L/2)*(W/2)*\pi)$	
Shape Factor6 (SF6)	$SF6=\sqrt{((4AS)/ \pi)}$	(Visen et al. 2002)
Shape Factor7 (SF7)	$SF7=AS/L^3$	(Chaugule et Mali, 2014)
Shape Factor8 (SF8)	$SF8=SF6/L$	(Arefi et al. 2011)
Shape Factor9 (SF9)	$SF9=PL^2/4\pi*AS$	(Shouche <i>et al.</i> , 2001).
Shape Factor11 (SF10)	$SF10=4\pi*AS/PL$	(Mebatsionet <i>al.</i> ,2013).

V. Classification d'images et Analyse de données

Nous avons expérimenté plusieurs classificateurs pour l'échelle espèce (SLC) et pour l'échelle variété (VLC) ; dans un premier temps (Mnp A) nous avons impliqué tous les attributs morphologiques (forme et taille). Ensuite (Mnp B) par des méthodes de sélection des caractéristiques nous avons effectué une recherche des attributs les plus discriminants pour chaque niveau de classification. Par la suite nous avons refait les classifications par les attributs sélectionnés en fonction de chaque algorithme de sélection. L'analyse des données a été réalisée sur la plateforme de datamining WEKA® (Bouckaert *et al.*, 2013; Ian Witten, WAIKATO, 2016).

V.1. Les classificateurs (classifier)

En mode exploratoire, le choix des algorithmes, a été défini en premier lieu par le cas d'étude « classification » du fait que les classes sont catégoriques (nominales) ; ensuite de chaque famille des classificateurs disponible sur la plateforme WEKA nous avons choisi ceux qui sont compatibles avec nos données, il s'agit de 04 familles à savoir : *Bayes, Fonctions, Meta et Tree*. La classification est menée en mode *Cross validation* pour construire et valider chaque modèle de ML.

La sélection des classificateurs a été portée sur ceux qui ont affiché des performances de précision plus de 85% en niveau espèce. Les configurations ont été laissées sur l'option par défaut pour effectuer le choix initial, ensuite les trois meilleurs algorithmes retenus ont fait l'objet *d'Optimisation d'hyperparamètres (Fine tuning)*, une investigation poussée à la recherche de la meilleure configuration qui améliore la précision. Les classificateurs retenus sont (avec leurs appellation dans la plateforme WEKA):

- **Le classificateur Multilayer Perceptron** : est un classificateur de réseau de neurones qui utilise la rétro-propagation pour apprendre à un perceptron multicouche à classer les instances. Les paramètres du réseau peuvent également être surveillés et modifiés pendant le temps d'entraînement. Les nœuds de ce réseau sont tous sigmoïdes (sauf lorsque la classe est numérique, auquel cas les nœuds de sortie deviennent des unités linéaires sans seuil).
- **MLP Classifier** : Est un perceptron multicouche à une seule couche cachée, l'entraînement se fait en utilisant la classe d'optimisation de WEKA en minimisant la fonction de perte donnée, en plus d'une pénalité quadratique.
- **Le Support Vector Machine** : Implémente l'algorithme d'optimisation minimale séquentielle (SMO) de John Platt (1998) pour l'entraînement d'un classificateur à vecteurs de support. Cette implémentation remplace globalement toutes les valeurs manquantes et transforme les caractéristiques nominales en caractéristiques binaires. Les problèmes multi-classes sont résolus en utilisant la classification par paires (Hastie et Tibshirani, 1998).
- **Random Forest** : Classe pour construire une forêt d'arbres aléatoires. (Breiman, 2001)
- **Bagging** (weka.classifiers.meta.Bagging) : Classe pour fixer un classificateur afin de réduire la variance.(Breiman, 1996).
- **meta.AttributeSelectedClassifier**: la dimensionnalité des données d'apprentissage et de test est réduite par la sélection d'attributs avant d'être transmise à un classificateur.
- **meta.ClassificationViaRegression** : Classe en utilisant des méthodes de régression. La classe est binarisée et un modèle de régression est construit pour chaque valeur de classe.

V.2. La sélection des Attributs

Le tableau 5, regroupe les différentes méthodes d'évaluation utilisées pour la sélection des attributs, tandis que le tableau 6 regroupe les méthodes de recherche correspondantes aux méthodes d'évaluation. La description des différentes méthodes et algorithmes est détaillée sur le site : *Weka Application Programming Interface*.

(<https://weka.sourceforge.io/doc.dev/overview-summary.html>).

Tableau 5 : Méthodes d'évaluation des attributs

N°	Méthode d'évaluation	Description
1	CfsSubsetEval	Évalue la valeur d'un sous-ensemble d'attributs en tenant compte de la capacité prédictive individuelle de chaque caractéristique ainsi que du degré de redondance entre elles. (Hall and Smith, 1998)
2	ClassifierAttributeEval	Évalue la valeur d'un attribut à l'aide d'un classificateur qui doit être le spécifier dans la configuration. Dans notre cas nous avons choisi la configuration par <i>défaut</i> .
3	InfoGainAttributeEval	Évalue le mérite d'un attribut en mesurant le ratio de gain par rapport à la classe.
4	OneRAttributeEval	Évalue le mérite d'un attribut par le classificateur OneR
5	ReliefFAttributeEval	Évalue le mérite d'un attribut en échantillonnant à plusieurs reprises une instance et en considérant la valeur de l'attribut donné pour l'instance la plus proche de la même classe et d'une classe différente. Peut fonctionner sur des données de classe discrètes et continues.

Tableau 6: Méthode de recherche pour la sélection et le classement des attributs

N°	Méthode de recherche	description
1	GreedyStepwise Utilisée par la méthode d'évaluation 1	Effectue une recherche à vide vers l'avant ou vers l'arrière dans l'espace des sous-ensembles d'attributs. Elle produit une liste classée d'attributs en parcourant l'espace d'un côté à l'autre et en enregistrant l'ordre dans lequel les attributs sont sélectionnés.
2	Attribute ranking Utilisée par la méthode d'évaluation 2,3,4,5	Classe les attributs selon leurs évaluations individuelles. À utiliser conjointement avec les évaluateurs d'attributs (ReliefF, GainRatio, Entropy, etc.).

On a aussi sélectionné manuellement les attributs extraits par SmartGrain ® de ceux qu'on calculé et on les a investigués séparément.

V.3. L'évaluation de la classification

A la fin de chaque classification on a opté à prendre en considération les informations suivantes :

- Les résultats de la comparaison entre les classificateurs sur WekaExperiment
- Les valeurs de Précision et F-mesure, Recall;
- La Matrice de confusion ;
- La liste les caractéristiques les plus discriminants.

Résultats et discussions

VI. La classification en niveau « Espèce »

Le tableau 7 représente la synthèse des résultats des deux manipulations de la classification du niveau espèce, ainsi que les classificateurs optimisés qui ont marqué la plus grande performance par rapport aux autres classificateurs investigués à savoir : le *MLPClassifier* , *MultilayerPerceptron* et *RandomForest*.

Tableau 7 : Résumé des résultats de la classification du niveau « Espèce »

	SLC	MLPClassifier	MultilayerPerceptron	RandomForest
	<i>MnpA (all attributs)</i>	87,24%	87,27%	86,76%
<i>Mnp2</i>	<i>SmartGrain</i>	87,16%	87,16%	86,71%
	<i>ShapeFactor</i>	87,22%	87,19%	86,67%
	<i>2-3-5-8-13-15</i>	86,95%	87,08%	86,76%
	<i>13-3-2-5-6-15</i>	87,15%	87,18%	86,68%
	<i>11-6-5-8-10-14</i>	82,77%	82,75%	81,30%

La comparaison montre que le *MultilayerPerceptron* est légèrement plus performant que les autres classificateurs avec les attributs : *PL, L, LWR, CS, SF7, SF9*.

Dans ce qui suit, nous détaillons les résultats obtenus.

VI.1. Manipulation A -SLC, la classification sans sélection d'attributs

VI.1.1 Pré-sélection des classificateurs en SLC

Le tableau 8 représente les résultats de l'expérimentation tenue sur la plateforme WEKA® pour comparer et pré-choisir entre les différents classificateurs selon la configuration (**08 classificateurs, 05 Folds, 05 répétitions**), les résultats montrent des différences statistiquement significative (à 5%).

Tableau 8: résultat de la comparaison entre les différents algorithmes de classification (SLC)

Analysing	(5) funct	(1) bayes	(2) meta.	(3) meta.	(4) trees	(6) funct	(7) funct	(8) funct
Percent_correct	87.04	85.33 *	86.35 *	86.64 *	86.70	87.04	86.03 *	86.10 *
True_positive_rate	0.88	0.84 *	0.87	0.87	0.88	0.89	0.86	0.86 *
False_positive_rate	0.14	0.13 *	0.15 v	0.14	0.14 v	0.15	0.14	0.14 v
IR_precision	0.88	0.88	0.87 *	0.88	0.88 *	0.88	0.88	0.87 *
IR_recall	0.88	0.84 *	0.87	0.87	0.88	0.89	0.86	0.86 *
F_measure	0.88	0.86 *	0.87 *	0.88 *	0.88	0.88	0.87 *	0.87 *

Key:

*: The specific result is statistically worse than the baseline scheme at the significance level specified (currently 0.05).

v: The specific result is statistically better than the baseline scheme at the significance level specified (currently 0.05).

(1) bayes.NaiveBayes " 5995231201785697655

(2) meta.AttributeSelectedClassifier '-E "\CfsSubsetEval -P 1 -E 1\'' -S "\BestFirst -D 1 -N 5\'' -W trees.J48 -- -C 0.25 -M 2' -1151805453487947577

(3) meta.ClassificationViaRegression '-W trees.M5P -- -M 4.0 -num-decimal-places 4' 4500023123618669859

(4) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698

(5) functions.MLPClassifier '-N 2 -R 0.01 -O 1.0E-6 -P 1 -E 1 -S 1 -L functions.loss.SquaredError -A functions.activation.ApproximateSigmoid' -3297474276438394644

(6) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a' -5990607817048210779

(7) functions.SGD '-F 0 -L 0.01 -R 1.0E-4 -E 500 -C 0.001 -S 1' -3732968666673530290

(8) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "\functions.supportVector.PolyKernel -E 1.0 -C 250007\'' -calibrator "\functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4\'' -6585883636378691736

En fonction des paramètres *F_Measure*, *Recall* et *la precision* en case verte (tableau 8) nous avons pré-sélectionné les classificateurs en ordre décroissant à savoir le *MLPClassifier(5)*, *MultilayerPerceptron (6)* et *RandomForest (4)* respectivement.

Les classificateurs choisis précédemment ont subi des séries d'expérimentations à la recherche de la meilleure configuration avec le même rapport *05 Folds*, *05 répétitions* et *n configuration* pour chaque classificateur en fonction de la capacité du calcul (*la plateforme crash à la limite de mémoire demandée*).

VI.1.2 Optimisation d'hyperparamètres des classificateurs (MnpA)

Les tableaux 9-11, montrent les résultats finaux obtenus de l'Optimisation (Fine tuning) de chaque classificateur, les configurations retenues sont inscrites sous chaque tableau. (Ce qui est en vert c'est la configuration retenue et en jaune c'est le paramètre modifié).

Tableau 9: Optimisation (Fine tuning) du MLPClassifier (SLC)

Analysing	(1) funct	(2) funct	(3) funct	(4) funct	(5) funct	(6) funct	(7) funct	(8) funct	(9) funct
Percent_correct	87.23	87.23	87.29	87.32	87.18	87.23	87.31	87.34	87.36
True_positive_rate	0.88	0.88	0.88	0.88 v	0.88 *	0.88	0.88	0.88	0.88
False_positive_rate	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
IR_precision	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
IR_recall	0.88	0.88	0.88	0.88 v	0.88 *	0.88	0.88	0.88	0.88
F_measure	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88

Config. retenue: (1) functions.MLPClassifier '-N 4 -R 0.01 -O 1.0E-6 -P 1 -E 1 -S 1 -L functions.loss.SquaredError -A functions.activation.Softplus' -3297474276438394644

Tableau 10: Optimisation (Fine tuning) du *functions.MultilayerPerceptron* (SLC)

Analysing	(1) functio	(2) funct	(3) funct	(4) funct
Percent_correct	87.19	87.21	87.21	87.23
True_positive_rate	0.89	0.89	0.89	0.90
False_positive_rate	0.15	0.15	0.15	0.15
IR_precision	0.87	0.87	0.87	0.87
IR_recall	0.89	0.89	0.89	0.90
F_measure	0.88	0.88	0.88	0.88

Config. retenue : (3) functions.MultilayerPerceptron '-L 0.03 -M 0.2 -N 700 -V 0 -S 0 -E 20 -H a' -5990607817048210779

Tableau 11: Fine tuning du RandomForest (SLC)

Analysing	(1) trees	(2) trees	(3) trees
Percent_correct	86.71	86.74	86.79
True_positive_rate	0.88	0.88	0.88
False_positive_rate	0.14	0.14 *	0.14
IR_precision	0.88	0.88 v	0.88
IR_recall	0.88	0.88	0.88
F_measure	0.88	0.88	0.88

Config. retenue : (2) trees.RandomForest '-P 100 -I 120 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698

VI.1.3 Résultats Détaillés de la Manipulation A :

Les tableaux 12-14 montrent les résultats obtenus par chaque classificateur Optimisé pour le niveau espèce en impliquant tous les attributs. Les résultats montrent que les classificateurs Optimisés de la famille « *Function* » affichent une légère différence en performance avec celui de la famille de « *Tree* ». Les valeurs de la *précision* en ordre décroissant pour *functions.MultilayerPerceptron* et *functions.MPLClassifier* et *tree.RandomForest* sont de **87,3%** , **87.2%** et **86.8 %** respectivement. Les paramètres *F-mesure* et *Recall* sont restés dans la même valeur de précision correspondante à chaque classificateur.

La valeur de précision pour la classification du Blé dur (HW) par le classificateur *functions.MPLClassifier* (tableau 12) est de 88,1% tandis que pour le Blé tendre (BW) elle est de 86.2%.

Tableau 12: Résultats détaillés de *functions.MPLClassifier*

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
HW	0,881	0,137	0,881	0,881	0,881	0,744	0,944	0,95
BW	0,863	0,119	0,862	0,863	0,862	0,744	0,944	0,933
Avg.	0,872	0,129	0,872	0,872	0,872	0,744	0,944	0,942

La valeur de précision pour le classification du blé dur (HW) par le classificateur *functions.MultilayerPerceptron* (tableau 13) est de 88,2% tandis que pour le Blé tendre (BW) elle est de 86.2%.

Tableau 13: Résultats détaillés de *functions.MultilayerPerceptron*

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
HW	0,88	0,136	0,882	0,88	0,881	0,744	0,945	0,951
BW	0,864	0,12	0,862	0,864	0,863	0,744	0,945	0,935
Weighted A	0,873	0,128	0,873	0,873	0,873	0,744	0,945	0,944

La valeur de précision pour le classification du blé dur (HW) par le classificateur *tree.RandomForest* (Tableau 14) est de 87,6% tandis que pour le Blé tendre (BW) elle est de 85.8%.

Tableau 14: Résultats détaillés de *tree.RandomForest*

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
HW	0,878	0,144	0,876	0,878	0,877	0,734	0,939	0,945
BW	0,856	0,122	0,858	0,856	0,857	0,734	0,939	0,926
Weighted A	0,868	0,134	0,868	0,868	0,868	0,734	0,939	0,936

I.1.1.D Matrices de confusion manipulation A

D'une manière générale le blé dur est mieux classé que le blé tendre ; mais afin d'afficher les détails de la classification, la matrice de confusion visualise clairement les grains correctement classés et les grains incorrectement classés, et la spécialisation d'un classificateur à mieux identifier une espèce.

Pour mieux visualiser cette partie, la matrice de confusion SLC a révélé le détail de classification pour chaque espèce. D'après le tableau 15 des matrices de confusion, une partie du blé dur est classée dans le blé tendre et vice-versa, tous les classificateurs ont tendance à mieux classer le blé dur que le blé tendre.

Tableau 15: Matrices de Confusion des trois Classificateurs (SLC-Manipulation A)

functions.MLPClassifier				functions.MultilayerPerceptron					
<i>MnpA</i>				<i>MnpA</i>					
=== Confusion Matrix ===				=== Confusion Matrix ===					
Classés		a	b	Précision		Précision			
HW	a	38088	5153	88,08%	HW	a	38050	5191	88,00%
BW	b	5128	32233	86,27%	BW	b	5070	32291	86,43%
trees.RandomForest				trees.RandomForest					
<i>MnpA</i>				<i>MnpA</i>					
=== Confusion Matrix ===				=== Confusion Matrix ===					
Classés		a	b	Précision		Précision			
HW	a	37953	5288	87,77%	HW	a	37953	5288	87,77%
BW	b	5383	31978	85,59%	BW	b	5383	31978	85,59%

VI.2. Manipulation B-SLC, la classification par sélection d'attributs

Les résultats de la sélection des attributs pour le niveau espèce et les performances de classification par les différentes combinaisons (attributs-classificateurs) sont obtenus dans cette manipulation.

Les caractéristiques sélectionnées varient selon la méthode utilisée et la précision a diminué par rapport à la première expérimentation. Le principe de la sélection est basé sur la corrélation entre les caractéristiques. De nombreux attributs étaient très corrélés avec d'autres, et si l'un de ces attributs était sélectionné le reste des attributs ne contribueront pas de manière significative dans la classification et ne seront pas pris en considération.

VI.2.1 Résultats de la sélection des attributs

Le tableau 16 montre les résultats de la sélection des attributs par chaque méthode qu'on a exploité dans la plateforme WEKA®, on a aussi manuellement sélectionné les attributs extraits par SmartGrain ® de ceux qu'on a calculé et on les a investigués séparément.

Sur le tableau 16 les attributs surlignés en toutes couleurs et en gras sont les plus choisis à savoir : **PL, L, LWR, CS, SF2, SF7, SF9**, ils ont été choisis au moins deux fois par une des méthodes d'évaluation. Les autres surlignés en rouge, ont été sélectionnés au plus une seule fois ou non parmi les 06 premiers attributs (dans le cas des méthodes recherche *Attribute ranking*).

Tableau 16: Résultats de la Sélection des attributs (SLC)

Méthode d'évaluation	Attributs sélectionnés					
<i>Des attributs de SmatGrain</i>	1. AS	2. PL	3. L	4. W	5. LWR	6. CS
<i>Des attributs Facteurs de forme calculés</i>	7. SF1	8. SF2	9. SF3	10. SF4	11. SF5	12. SF6
	13. SF7	14. SF8	15. SF9	16. SF10		
<i>CfsSubsetEval</i>	2. PL	3. L	5. LWR	8. SF2	13. SF7	15. SF9
<i>ClassifierAttributeEval</i>	13. SF7	3. L	2. PL	5. LWR	15. SF9	6. CS
<i>InfoGainAttributeEval</i>	13. SF7	3. L	5. LWR	15. SF9	6. CS	2. PL
<i>OneRAttributeEval</i>	13. SF7	3. L	2. PL	5. LWR	6. CS	15. SF9
<i>ReliefFAttributeEval</i>	11. SF5	6. CS	5. LWR	8. SF2	10. SF4	14. SF8

Les Méthodes d'évaluation *ClassifierAttributeEval*, *InfoGainAttributeEval* et *OneRAttributeEval* ont choisi les mêmes attributs à savoir : **PL, L, LWR, CS, SF7, SF9**.

Les facteurs de forme sont les moins choisis parmi tous les attributs, ceux de SmartGrain sont les plus figurants dans les différentes listes.

VI.2.2 Performances des classificateurs (SLC)

Le tableau 17 visualise les valeurs de précision pour chaque classificateur pour chaque classe et chaque groupe d'attributs. Les valeurs de précisions de la Manipulation B sont plus au moins proches à l'exception des valeurs de la classification avec les attributs sélectionnés par *ReliefFAttributeEval*.

Tableau 17: Visualisation des Performance de chaque classificateur en fonction des attributs et des classes

	SLC	MLPClassifier			MultilayerPerceptron			RandomForest		
		Wheat	HW	BW	Wheat	HW	BW	Wheat	HW	BW
	<i>MnpA (all attributs)</i>	87,24%	88,08%	86,27%	87,27%	88,00%	86,43%	86,76%	87,77%	85,59%
<i>MnpB</i>	<i>SmartGrain</i>	87,16%	87,99%	86,20%	87,16%	87,54%	86,71%	86,71%	87,68%	85,59%
	<i>ShapeFactor</i>	87,22%	88,03%	86,28%	87,19%	87,59%	86,73%	86,67%	87,66%	85,52%
	<i>2-3-5-8-13-15</i>	86,95%	87,73%	86,04%	87,08%	87,22%	86,92%	86,76%	87,75%	85,60%
	<i>13-3-2-5-6-15</i>	87,15%	87,98%	86,18%	87,18%	87,77%	86,51%	86,68%	87,79%	85,38%
	<i>11-6-5-8-10-14</i>	82,77%	86,25%	78,75%	82,75%	86,41%	78,51%	81,30%	84,71%	77,36%

Dans le tableau 18, une tendance globale des classificateurs à mieux classer le blé dur que le blé tendre ; le Classificateur *functions.MPLClassifier* a tendance à mieux classer le Blé dur, tandis que pour le blé tendre ce sont le classificateur *functions.MultilayerPerceptron* puis le *tree.RandomForest*.

Les matrices de confusion donnent plus de détails pour chaque combinaison (attribut/classificateur).

Tableau 18: Visualisation des Performance de chaque classificateur en fonction des classes

	SLC	MLPClassifier			MultilayerPerceptron			RandomForest		
		Wheat	HW	BW	Wheat	HW	BW	Wheat	HW	BW
	<i>MnpA (all attributs)</i>	87,24%	88,08%	86,27%	87,27%	88,00%	86,43%	86,76%	87,77%	85,59%
<i>Mnp2</i>	<i>SmartGrain</i>	87,16%	87,99%	86,20%	87,16%	87,54%	86,71%	86,71%	87,68%	85,59%
	<i>ShapeFactor</i>	87,22%	88,03%	86,28%	87,19%	87,59%	86,73%	86,67%	87,66%	85,52%
	<i>2-3-5-8-13-15</i>	86,95%	87,73%	86,04%	87,08%	87,22%	86,92%	86,76%	87,75%	85,60%
	<i>13-3-2-5-6-15</i>	87,15%	87,98%	86,18%	87,18%	87,77%	86,51%	86,68%	87,79%	85,38%
	<i>11-6-5-8-10-14</i>	82,77%	86,25%	78,75%	82,75%	86,41%	78,51%	81,30%	84,71%	77,36%

VI.2.3 Matrices de confusion de chaque classification

Les tableaux 19-23 regroupent les matrices de confusion de chaque classificateur obtenu en utilisant chaque groupe d'attributs sélectionnés ; les matrices de confusions montrent en détail les erreurs de classement des grains.

Tableau 19: matrices de confusion avec les attributs de Smart Grain

functions.MLPClassifier				SmartGrain			
=== Confusion Matrix ===				87,16%			
Classés		a	b	Précision			
HW	a	38049	5192	87,99%			
BW	b	5157	32204	86,20%			

trees.RandomForest				SmartGrain			
=== Confusion Matrix ===				86,71%			
Classés		a	b	Précision			
HW	a	37914	5327	87,68%			
BW	b	5383	31978	85,59%			

functions.MultilayerPerceptron				SmartGrain			
=== Confusion Matrix ===				87,16%			
Classés		a	b	Précision			
HW	a	37854	5387	87,54%			
BW	b	4965	32396	86,71%			

Le tableau 20 montre les Matrices de confusion de la classification obtenues en utilisant les *facteurs de forme* calculés ; le classificateur *Functions.MLPClassifier* affiche une précision meilleure que les autres classificateurs,

Tableau 20: Matrices de confusion « Facteurs de formes calculés »

functions.MLPClassifier				ShapeFactor			
=== Confusion Matrix ===				87,22%			
Classés		a	b	Précision			
HW	a	38066	5175	88,03%			
BW	b	5126	32235	86,28%			

trees.RandomForest				ShapeFactor			
=== Confusion Matrix ===				86,67%			
Classés		a	b	Précision			
HW	a	37907	5334	87,66%			
BW	b	5410	31951	85,52%			

functions.MultilayerPerceptron				ShapeFactor			
=== Confusion Matrix ===				87,19%			
Classés		a	b	Précision			
HW	a	37874	5367	87,59%			
BW	b	4957	32404	86,73%			

Le tableau 21 montre les Matrices de confusion de la classification en utilisant les attributs sélectionnés par la méthode d'évaluation « *CfsSubsetEval* » en conjonction avec *GreedyStepwise*

Tableau 21 : Matrices de confusion « CfsSubsetEval »

functions.MLPClassifier				<i>CfsSubsetEval</i>	
=== Confusion Matrix ===				86,95%	
Classés	a	b	Précision		
HW	a	37936	5305	87,73%	
BW	b	5214	32147	86,04%	

trees.RandomForest				<i>CfsSubsetEval</i>	
=== Confusion Matrix ===				86,76%	
Classés	a	b	Précision		
HW	a	37946	5295	87,75%	
BW	b	5379	31982	85,60%	

functions.MultilayerPerceptron				<i>CfsSubsetEval</i>	
=== Confusion Matrix ===				87,08%	
Classés	a	b	Précision		
HW	a	37716	5525	87,22%	
BW	b	4888	32473	86,92%	

Le tableau 22 montre les Matrices de confusion de la classification en utilisant les attributs sélectionnés par la méthode d'évaluation « *ClassifierAttributeEval*, *InfoGainAttributeEval* et *OneRAttributeEval* » en conjonction avec la méthode de recherche « Attribute ranking » ; ces derniers ont sélectionné les mêmes attributs, mais classés en ordre différents.

Tableau 22: Les matrices de confusion des ClassifierAttributeEval, InfoGainAttributeEval et OneRAttributeEval

functions.MLPClassifier				<i>OneRAttributeEval</i>	
=== Confusion Matrix ===				87,15%	
Classés	a	b	Précision		
HW	a	38045	5196	87,98%	
BW	b	5163	32198	86,18%	

trees.RandomForest				<i>OneRAttributeEval</i>	
=== Confusion Matrix ===				86,68%	
Classés	a	b	Précision		
HW	a	37962	5279	87,79%	
BW	b	5461	31900	85,38%	

functions.MultilayerPerceptron				<i>OneRAttributeEval</i>	
=== Confusion Matrix ===				87,18%	
Classés	a	b	Précision		
HW	a	37952	5289	87,77%	
BW	b	5041	32320	86,51%	

Les attributs sélectionnés par la méthode *ReliefAttributeEval*, ont donné une faible précision par rapport aux autres combinaisons (tableau 23).

Tableau 23: Les matrices de confusion de ReleFAttributeEval

functions.MLPClassifier				ReliefAttril	
=== Confusion Matrix ===				82,77%	
Classés		a	b	Précision	
HW	a	37295	5946	86,25%	
BW	b	7941	29420	78,75%	

trees.RandomForest				ReliefAttributeEval	
=== Confusion Matrix ===				81,30%	
Classés		a	b	Précision	
HW	a	36629	6612	84,71%	
BW	b	8457	28904	77,36%	

functions.MultilayerPerceptron				ReliefAttrib	
=== Confusion Matrix ===				82,75%	
Classés		a	b	Précision	
HW	a	37366	5875	86,41%	
BW	b	8028	29333	78,51%	

VII. La classification en niveau « Variété »

Le tableau 24 regroupe les performances de chaque classificateur par chaque groupe d'attributs sélectionnés. En référence avec la classification en niveau espèce la précision de classification a varié considérablement entre le niveau espèce (87%) et le niveau variété (~60).

Tableau 24: Résultats de la classification en niveau variété

VLC	functions.SMO	functions.MLPClassifier	functions.MultilayerPerceptron
<i>All attributs</i>	56,70%	59,46%	59,44%
<i>SmartGrain</i>	56,06%	58,63%	58,39%
<i>ShapeFactor</i>	56,29%	58,59%	58,95%
<i>2-3-5-8-13-15</i>	55,90%	58,38%	58,16%
<i>13-3-2-5-6-15</i>	55,78%	58,51%	58,41%
<i>11-6-5-8-10-14</i>	47,04%	52,16%	52,36%
<i>CfsSubsetEval</i>	56,37%	59,39%	58,63%
<i>ClassifierAttributeEval</i>	55,95%	59,12%	58,36%
<i>InfoGainAttributeEval</i>	55,78%	59,24%	58,41%
<i>OneRAttributeEval</i>	55,88%	59,10%	58,27%
<i>ReliefFAttributeEval</i>	56,19%	59,38%	58,60%

VII.1. Manipulation A- VLC, sans sélection d'attributs

VII.1.1 Performances et Pré -Choix des classificateurs :

Le tableau 25 représente les résultats de l'expérimentation tenue sur la plateforme WEKA® pour comparer et pré-choisir entre les différents classificateurs selon la configuration (**08 classifieurs, 05 Folds, 05 répétitions**), les résultats montrent les différences statistiquement significative (à 5%).

Sur la plateforme WEKA®, pour le cas du « niveau Variété », on a utilisé la méthode « *The ranking test* » qui classe le classificateur en fonction de ses gains et ses pertes par rapport aux autres classificateurs.

Chaque case du tableau 25 correspond au classificateur qui apparaît pour chaque paramètre et chaque position. Le choix a été effectué en fonction du nombre d'apparence dans les 03 premières places. Il s'agit des classificateurs :

functions.SMO, *functions.MultilayerPerceptron* et *functions.MLPClassifier*, on note que les classificateurs : *SGD* et le *ClassificationViaRegression* ne fonctionnaient pas en *multi-Nominale-classes* mode.

Tableau 25: Résultats de la comparaison entre les différents algorithmes de classification (VLC)

Rank	Percent_correct	True_positive_rate	False_positive_rate	IR_precision	IR_recall	F_measure
1	3	6	6	1	6	6
2	5	5	4	6	5	5
3	4	4	2	5	4	4
4	1	3	5	4	3	3
5	6	2	3	3	2	2
6	2	1	1	2	1	1

(1) bayes.NaiveBayes
 (2) meta.AttributeSelectedClassifier
 (3) trees.RandomForest
 (4) functions.MLPClassifier
 (5) functions.MultilayerPerceptron
 (6) functions.SMO '

VII.1.2 Optimisation d'hyperparamètres des classificateurs (VLC)

Les configurations des classificateurs retenues pour le niveau de classification variétale sont comme suit :

- ***functions.SMO:***

```
weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator
"weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
```

- ***functions.MultilayerPerceptron:***

```
weka.classifiers.functions.MultilayerPerceptron -L 0.03 -M 0.2 -N 700 -V 0 -S 0 -E 20 -H a
```

- ***Le.MLPClassifier :***

```
weka.classifiers.functions.MLPClassifier -N 6 -R 0.01 -O 1.0E-6 -P 1 -E 1 -S 1 -L
weka.classifiers.functions.loss.SquaredError -A weka.classifiers.functions.activation.Softplus
```

VII.1.3 Résultats détaillés de la classification

Les tableaux 26-31 montrent le détail de la classification de chaque classificateur avec la configuration optimisée, suivi par la matrice de confusion qui visualisent en détail les erreurs de classification et la confusion entre les variétés de blé.

Tableau 26: résultats détaillés par classe du classificateur *functions.SMO*

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Simeto	0,679	0,184	0,609	0,679	0,642	0,48	0,805	0,557
ARZ	0,687	0,144	0,599	0,687	0,64	0,519	0,843	0,543
Vitron	0,354	0,124	0,474	0,354	0,405	0,256	0,734	0,39
HD	0,52	0,129	0,54	0,52	0,53	0,396	0,777	0,446
Avg.	0,567	0,148	0,559	0,567	0,559	0,417	0,791	0,489

Le classificateur *functions.SMO*, classe mieux la variété Simeto et ARZ tandis que les variétés HD et Vitron sont moins classées. Une partie de la variété HD est classée en Vitron ce qui explique la chute de précision de classification par rapport au niveau espèce. Les mêmes observations pour les autres classificateurs.

Tableau 27: Matrice de confusion du classificateur *functions.SMO*

functions.SMO						
=== Confusion Matrix ===						56,70%
Classés	>>>	a	b	c	d	Precision
Simeto	a	16263	1417	4530	1752	68%
ARZ	b	619	13169	1141	4246	69%
Vitron	c	8889	1519	6818	2053	35%
HD	d	939	5885	1908	9454	52%

Les classificateurs *function.MLPClassifier* et *function.MultilayerPerceptron* classent mieux la variété ARZ, la HD et le Vitron, mais moins le Simeto que le classificateur *functions.SMO*.

Tableau 28: Résultats détaillés par classe du classificateur *function.MLPClassifier*

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Simeto	0,591	0,131	0,657	0,591	0,622	0,475	0,85	0,718
ARZ	0,685	0,112	0,656	0,685	0,67	0,564	0,897	0,73
Vitron	0,513	0,169	0,488	0,513	0,5	0,338	0,813	0,503
HD	0,59	0,128	0,573	0,59	0,582	0,457	0,858	0,628
Avg.	0,595	0,135	0,597	0,595	0,595	0,46	0,854	0,649

Tableau 29: Matrice de confusion du classificateur *function.MLPClassifier*

functions.MLPClassifier								
==== Confusion Matrix ====								59,46%
Classés	>>>	a	b	c	d	Precision		
Simeto	a	14158	1138	6806	1860	59%		
ARZ	b	364	13134	1411	4266	68%		
Vitron	c	6457	1054	9898	1870	51%		
HD	d	586	4697	2168	10735	59%		

Tableau 30: Résultats détaillés par classe du classificateur *function.MultilayerPerceptron*

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Simeto	0,574	0,122	0,666	0,574	0,617	0,474	0,851	0,721
ARZ	0,673	0,107	0,662	0,673	0,668	0,563	0,895	0,733
Vitron	0,538	0,179	0,485	0,538	0,51	0,346	0,812	0,504
HD	0,598	0,132	0,57	0,598	0,583	0,458	0,857	0,627
Avg.	0,594	0,134	0,6	0,594	0,596	0,461	0,854	0,651

Tableau 31: Matrice de confusion du classificateur *function.MultilayerPerceptron*

functions.MultilayerPerceptron								
==== Confusion Matrix ====								59,44%
Classés	>>>	a	b	c	d	Precision		
Simeto	a	13762	1089	7272	1839	57%		
ARZ	b	318	12908	1529	4420	67%		
Vitron	c	5947	1006	10363	1963	54%		
HD	d	626	4486	2197	10877	60%		

VII.2. Manipulation B- VLC

VII.2.1 Résultats de la Sélection des attributs (VLC)

Les attributs surlignés en toutes couleurs et en gras (tableau 32) sont les plus choisis à savoir : **PL, L, W, LWR, CS, SF2, SF7, SF9**, ils ont été choisis au moins deux fois par une des méthodes d'évaluation. On note la sélection de l'attribut **W** dans ce niveau (VLC).

Les autres attributs surlignés en rouge, ont été sélectionnés au plus une seule fois ou non parmi les 06 premiers attributs (dans le cas des méthodes recherche *Attribute ranking*).

Tableau 32 : Résultats de la Sélection des attributs (VLC)

VLC						
Méthode d'évaluation	Attributs sélectionnés					
<i>Des attributs de SmatGrain</i>	1. AS	2. PL	3. L	4. W	5. LWR	6. CS
<i>Des attributs Facteurs de forme calculés</i>	7. SF1	8. SF2	9. SF3	10. SF4	11. SF5	12. SF6
	13. SF7	14. SF8	15. SF9	16. SF10		
<i>CfsSubsetEval</i>	2. PL	3. L	5. LWR	6. CS	8. SF2	13. SF7
<i>ClassifierAttributeEval</i>	16. SF10	15. SF9	6. CS	5. LWR	4. W	3. L
<i>InfoGainAttributeEval</i>	13. SF7	3. L	2. PL	5. LWR	6. CS	15. SF9
<i>OneRAttributeEval</i>	3. L	4. W	2. PL	13. SF7	5. LWR	1. AS
<i>ReliefFAttributeEval</i>	11. SF5	6. CS	10. SF4	5. LWR	14. SF8	12. SF6

VII.2.2 Performances des classifications en (VLC)

Les tableaux 33 et 34 regroupent les résultats de la classification VLC, ces deux tableaux visualisent, par nuance de couleurs pondérées, la variation des valeurs de précision par rapport à la valeur inférieure et supérieure enregistrées pour toutes les catégories d'attributs, classes et classificateurs.

Tableau 33 : visualisation des résultats obtenues de la classifications VLC en intera-catégorie

	functions.SMO					functions.MLPClassifier					functions.MultilayerPerceptron				
	Avr	Simeto	Vitron	ARZ	HD	Avr	Simeto	Vitron	ARZ	HD	Avr	Simeto	Vitron	ARZ	HD
<i>All attributs</i>	56,70%	68%	35%	69%	52%	59,46%	59%	51%	68%	59%	59,44%	57%	54%	67%	60%
<i>SmartGrain</i>	56,06%	70%	31%	71%	48%	58,63%	59%	50%	69%	56%	58,39%	58%	51%	70%	55%
<i>ShapeFactor</i>	56,29%	70%	31%	69%	52%	58,59%	58%	51%	71%	55%	58,95%	59%	51%	67%	59%
<i>CfsSubsetEval</i>	56,37%	70%	31%	68%	53%	59,39%	59%	51%	68%	60%	58,63%	57%	51%	68%	58%
<i>ClassifierAttributeEval</i>	55,95%	71%	29%	69%	50%	59,12%	58%	52%	70%	58%	58,36%	57%	51%	70%	56%
<i>InfoGainAttributeEval</i>	55,78%	70%	31%	69%	50%	59,24%	59%	50%	69%	59%	58,41%	57%	51%	68%	58%
<i>OneRAttributeEval</i>	55,88%	67%	35%	73%	47%	59,10%	58%	52%	68%	58%	58,27%	54%	55%	69%	56%

Tableau 34: visualisation intra classe de la précision de chaque classificateur

	functions.SMO					functions.MLPClassifier					functions.MultilayerPerceptron				
	Avr	Simeto	Vitron	ARZ	HD	Avr	Simeto	Vitron	ARZ	HD	Avr	Simeto	Vitron	ARZ	HD
<i>All attributs</i>	56,70%	68%	35%	69%	52%	59,46%	59%	51%	68%	59%	59,44%	57%	54%	67%	60%
<i>SmartGrain</i>	56,06%	70%	31%	71%	48%	58,63%	59%	50%	69%	56%	58,39%	58%	51%	70%	55%
<i>ShapeFactor</i>	56,29%	70%	31%	69%	52%	58,59%	58%	51%	71%	55%	58,95%	59%	51%	67%	59%
<i>CfsSubsetEval</i>	56,37%	70%	31%	68%	53%	59,39%	59%	51%	68%	60%	58,63%	57%	51%	68%	58%
<i>ClassifierAttributeEval</i>	55,95%	71%	29%	69%	50%	59,12%	58%	52%	70%	58%	58,36%	57%	51%	70%	56%
<i>InfoGainAttributeEval</i>	55,78%	70%	31%	69%	50%	59,24%	59%	50%	69%	59%	58,41%	57%	51%	68%	58%
<i>OneRAttributeEval</i>	55,88%	67%	35%	73%	47%	59,10%	58%	52%	68%	58%	58,27%	54%	55%	69%	56%

VII.2.3 Matrices de confusion issues des différentes combinaisons d'attributs

Les tableaux 35-38, regroupent les matrices de confusion de chaque classificateur obtenu en utilisant chaque catégorie d'attribut sélectionnés ; ces matrices montrent en détail les erreurs de classement des grains.

Tableau 35: Les matrices de confusion issues de la classification par les attributs de SmartGrain

functions.SMO							SmartGrain						
=== Confusion Matrix ===							56,06%						
Classés	>>>	a	b	c	d	Precision	Classés	>>>	a	b	c	d	Precision
Simeto	a	16744	1600	3959	1659	70%	Simeto	a	14095	1157	6671	2039	59%
ARZ	b	584	13639	1244	3708	71%	ARZ	b	322	13324	1499	4030	69%
Vitron	c	9501	1719	6036	2023	31%	Vitron	c	6456	1027	9651	2145	50%
HD	d	1049	6646	1727	8764	48%	HD	d	597	5231	2171	10187	56%

functions.MultilayerPerceptron							SmartGrain						
=== Confusion Matrix ===							58,39%						
Classés	>>>	a	b	c	d	Precision	Classés	>>>	a	b	c	d	Precision
Simeto	a	13950	1176	6894	1942	58%	Simeto	a	13950	1176	6894	1942	58%
ARZ	b	348	13329	1484	4014	70%	ARZ	b	348	13329	1484	4014	70%
Vitron	c	6325	1133	9776	2045	51%	Vitron	c	6325	1133	9776	2045	51%
HD	d	652	5365	2164	10005	55%	HD	d	652	5365	2164	10005	55%

Tableau 36: Les matrices de confusion issues de la classification par les attributs de formes calculés

functions.SMO							ShapeFactor						
=== Confusion Matrix ===							56,29%						
Classés	>>>	a	b	c	d	Precision	Classés	>>>	a	b	c	d	Precision
Simeto	a	16843	1425	3991	1703	70%	Simeto	a	13794	1165	7138	1865	58%
ARZ	b	645	13137	1079	4314	69%	ARZ	b	336	13550	1423	3866	71%
Vitron	c	9710	1563	5933	2073	31%	Vitron	c	6312	1084	9810	2073	51%
HD	d	980	5872	1877	9457	52%	HD	d	526	5326	2262	10072	55%

functions.MultilayerPerceptron							ShapeFactor						
=== Confusion Matrix ===							58,95%						
Classés	>>>	a	b	c	d	Precision	Classés	>>>	a	b	c	d	Precision
Simeto	a	14118	1103	6741	2000	59%	Simeto	a	14118	1103	6741	2000	59%
ARZ	b	442	12875	1388	4470	67%	ARZ	b	442	12875	1388	4470	67%
Vitron	c	6370	1021	9763	2125	51%	Vitron	c	6370	1021	9763	2125	51%
HD	d	817	4621	1987	10761	59%	HD	d	817	4621	1987	10761	59%

Tableau 37: Les matrices de confusion issues de la classification par les attributs sélectionnés par CfsSubsetEval

functions.SMO							<i>CfsSubsetEval</i>	
=== Confusion Matrix ===							56,37%	
Classés	>>>	a	b	c	d	Precision		
Simeto	a	16838	1384	3955	1785	70%		
ARZ	b	543	13002	1255	4375	68%		
Vitron	c	9686	1495	6011	2087	31%		
HD	d	1125	5743	1735	9583	53%		

functions.MLPClassifier							<i>CfsSubsetEval</i>	
=== Confusion Matrix ===							59,39%	
Classés	>>>	a	b	c	d	Precision		
Simeto	a	14059	1136	6815	1952	59%		
ARZ	b	345	13124	1404	4302	68%		
Vitron	c	6365	1052	9858	2004	51%		
HD	d	584	4663	2114	10825	60%		

functions.MultilayerPerceptron							<i>CfsSubsetEval</i>	
=== Confusion Matrix ===							58,63%	
Classés	>>>	a	b	c	d	Precision		
Simeto	a	13760	1142	7066	1994	57%		
ARZ	b	334	13106	1424	4311	68%		
Vitron	c	6274	1074	9792	2139	51%		
HD	d	553	4784	2252	10597	58%		

Tableau 38: Les matrices de confusion issues de la classification par les attributs sélectionnés par ReliefFAttributeEval

functions.SMO							<i>ClassifierAttributeEval</i>	
=== Confusion Matrix ===							55,95%	
Classés	>>>	a	b	c	d	Precision		
Simeto	a	17036	1525	3693	1708	71%		
ARZ	b	578	13302	1235	4060	69%		
Vitron	c	9901	1655	5620	2103	29%		
HD	d	1181	6256	1608	9141	50%		

functions.MLPClassifier							<i>ClassifierAttributeEval</i>	
=== Confusion Matrix ===							59,12%	
Classés	>>>	a	b	c	d	Precision		
Simeto	a	13791	1273	7017	1881	58%		
ARZ	b	299	13363	1411	4102	70%		
Vitron	c	6228	1178	9941	1932	52%		
HD	d	613	4928	2089	10556	58%		

functions.MultilayerPerceptron							<i>ClassifierAttributeEval</i>	
=== Confusion Matrix ===							58,36%	
Classés	>>>	a	b	c	d	Precision		
Simeto	a	13735	1277	7016	1934	57%		
ARZ	b	305	13393	1387	4090	70%		
Vitron	c	6156	1184	9797	2142	51%		
HD	d	539	5345	2191	10111	56%		

VIII. Discussion

La classification des grains de blé impliquant l'ensemble des attributs morphologiques nous a permis d'évaluer la performance de chaque classificateur et de faire sortir les meilleurs classificateurs adéquats à chaque niveau de classification (tableaux 7-11 et 25-31).

En matière d'attribut, plus on a d'attributs plus la précision d'un classificateur donné est meilleure (Tableaux 7, 17, 24). Cependant la sélection des attributs nous a permis de tester séparément chaque combinaison d'attributs sélectionnés, ce qui a permis d'évaluer à la fois chaque classificateur et d'évaluer aussi l'impact de chaque catégorie d'attributs dans la classification (Tableaux 17, 24 et 33). D'une manière générale, sur le plan taille des grains représentée par la longueur et la largeur ces derniers sont des attributs importants, la surface et le périmètre ajoute un poids en plus pour améliorer la précision. Quant à la forme, elle est importante aussi vu que les facteurs de forme donnent une précision proche aux deux catégories d'attributs expérimentés.

En comparant les résultats des attributs obtenus par smart grain avec ceux calculés on constate qu'il y'a une légère différence, même si on les compare aux résultats obtenus avec l'ensemble des attributs, cette dernière est mieux sur le plan précision. Ceci démontre que chaque information est importante pour l'identification des grains, le poids (l'impact) de chaque attribut est différent comme le montre leur classement dans les tableaux (7 et 24).

D'un point de vue biologique la morphologie des grains peut varier considérablement en largeur et en surface qu'en longueur, cette dernière semble être plus stable et homogène. A la fin de la phase de remplissage des grains les dimensions des grains doivent avoir leurs valeurs maximales, puis en phase de maturation et séchage les dimensions axiales vont légèrement diminuer, mais la variation pour la longueur est faible par rapport à la largeur. C'est ce que nous avons constaté dans une étude sur la variabilité morphologique des grains .

Les résultats montrent aussi que les caractéristiques morphologiques expérimentées ne sont pas en mesure de déterminer l'identité du grain avec une précision acceptable (+95%). Ceci peut être expliqué par une insuffisance d'ordre qualitatif et quantitatif en information. L'insuffisance qualitative est liée d'un côté à l'étape de l'extraction des attributs par le SmartGrain (segmentation) et d'un autre côté à la nature des attributs utilisés, cette dernière est

à l'origine de l'insuffisance quantitative pour les attributs morphologiques, il fallait ajouter l'épaisseur du grain et d'autres facteurs de forme comme les angles et les courbures des extrémités des grains. Il fallait aussi explorer d'autres catégories d'attributs comme la couleur la texture ou d'autres bandes du spectre (visible et invisible).

Nos résultats sont loin des valeurs de certains travaux qui ont eu des précisions entre 90 et 100%, sur la base des attributs morphologiques, ceci peut être expliqué par la quantité de données impliquées. **Zapotoczny (2011)** a utilisé une base de données de 6500 grains, qui est plus faible que la nôtre et il a obtenu une précision de 98-100%, ce qui semble un problème de surapprentissage (overfitting). Tandis que **S. Majumdar et D. S. Jayas 2000** qui ont utilisé une base de données de 40 000 grains ils ont obtenu 98 et 97 % pour deux classes de blé *Western Red Spring et Western Amber Durum*, ceci peut être expliqué par l'homogénéité intra-classe et du fait qu'ils ont utilisé deux variétés distinctes. Pour le niveau variétale, nos résultats sont plus faibles que ceux obtenus par (Khoshroo and Arefi, 2014), ils ont obtenu 85.72% mais il ont utilisé une faible base de données (500 grains) et/ou les variétés sont plus ou moins distinctes ce qui a rendu la classification faisable. Dans notre cas il peut s'agir aussi d'un problème d'apprentissage faible (Underfitting) et des problèmes de paramétrage des classificateurs.

Sur le plan agronomique la variété ARZ est facilement classée par rapport aux autres variétés, ce qui indique que son modèle morphologique est plus stable et moins sensible aux facteurs de changement qui peuvent l'affecter durant son développement ; cette idée est confirmée dans l'étude de la variabilité morphologique qu'on a menée pour l'élaboration d'une carte variétale de blé ; Le coefficient de variation et l'écart-type sont plus faibles pour cette variété.

D'une manière plus pragmatique, la classification Shallow par des attributs morphologiques n'est pas satisfaisante en matière de précision pour les deux niveaux de classification SLC et VLC. Mais les données collectées peuvent être utiles pour d'autres problématiques d'ordre quantitatif et qualitatif.

IX. Conclusion de l'expérimentation N°1

Dans cette expérimentation nous avons investigué la possibilité de reconnaître l'identité des grains de blé à la base de leurs caractéristiques morphologiques à savoir leurs tailles et formes ; ces dernières ont été extraites par Computer Vision. C'est pour cette raison qu'on l'a appelé Shallow approche, car on a suivi la méthode classique de la Computer Vision qui consiste en ces étapes : l'acquisition d'image, l'extraction des attributs (the handcraft features extraction) et la classification par ML.

L'identification concerne des grains de quatre variétés de blé VITRON et SIMETO pour l'espèce blé dur, ARZ et HD pour l'espèce blé tendre. L'expérimentation a été réalisée sur la puissante et riche plateforme de Datamining Weka, qui nous a permis de déployer notre méthodologie pour déterminer le classificateur et les attributs les plus discriminants.

Parmi plusieurs algorithmes, quatre classificateurs ont été choisis : SupportVectorMachine (SMO) , RandomForest (RF), MLPClassifier et MultilayerPerception, ces deux derniers ont donné la meilleure précision (> 87%) pour le niveau espèce avec toutes les caractéristiques morphologiques, la configuration de test était fixée à : 05 plis (folds) en validation croisée (cross validation).

À travers cette expérimentation, Nous avons constaté et confirmé que la classification du niveau « Variété » est plus difficile que celle du niveau « Espèce » ; cette difficulté s'explique bien évidemment par le degré de rapprochement (ressemblance) taxonomique inter-espèce (Blé dur-Blé tendre), qui est plus forte en inter-variété.

Les classificateurs ont des comportements variables aussi au même niveau de classification, dont ils classaient des variétés mieux que les autres, qui est expliqué par la variabilité intra-variété ; cette dernière avait, en plus du nombre d'attributs, un impact sur la précision des classificateurs, qui n'ont pas permis d'atteindre les résultats escomptés ; même si le niveau espèce est encourageant, le niveau variétal reste un vrai challenge pour nous, il est question de chercher d'autres alternatives et approches :

- Utiliser plus d'attributs de nature différente couleur, texture
- Utiliser un autre type de signal, spectre, infra-rouge ... n'est pas disponible.
- Utiliser une autre approche de classification d'images par ML en gardant les mêmes images (RGB), moins chère.

Chapitre 3: EXPÉRIIMENTATION N°2

Deep Learning approches

I. Introduction

La deuxième expérimentation représente l'implémentation de l'approche Deep Learning basée sur l'entraînement des réseaux de neurones convolutionnels profonds (CNNs). L'objectif ultime est de réaliser une classification à l'échelle variétale avec une bonne précision qui doit être égale ou supérieure à 95%.

Le protocole expérimental est représenté graphiquement sur la figure 42. La phase de Deep classification est effectuée par l'utilisation de la base de données « *single grain images* ».

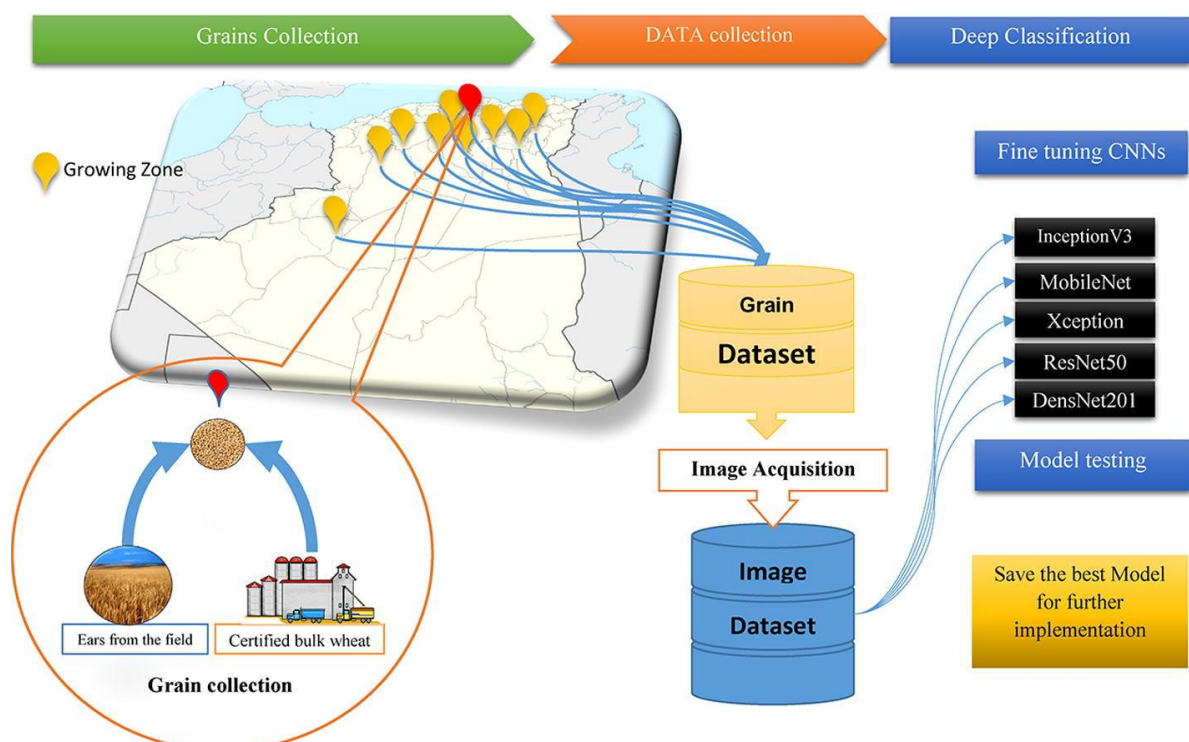


Figure 42: Représentation graphique du protocole de l'expérimentation *Deep Learning Classification*

Une partie de cette expérimentation a fait l'objet d'une publication scientifique (Laabassi *et al.*, 2021). A noter que cette expérimentation de Deep Learning a été effectuée à plusieurs reprises sur les trois années (2017-2019). Comme indiqué dans la figure 43, pour la première année 2017 on a formé et testé les modèles avec la base de données 2017. Quant aux deux autres années nous avons utilisé la base de données de l'année précédente associée à celle l'année en cours (2017-2018 ; 2018-2019). En définitif nous avons construit et testé les meilleurs Modèles CNNs sur l'ensemble de la base de données des 03 années (2017-2018-2019).

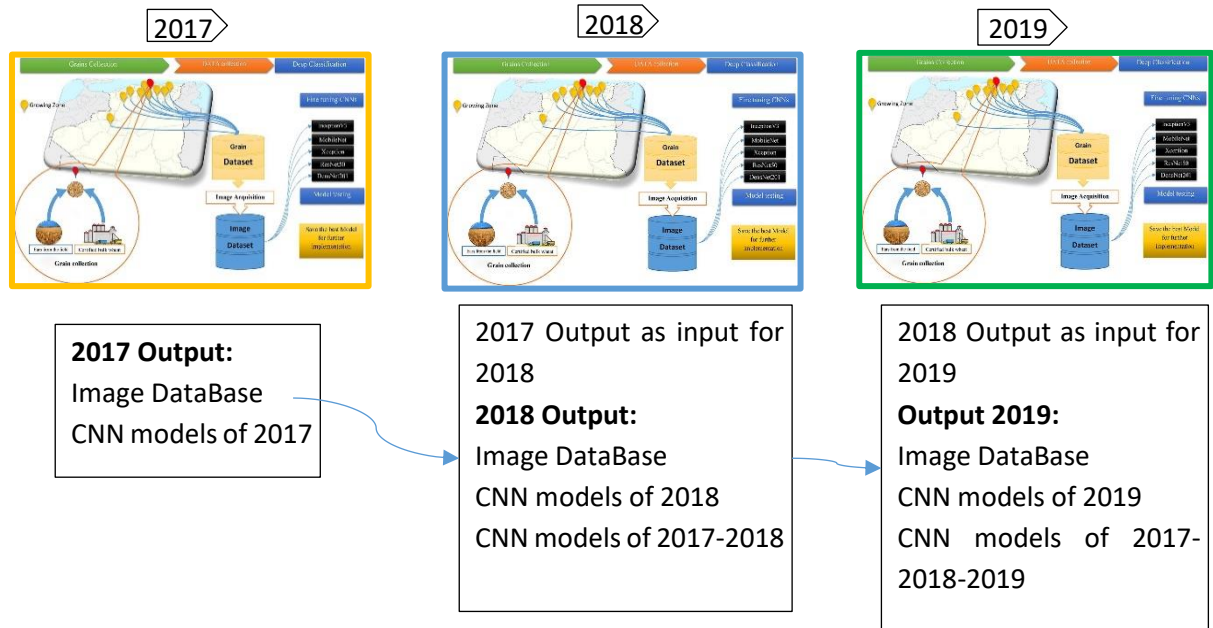


Figure 43: Progression du protocole expérimental et l'enchaînement de l'apprentissage sur 03 années.

II. La base de données Images

Nous avons utilisé la base de données des *images de grain individuel* préparée pour cette expérimentation de Deep Learning (chapitre1 partie expérimentale). Les performances des CNNs (Figure 44) s'améliorent avec l'augmentation de la taille de la base de données (Alom *et al.*, 2019).

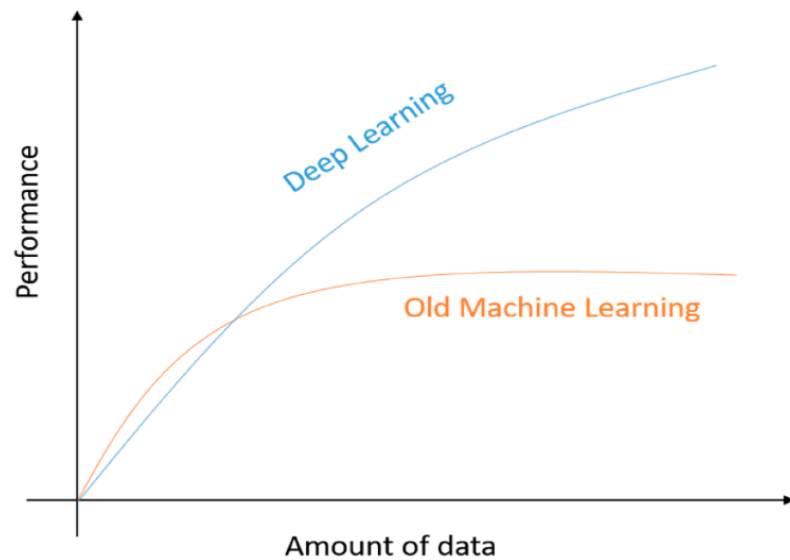


Figure 44: Evolution des performances des CNN en fonction de la quantité des données (Alom *et al.*, 2019)

A noter qu'aucune opération de traitement spécifique en vue d'amélioration ou de transformation ou d'augmentation (preprocessing and data augmentation) sur les images n'a été réalisée. A rappeler que nous avons limité au maximum les opérations supplémentaires, ainsi que nous avons laissé les données dans leur état brut (plus proche de la réalité de terrain).

III. Classificateur Deep Learning

Les réseaux de neurones convolutifs (ConvNets ou CNN) sont un type spécialisé de réseaux de neurones profonds (DNN) conçus pour traiter des données multidimensionnelles tels que des signaux (1D), des images (2D) et des vidéos (3D) (LeCun, Bengio et Hinton, 2015 ; LeCun, Bottou, Bengio et Haffner, 1998).

Les *ConvNets* contiennent au minimum deux à plusieurs centaines de *couches de convolutions* avec des *fonctions d'activation* non linéaires, telles que *Sigmoid, Tanh et ReLU* (Unités linéaires rectifiées), appliquées aux résultats.

Un ensemble différent de *filtres de convolution* est appliqué sur chaque couche, puis la sortie des convolutions est combinée pour maintenir la connectivité locale entre les neurones des couches adjacentes. Cette connectivité locale permet à chaque neurone d'être connectée uniquement à un petit sous-ensemble local de l'image donnée, ce qui contribue à réduire le nombre de paramètres dans l'ensemble du réseau ainsi qu'à rendre le calcul plus efficace (Chen, Wu, Fan, Sun, & Naoï, 2014).

Une structure ConvNet à couches aussi profondes permet au réseau d'apprendre automatiquement les meilleurs attributs pendant le processus de formation et, dans la plupart des cas, surpassera les extracteurs de attributs manuels (handcraft features) qui nécessitent généralement une compétence et des connaissances approfondies en ingénierie (Hu, Xia, Hu, & Zhang, 2015 ; LeCun et al., 2015).

Les figures (45-49) illustrent les diagrammes des Cinq architectures de CNNs qui ont été expérimentées, à savoir : InceptionV3, MobileNet, Xception, ResNet50, and DensNet201. (Mahdianpari *et al.*, 2018; Alom *et al.*, 2019)

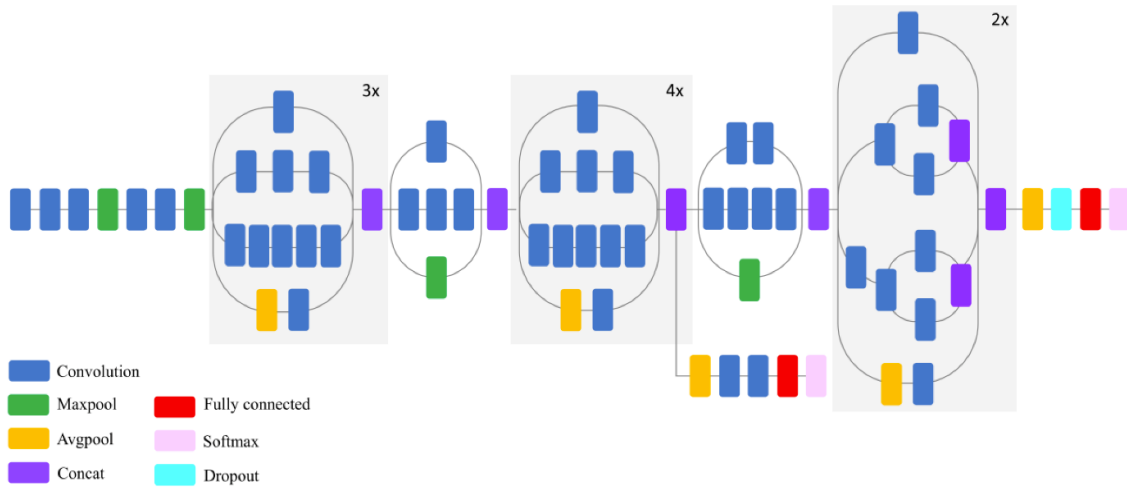


Figure 45: Diagramme schématique du modèle InceptionV3 (Mahdianpari *et al.*, 2018)

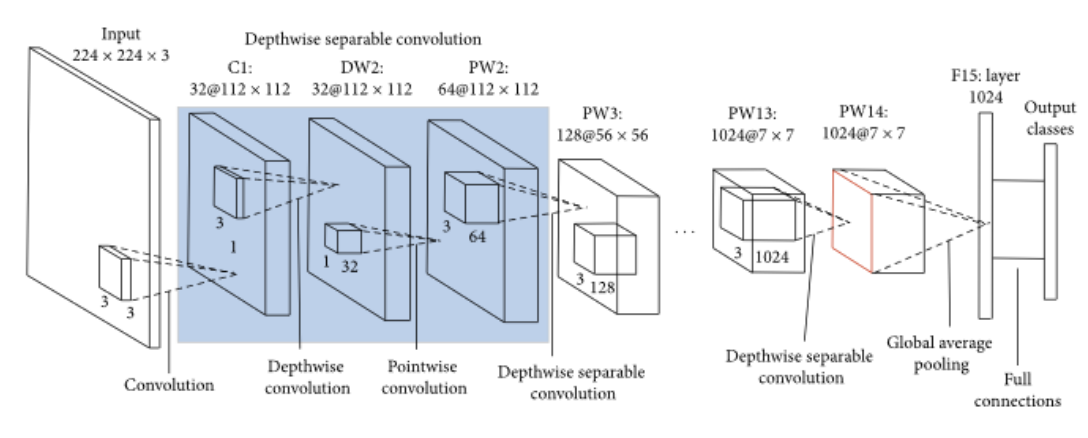


Figure 46 Diagramme schématique du modèle MobileNet (Howard *et al.*, 2017).

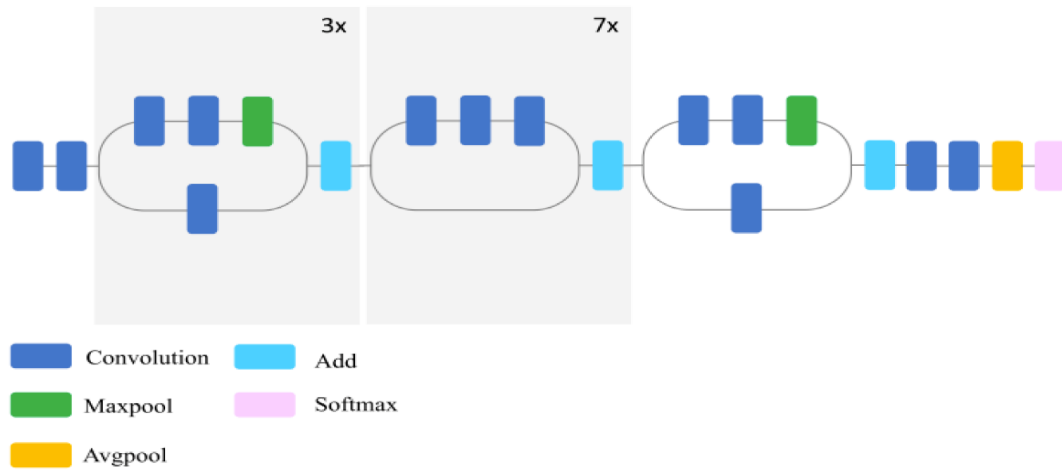


Figure 47: Diagramme schématique du modèle Xception (Mahdianpari *et al.*, 2018).

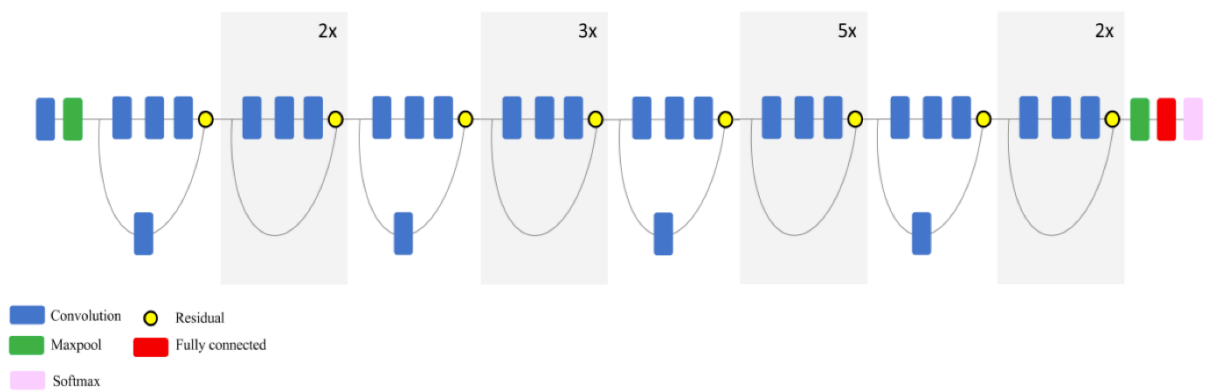


Figure 48: Diagramme schématique du modèle ResNet (Mahdianpari *et al.*, 2018).

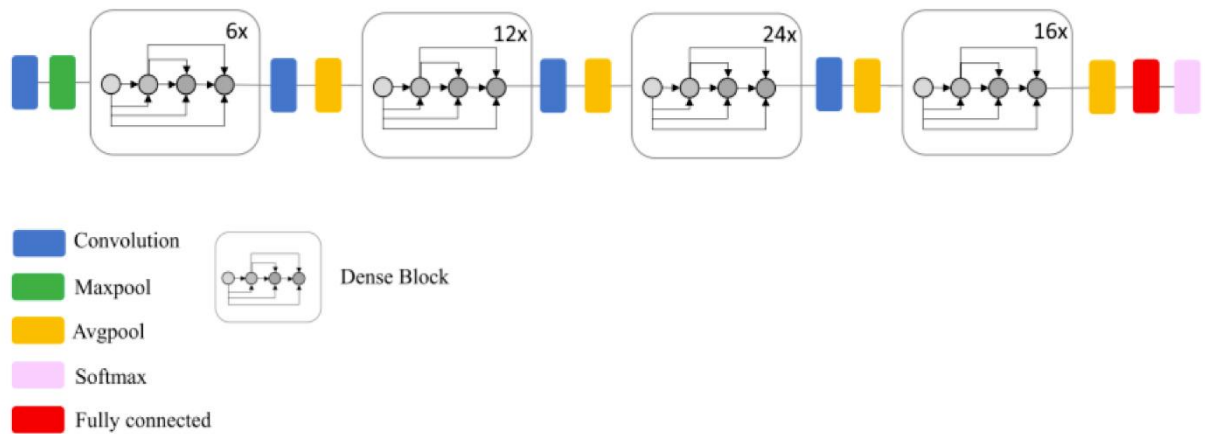


Figure 49: Diagramme schématique du modèle DenseNet (Mahdianpari *et al.*, 2018).

IV. La technique de « Transfert Learning »

Le Transfert Learning est un problème de recherche en ML traitant le *savoir acquis* par une machine sur un problème donné (task A) et l'application de ce savoir pour résoudre un autre problème (task B) de même nature (Figure 50).

Les architectures que nous avons utilisées à l'aide de la technique d'apprentissage par transfert (transfert Learning, fine-tuning) ont été pré-entraînées sur la base de données ImageNet⁵ qui contient des millions d'images et 1000 classes d'objets différents.

La stratégie d'apprentissage par transfert a été choisie car elle peut améliorer la précision tout en accélérant l'entraînement (Alom et al., 2019). L'apprentissage par transfert peut offrir des avantages importants pour l'identification automatisée et peut améliorer les modèles de classification à faible performance (Kaya et al., 2019).

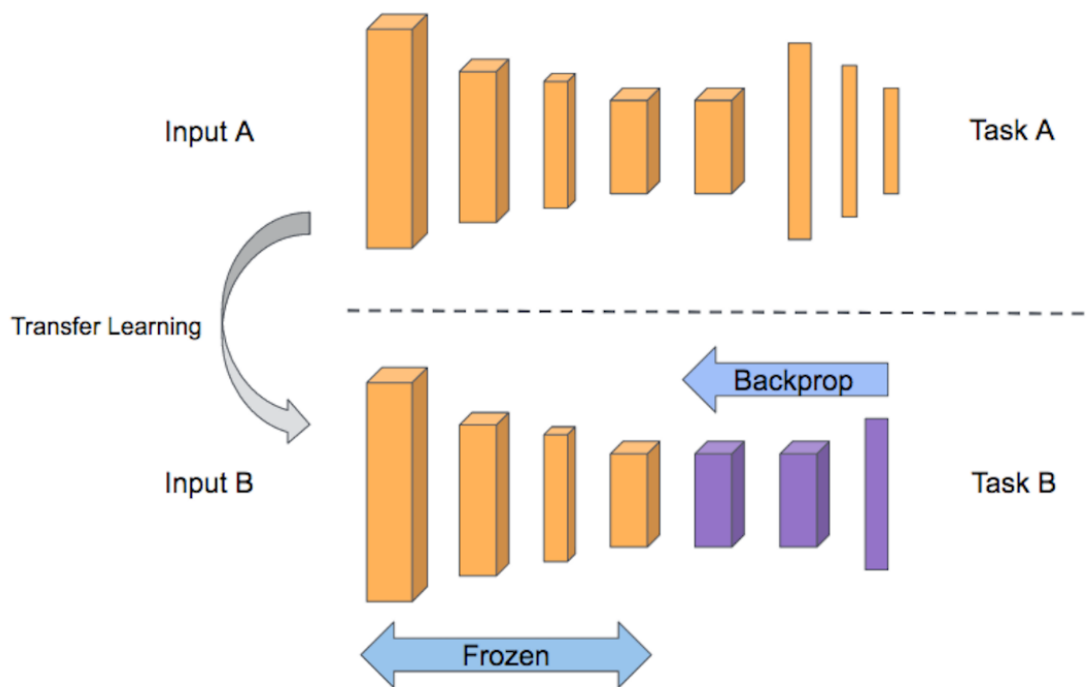


Figure 50: principe de la technique de Transfert Learning

<https://www.image-net.org>⁵

V. Structure des Modèles

V.1. Architecture

Pour construire le nouveau modèle, nous avons remplacé les couches supérieures d'origine des architectures standard par d'autres couches (figure 51). La partie restante de l'architecture est appelée « modèle de base ». Les couches de remplacement sont les suivantes :

1. Une couche *Average Pooling2D* qui réduit le risque de sur-apprentissage ;
2. Une couche *dense layer* de 1024 neurones est ajoutée avec fonction d'activation *ReLU* ;
3. La dernière couche de prédiction entièrement connectée avec la fonction d'activation *SoftMax*. Le nombre de neurones dans cette couche est de quatre selon le nombre de classes (Simeto, Vitron, ARZ, HD).

```
#Inception_V3 model
if(architecture_name == 'Inception_V3'):
    base_model= InceptionV3(weights='imagenet', include_top=False, input_shape=(299, 299, 3))
    x = base_model.output
    x = AveragePooling2D()(x)
    x = Dense(1024, activation='relu')(x)
    predictions = Dense(class_number, activation='softmax')(x)
    model= Model(inputs=base_model.input, outputs=predictions)
    print('ok')
```

Figure 51: Extrait du code de Transfert Learning pour l'architecture Inception V3

Concernant la deuxième partie de l'expérimentation, impliquant l'ensemble de la base de données des 3 années de collecte, nous avons choisi d'investiguer les architectures InceptionV3 et MobileNet. On a ajouté des couches denses layer de 128 neurones à Inception V3 et pour le modèle de base de MobileNet une couche *Flatten*, une couche de *Dropout* et une dense layer de 512 neurones et une (Figure 52).

```

architecture_name="InceptionV3 model"
optimizer_name = "SGD"
from keras.applications.inception_v3 import InceptionV3, preprocess_input
input_shape = (299, 299, 3)
base_model= InceptionV3(weights='imagenet', include_top=False, input_shape=input_shape)
x = base_model.output
x = GlobalAveragePooling2D(data_format='channels_last')(x)
x = Dense(1024, activation='relu')(x)
x = Dense(128, activation='relu')(x)
predictions = Dense(class_number, activation='softmax')(x)
model= Model(inputs=base_model.input, outputs=predictions)
print('ok')

```

```

architecture_name = "Custum_MobileNet"
optimizer_name = "SGD"#"Adam" #SGD
from keras.applications.mobilenet import MobileNet, preprocess_input
base_model = MobileNet(weights= 'imagenet', include_top= False, input_shape=(100,200,3))
x = base_model.output
x = GlobalAveragePooling2D()(x)
#x = AveragePooling2D((4, 4), name='avg_pool')(x)
x = Flatten()(x)
x = Dense(1024, activation='relu')(x)
x = Dropout(0.2)(x)
x = Dense(512, activation='relu')(x)
predictions = Dense(class_number, activation='softmax')(x)
model = Model(inputs=base_model.inputs, outputs=predictions)

```

Figure 52 : Extrait de code de InceptionV3 et MobileNet (fine tuned).

V.2. Hyperparamaters

Concernant la première d'expérimentation (année 2017), le *batch size* a été fixé à 32 et le nombre d'époques était de 100 *epochs*. L'optimisation du modèle est réalisée à l'aide de l'algorithme de descente de gradient *Adam optimizer* avec un taux d'apprentissage de 0,001. De plus, nous avons utilisé la méthode de régularisation de l'arrêt précoce (*early stopping regularization*) qui permet d'éviter le surapprentissage (*overfitting*). L'apprentissage est arrêté lorsque la fonction de perte (*validation_loss*) ne s'améliore pas avec un délai (*patience*) de 30 epochs. Tous les scripts sont écrits en langage de programmation *Python 3*. Nous avons utilisé le Framework Keras 2.2⁶, pour à sa simplicité. Nous avons utilisé *Tensorflow 1.9*⁷ en tant que

⁶ <https://www.keras.io/>

⁷ www.tensorflow.org

backend Keras. Nos premières expériences (sur la base de données de 2017) ont été réalisées sur une machine à l'université de MONS (Belgique) ayant les spécifications suivantes :

- Ram: 64 GB
- GPU: 4 x Nvidia Geforce 1080ti 11Gb
- CPU: Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz

Nous avons utilisé aussi la plateforme online « Kaggle » (cloud computing) pour réaliser la suite de nos expérimentations sur les différentes architectures, avec les caractéristiques suivantes :

- Device name: Tesla P100-PCIE-16GB;

Sur la configuration (script) du training, nous avons apporté les modifications :

- Learning rate : 0.0005 pour InceptionV3 et 0.0001 pour MobileNet ;
- L'utilisation de l'*optimizer SGD* (Stochastic Gradient Descent).

VI. Résultats et discussion

Nous allons présenter en premier lieu les premiers résultats que nous avons obtenus avec les data de l'année 2017 et en deuxième lieu les résultats de toutes les années confondues avec les CNNs choisis à savoir *InceptionV3 et MobileNet*.

VI.1. Phase de build model

Le tableau 39 présente les valeurs de la précision d'entraînement et de validation des différentes architectures de CNNs. Tous les modèles ont atteint une précision de validation supérieure ou égale à 97 %. La meilleure valeur de précision de validation a été obtenue avec une légère différence par DenseNet201 de 98,30 %, le Xception de 98,22 % et MobileNet de 98,10 %, tandis que ResNet50 a enregistré la plus faible valeur de précision de validation de 97,55 %. Nous remarquons que le modèle MobileNet prend moins de taille dans la construction du modèle VLC par rapport à ResNet50.

Tableau 39: résultats de l'étape de construction du modèle pour la VLC

Architecture	Building Model (database de 2017)				Ranking
	Training results		Validation results		
	accuracy (%)	loss	accuracy (%)	loss	
InceptionV3	99.92	0.003	97.55	0.13	4
MobileNet	99.84	0.005	98.10	0.12	3
Xception	99.86	0.004	98.22	0.10	2
ResNet50	99.53	0.020	97.30	0.13	5
DenseNet201	97.92	0.010	98.30	0.07	1

La figure 53 montre la variation par époque de la courbe de la valeur de perte et de précision dans les étapes d'entraînement et de validation pour les trois premiers CNNs.

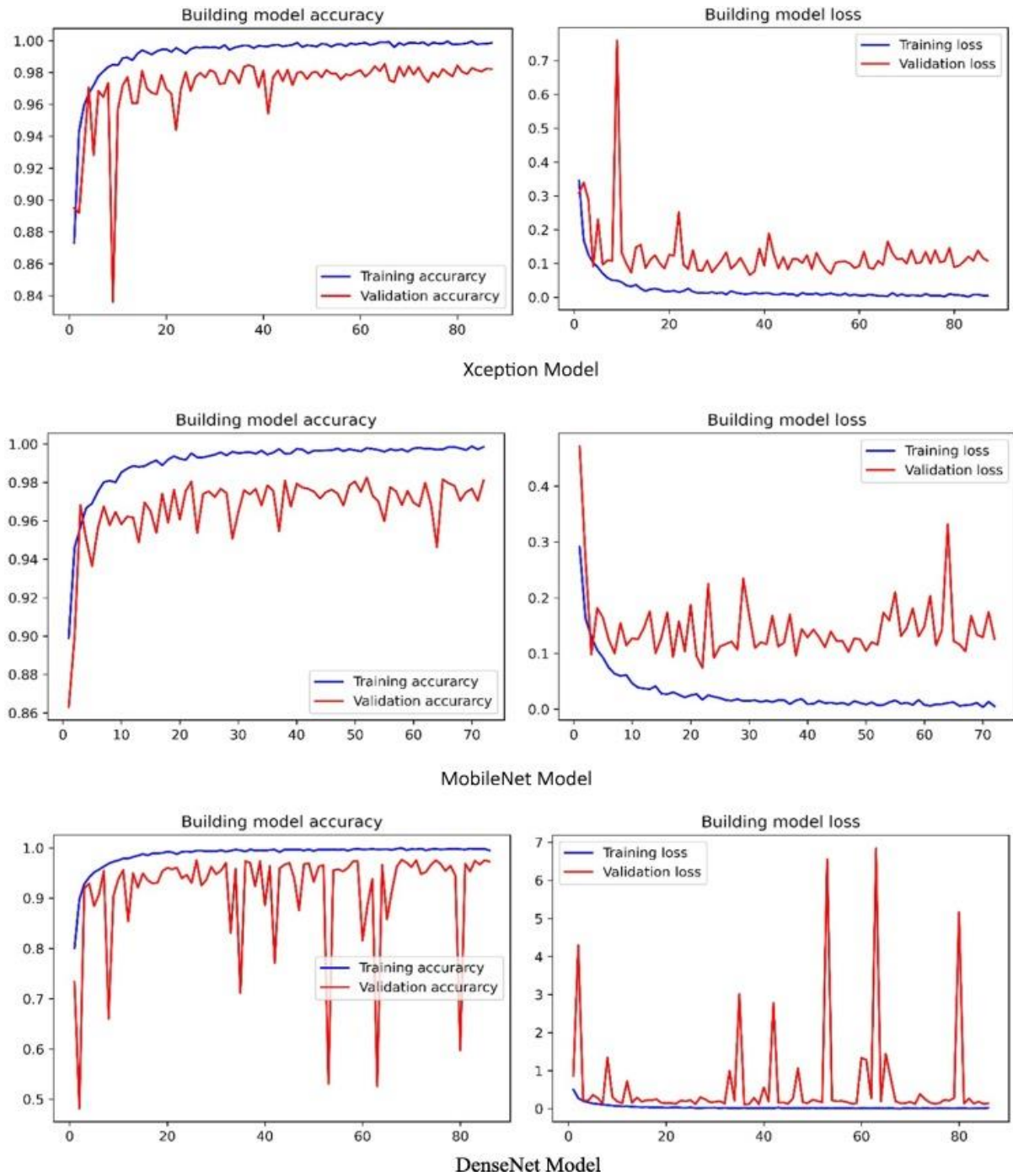


Figure 53: Courbes d'apprentissages et les performances des différents CNNs

VI.2. Test du modèle

Pendant la phase de test, tous les CNNs ont atteint des valeurs de précision comprises entre 94 et 95 % (année 2017). Ces résultats avaient diminué par rapport à ceux de la formation et de l'étape de validation. Les meilleurs scores de test ont été obtenus en superposition par les mêmes architectures que lors de l'étape de validation, mais la meilleure précision a été obtenue par DensNet201 de 95,68 %, suivi par InceptionV3 avec 95,62 % et enfin MobileNet de 95,49 %.

Le tableau 40 indique la précision de chaque architecture pour chaque variété et le CNN le plus approprié pour chaque variété et espèce de blé. Le meilleur classificateur CNN pour Vitron var est DensNet201 avec une valeur de précision de 80 %. En revanche, la même architecture atteint une grande précision lors de la classification de HD avec une valeur de 99,50 %. InceptionV3 classe mieux la Semito que les autres variétés de blé avec 98,93 %. ARZ avait été la mieux classée par toutes les architectures et le Xception avait atteint la valeur de précision la plus élevée de 99,87 %. En ce qui concerne le niveau des espèces de blé, ils sont mieux classés par InceptionV3 avec une précision de test de 99,64 %, avec 99,59 % pour le blé dur et 99,71 % pour le blé tendre.

Tableau 40 : Résultats du test des différents CNNs avec les données de collecte de 2017

Classes\CNNs	InceptionV3	MobileNet	Xception	ResNet50	DenseNet201	1st	2nd
Vitron	78.69	79.07	72.04	76.67	80.15	DenseNet201	MobileNet
Simeto	98.93	98.66	98.60	98.20	98.44	InceptionV3	MobileNet
HD	99.42	99.06	98.93	99.51	99.51	DenseNet201	ResNet50
ARZ	99.20	99.60	99.87	97.60	98.94	Xception	MobileNet
VLC Test	95.62	95.49	94.23	94.87	95.68	DenseNet201	InceptionV3
Hard Wheat	99.59	99.34	99.04	99.42	99.53	InceptionV3	DenseNet201
Bread Wheat	99.71	99.69	99.64	99.40	99.64	InceptionV3	MobileNet
SLC Test	99.64	99.49	99.23	99.41	99.56	InceptionV3	DenseNet201

La matrice de confusion (Tableaux 42-44) donne une vue claire des performances de VLC et indique les objets mal classés.

Tableau 41 :matrice de confusion du modèle *Xception*

Real class \ Predict class	Vitron	Simeto	HD	ARZ
Vitron	1136	398	17	26
Simeto	44	3666	2	6
HD	3	10	3049	20
ARZ	0	1	0	750
<i>Xception</i>				
Correct Class	8601	Wrong Class		527
Accuracy (%)	94.23	Error (%)		5.773

Tableau 42 : matrice de confusion du modèle *InceptionV3*

Real class \ Predict class	Vitron	Simeto	HD	ARZ
Vitron	1241	322	11	3
Simeto	32	3678	2	6
HD	5	4	3064	9
ARZ	2	0	4	745
<i>Inception V3</i>				
Correct Class	8728	Wrong Class		400
Accuracy	95.618	Error		4.382

Tableau 43: matrice de confusion du modèle *MobileNet*

Real class \ Predic class	Vitron	Simeto	HD	ARZ
Vitron	1247	301	15	14
Simeto	44	3668	0	6
HD	3	7	3053	19
ARZ	0	2	1	748
<i>MobileNet</i>				
Correct Class	8716	Wrong Class		412
Accuracy	95.486	Error		4.514

Tableau 44: matrice de confusion du modèle *DenseNet201*

Real class \ Predict class	Vitron	Simeto	HD	ARZ
Vitron	1264	294	8	11
Simeto	52	3660	3	3
HD	6	3	3067	6
ARZ	0	5	3	743
<i>DenseNet</i>				
Correct Class	8734	Wrong Class		394
Accuracy	95.684	Error		4.316

VI.3. Performance des modèles CNNs sur la base de données 2017-2019

Le tableau 45 résume les valeurs et les performances de « train/validation » et le test des modèles entraînés. Une amélioration de 2% en valeur de test est enregistrée pour les deux architectures InceptionV3 et Mobile Net.

Tableau 45: les résultats du test de InceptionV3 et MobileNet avec les données de collecte 2017-2019

CNN	Training		Validation		Test (%)
	accuracy (%)	loss	accuracy (%)	loss	
InceptionV3	99	5.0081e-04	98	0.082	98,2
MobileNet	99	0.02	97.16	0.092	

La figure 54 montre la visualisation graphique de la variation de l'accuracy et loss en fonction des époques d'apprentissage ; l'allure des courbes est plutôt stable par rapport aux courbes précédentes surtout pour l'architecture MobileNet.

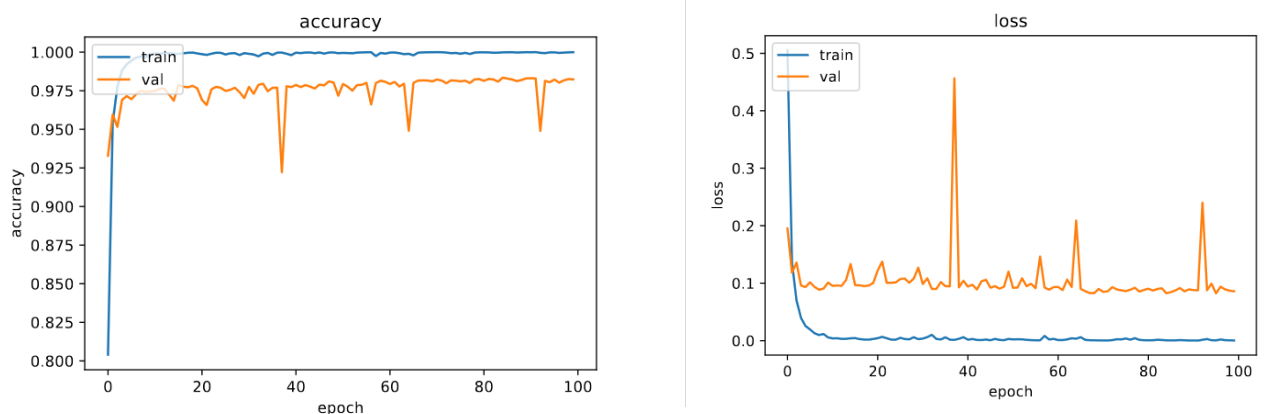


Figure 54: la variation des courbe d'accuracy et loss du Modèle **InceptionV3** modifié

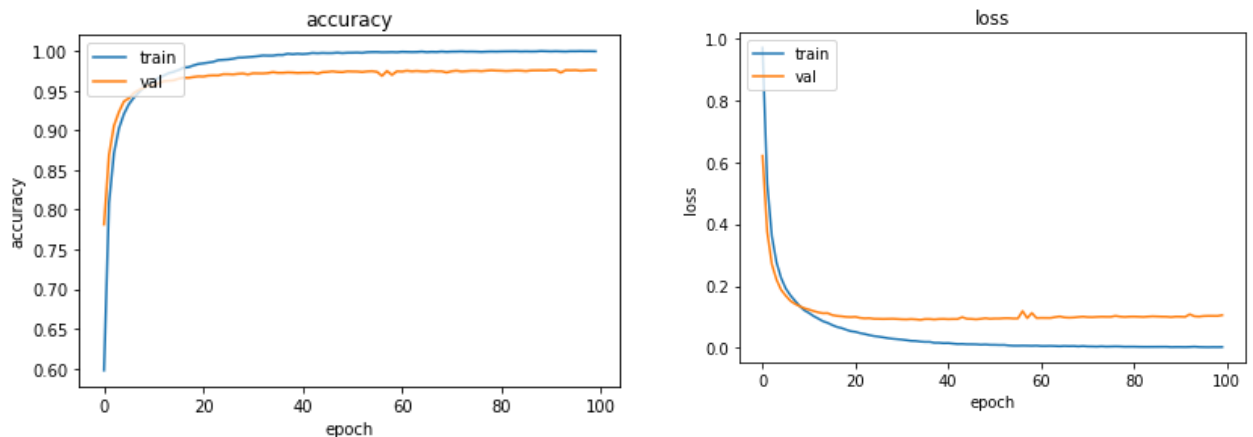


Figure 55: la variation des courbe d'accuracy et loss du Modèle **MobileNet** modifié

Quant aux résultats détaillés du test des deux modèles, les tableaux 46-49 regroupent la matrice de confusion et le rapport de classification.

Tableau 46: matrice de confusion du modèle **InceptionV3** modifié

Real class \ Predict class	Vitron	Simeto	HD	ARZ
Vitron	2007	26	4	3
Simeto	36	2764	4	17
HD	7	7	2797	45
ARZ	2	2	40	2505
InceptionV3				
Correct Class	10073	Wrong Class		193
Accuracy	0,981	Error		0,019

Tableau 47: *Classification report* du modèle **InceptionV3** modifié

	Precision	Recall	F1-score	Support
0	0.98	0.98	0.98	2040
1	0.99	0.98	0.98	2821
2	0.98	0.98	0.98	2856
3	0.97	0.98	0.98	2549
accuracy			0.98	10266
macro avg	0.98	0.98	0.98	10266
weighted avg	0.98	0.98	0.98	10266

{'ARZ': 0, 'HD': 1, 'Simeto': 2, 'Vitron': 3}

Tableau 48: matrice de confusion du modèle **MobileNet** modifié

Real class \ Predict class	Vitron	Simeto	HD	ARZ
Vitron	1988	45	4	3
Simeto	46	2737	11	27
HD	8	12	2777	59
ARZ	2	13	43	2491
MobileNet				
Correct Class	9993	Wrong Class		273
Accuracy	0,973	Error		0,027

Tableau 49: *Classification report* du modèle **MobileNet** modifié

	Precision	Recall	F1-score	Support
0	0.97	0.97	0.97	2040
1	0.98	0.97	0.97	2821
2	0.98	0.97	0.98	2856
3	0.97	0.98	0.97	2549
accuracy			0.97	10266
macro avg	0.97	0.97	0.97	10266
weighted avg	0.97	0.97	0.97	10266

{'ARZ': 0, 'HD': 1, 'Simeto': 2, 'Vitron': 3}

Les résultats montrent que les modèles InceptionV3 et MobileNet gardent toujours leurs capacités à effectuer la VLC avec une bonne précision et aussi plus de données ; ces dernières ont contribué positivement à l'amélioration de la précision de la classification.

VI.4. Pheno-Deep discussion

L'approche Deep Learning permet de réaliser simultanément les opérations d'extraction des attributs et la classification. Nous avons vérifié les performances et la capacité de cinq architectures CNNs à accomplir les tâches VLC. Les résultats démontrent la capacité des CNNs à classer correctement les variétés de blé à un taux supérieur à 95 %. Pour la base de données

de 2017, les trois meilleurs modèles pour VLC sont respectivement le DensNet201 (95,68 %) suivi par InceptionV3 (95,62 %) et enfin MobileNet (95,49 %).

De plus, en ce qui concerne le SLC (niveau espèce) l'InceptionV3 atteint la meilleure précision de test de 99,64 %. L'Inception V3 et le MobileNet 201 ont affiché une amélioration pour la base de données 2017-2019 avec une précision en test de 98.2% et 97.3% respectivement pour la VLC.

Une nette amélioration de la précision pour réaliser la VLC en comparaison à la méthode Shallow, cette amélioration s'explique d'un point de vue Biodigital par le nombre et la nature des attributs exploités dans chaque approche, vu que le Deep learning récupère le plus grand nombre d'attributs dans ces différentes couches d'extraction et en même temps ces attributs sont de nature diversifiée.

Le taux de précision obtenu dans notre expérimentation de Deep learning est supérieur à celui obtenu par Kozłowski et al. (2019) qui ont obtenu un taux de précision supérieur à 93% pour la classification des grains d'orges.

Les résultats montrent également une inter-dissimilarité de chaque CNN à traiter chaque classe avec la même précision (tableau 40) comme dans l'approche Shallow. Par conséquent, cette aptitude peut être qualifiée comme *la spécialisation* d'un CNN spécifique à classer correctement une variété spécifique.

D'un point de vue caractère biologique, notre travail représente une étude à l'échelle de l'organe (Fahimirad and Ghorbanpour, 2019) à savoir le grain. Ce dernier est un phénotype spécifique à l'espèce Triticum. Le grain contient des caractéristiques spécifiques qui ont permis de distinguer chaque variété d'une autre. Les images que nous avons prises pour chaque grain contiennent des informations visuelles sur les traits observables qui ont été convertis en groupes de pixels. On peut donc considérer que les informations contenues dans la face dorsoventrale du grain de blé étaient suffisantes pour que les CNNs construisent un modèle de grain pour chaque variété de blé en effectuant une auto-extraction de toutes les caractéristiques potentielles clés. En effet, il s'agit d'une capacité de l'apprentissage profond permettant l'abstraction des informations relatives au grain de blé.

Les matrices de confusion ont donné une vision claire sur les performances en VLC et indiquent les objets mal classés. L'erreur de classement est proportionnelle à la ressemblance entre des individus de deux classes différentes ; nous expliquons cela par combien certaines variétés sont plus proches les unes des autres ; la Vitron est la variété la plus mal classée. Les variétés de blé

tendre ont été les mieux classées. Cela pourrait être corrélé à la zone de culture et à la taille de l'ensemble de données. L'impact d'une variation temporelle est positif pour le CNN car elle permet d'augmenter la diversité des données et d'enrichir la carte des attributs en profondeur (Deep Features Map).

VII. Conclusion de l'expérimentation N°2

Le Deep Learning a démontré dans cette expérimentation sa grande capacité à réaliser non seulement la SLC mais aussi la VLC avec une précision qui dépasse les 96%. Dans cette expérimentation le problème central de notre étude a été résolu avec moins de 5 % de marge d'erreur. Les différentes architectures CNNs ont été déployées suivant la méthode de Transfert Learning. Cette technique permet d'exploiter « l'apprentissage » de ces CNNs pré-entraînées sur la base de données ImageNet pour notre base de données des grains de blé.

Avec les adaptations nécessaires au Transfert Learning, l'expérimentation dans sa partie software nous a permis de préparer le script nécessaire pour l'exécution de la tâche « *train* » et « *test* » des différentes architectures avec leurs spécifications et réglages nécessaires.

L'implication de deux bases de données (base de 2017 et base de 2017-2019) a permis d'évaluer la stabilité des différents CNNs. Les CNN MobileNet et InceptionV3 ont démontré de bonnes performances sur les différents essais effectués ; des différences mineures ont été enregistrées pour les phases *train* et *test* mais sans affecter le constat principal de la faisabilité de la VLC et la spécialisation des architectures pour une classe donnée.

VIII. Conclusion Générale

L'efficacité et les conditions du travail s'améliorent en introduisant le CV, le ML et l'intelligence artificielle (IA) dans les processus et activités agronomiques. Toutes les opérations basées principalement sur la perception et la décision humaine pourraient être intelligentes et automatisées avec de nouvelles méthodes ergonomiques remodelées, plus efficaces, plus précises et plus traçables.

La question centrale de cette étude tournait autour de la possibilité de reconnaître l'identité des grains de blé, ce qui permet de connaître leurs variétés. Pour étudier ce problème, nous avons proposé et adapté une méthodologie de classification de niveau variétale basée sur la classification d'images, suivant les deux approches Shallow et Deep Learning.

Les deux approches partagent une conduite principale : la collecte des données, le traitement et l'analyse de ces données en problème de classification d'images, l'extraction des attributs discriminant étant la principale clé de réussite de l'apprentissage d'un modèle de classification ; s'ajoute à cela, comme clé de réussite de la classification, la quantité et la diversité des données collectées.

Une stratégie de collecte doit être prise en considération prioritairement pour tout déploiement de la vision artificielle ; la richesse et la taille de la base de données permettra de faire un apprentissage adéquat. Il faut que cette base de données soit la plus ouverte possible à différentes utilisations à travers une bonne labélisation des différentes captures de l'entité biologique dans ses divers interactions et états.

Les résultats que nous avons obtenus dans cette étude montrent la capacité remarquable des techniques Machine Learning à classer les grains de blé sur la base d'images RVB simples. La première expérimentation nous a permis de réaliser l'approche classique de la *classification d'images* « Shallow » en touchant toutes les opérations concernées : la capture, la segmentation, l'extraction des attributs, le data pre-processing et l'exploration des données numériques grâce à la plateforme de data Mining Weka. Dans l'expérimentation Shallow le potentiel du Machine Learning a été établi grâce aux réseaux de neurones multicouches (MLP) aussi nous avons souligné les limites fonctionnelles de cette approche de classification ; ces limites sont dues aux types d'attributs utilisés (Morphologique uniquement) ; la classification de niveau espèce peut atteindre les 85 % mais elle diminue énormément pour le niveau variétal à cause de la haute ressemblance entre les grains.

L'approche Deep Learning reste la plus privilégiée, vu que tous les CNNs affinés ont atteint de bonnes valeurs de précision au cours de la phase de test, comprises entre 94 % et 98.1 %. Les premiers modèles les mieux classés, selon leur précision de test, sont respectivement le InceptionV3 avec (98,2%) et enfin MobileNet (97,3 %).

Sur la base des résultats actuels, une première application immédiate serait dans le processus de test des semences, à travers la conception d'un dispositif intelligent basé sur la CNN la plus performante et qui identifie la variété de grain en temps réel. Nous avons considéré l'InceptionV3 comme le plus adéquat pour la classification de blé. Cependant, pour VLC, nous considérons que l'architecture MobileNet est la plus adéquate à mettre en œuvre sur ce dispositif. Ce modèle présente une petite taille et moins de temps de calcul ; c'est pourquoi il convient aux dispositifs embarqués intelligents, améliorant l'évaluation et la gestion de la qualité des grains de céréales.

Dans nos futurs travaux, nous tenterons de construire une approche DL pour prédire l'origine géographique du grain de blé afin d'avoir une traçabilité du produit de bout en bout. Nous prévoyons également d'améliorer nos résultats en utilisant des techniques de régularisation qui nous permettent d'améliorer la précision sans surapprentissage. Le taux d'apprentissage sera également défini automatiquement pour fournir des courbes de précision et de perte sans pics.

L'apprentissage automatique devrait contribuer à la fois au domaine du phénotypage et à celui de la génétique. Nous tenterons d'appliquer notre méthodologie à d'autres espèces végétales cibles et d'établir un lien entre les résultats de visualisation et l'analyse phénotypique.

Références Bibliographiques

AHDB (2021) *Wheat growth guide Managing wheat growth, Agriculture and Horticulture Development Board*. Available at: <https://ahdb.org.uk/knowledge-library/wheat-growth-guide> (Accessed: 30 September 2021).

Alom, M. Z. *et al.* (2019) ‘A State-of-the-Art Survey on Deep Learning Theory and Architectures’, *Electronics*, 8(3), p. 292. doi: 10.3390/electronics8030292.

Anantrasirichai, N. and Bull, D. (2021) ‘Artificial intelligence in the creative industries: a review’, *Artificial Intelligence Review*, pp. 1–68. doi: 10.1007/s10462-021-10039-7.

AOSA (2018) *AOSA Rules for Testing Seeds*. Available at: <https://www.analyzeseeds.com/publications/>.

Bao, Y. *et al.* (2019) ‘Rapid classification of wheat grain varieties using hyperspectral imaging and chemometrics’, *Applied Sciences (Switzerland)*, 9(19). doi: 10.3390/app9194119.

Barmeier, G. and Schmidhalter, U. (2016) ‘High-Throughput Phenotyping of Wheat and Barley Plants Grown in Single or Few Rows in Small Plots Using Active and Passive Spectral Proximal Sensing’, *Sensors (Basel, Switzerland)*, 16(11), pp. 1–14. doi: 10.3390/s16111860.

Barry, S. *et al.* (2017) *Precision to Decision – Current and Future State of Agricultural Data for Digital Agriculture in Australia*. Available at: [https://www.crdc.com.au/sites/default/files/P2D Current and future state of data - CSIRO Data 61 Final Report.pdf](https://www.crdc.com.au/sites/default/files/P2D%20Current%20and%20future%20state%20of%20data%20-%20CSIRO%20Data%2061%20Final%20Report.pdf).

Bauer, S. D., Korč, F. and Förstner, W. (2011) ‘The potential of automatic methods of classification to identify leaf diseases from multispectral images’, *Precision Agriculture*, 12(3), pp. 361–377. doi: 10.1007/s11119-011-9217-6.

Bechtel, D. B. *et al.* (2009) ‘CHAPTER 3: Development, Structure, and Mechanical Properties of the Wheat Grain’, in *WHEAT: Chemistry and Technology*. AACC International, Inc. (Grain Science References), pp. 51–95. doi: 10.1094/9781891127557.003.

Bewley, J. D. *et al.* (2013) *Seeds*. New York, NY: Springer New York. doi: 10.1007/978-1-4614-4693-4.

Bonjean, A. (2001) ‘Histoire de la culture des céréales et en particulier de celle du blé tendre (*Triticum aestivum* L.)’, *Dossier de l’environnement de l’INRA*, (21), pp. 29–37.

Bonneuil, C. and Thomas, F. (2012) *Semences : une histoire politique*. Charles Lé.

Bouckaert, R. *et al.* (2013) ‘WEKA Manual for Version 3-7-10’.

- Brahimi, M., Boukhalfa, K. and Moussaoui, A. (2017) 'Deep Learning for Tomato Diseases: Classification and Symptoms Visualization', *Applied Artificial Intelligence*, 31(4), pp. 299–315. doi: 10.1080/08839514.2017.1315516.
- Breiman, L. (1996) 'Bagging predictors', *Machine Learning 1996 24:2*, 24(2), pp. 123–140. doi: 10.1007/BF00058655.
- Breiman, L. (2001) 'Random Forests', *Machine Learning 2001 45:1*, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324.
- Chaugule, A. and Mali, S. N. (2014) 'Evaluation of Texture and Shape Features for Classification of Four Paddy Varieties', *Journal of Engineering (United States)*, 2014, pp. 1–8. doi: 10.1155/2014/617263v.
- Chen, Q., Lin, H. and Zhao, J. (2021) 'Computer Vision Technology in Food', in *Advanced Nondestructive Detection Technologies in Food*. Singapore: Springer Singapore, pp. 91–126. doi: 10.1007/978-981-16-3360-7_4.
- Chiara Delogu (2013) 'Validation of a new method: use of SDS-PAGE technique for the verification of Triticum and ×Triticosecale', *ISTA News Bulletin 147*, (145), p. 32. Available at: https://www.seedtest.org/upload/cms/user/STI_145_April2013_web.pdf.
- Choudhary, R. *et al.* (2009) 'Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples', *Biosystems Engineering*, 102(2), pp. 115–127. doi: 10.1016/j.biosystemseng.2008.09.028.
- Choudhary, R., Paliwal, J. and Jayas, D. S. Ñ. (2008) 'Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images', *Biosystems Engineering*, 99(3), pp. 330–337. doi: 10.1016/j.biosystemseng.2007.11.013.
- Copeland, L. O. and McDonald, M. B. (2001) *Principles of Seed Science and Technology, Principles of seed science and technology*. Boston, MA: Springer US. doi: 10.1007/978-1-4615-1619-4.
- Coppens, F. *et al.* (2017) 'Unlocking the potential of plant phenotyping data through integration and data-driven approaches', *Current Opinion in Systems Biology*, pp. 58–63. doi: 10.1016/j.coisb.2017.07.002.
- Davies, E. R. (2009) 'The application of machine vision to food and agriculture: a review', *The Imaging Science Journal*, 57(4), pp. 197–217. doi: 10.1179/174313109x454756.
- Davies, E. R. (2012) *Computer and Machine Vision, Computer and Machine Vision*. London: Elsevier (Texts in Computer Science). doi: 10.1016/C2010-0-66926-4.

- Davies, E.R. (2012) ‘Inspection of Cereal Grains’, in Davies, E. R. B. T.-C. and M. V. (Fourth E. (ed.) *Computer and Machine Vision*. Boston: Elsevier, pp. 553–577. doi: 10.1016/B978-0-12-386908-1.00021-5.
- Delwiche, S. R. (2010) ‘Analysis of grain quality at receipt’, in Wrigley, C. W. and Batey, I. L. (eds) *Cereal grains: Assessing and managing quality*. Woodhead Publishing Limited, pp. 267–310. Available at: <http://dx.doi.org/10.1533/978-1-84569-563-7.3.267>.
- Dexter, J. E. *et al.* (1989) ‘The structure and protein composition of vitreous, piebald and starchy durum wheat kernels’, *Journal of Cereal Science*, 10(1), pp. 19–32. doi: 10.1016/S0733-5210(89)80031-1.
- Dholakia, B. B. *et al.* (2003) ‘Molecular marker analysis of kernel size and shape in bread wheat’, *Plant Breeding*, 122(5), pp. 392–395.
- Dolata, P. and Reiner, J. (2018) ‘Barley Variety Recognition with Viewpoint-aware Double-stream Convolutional Neural Networks’, in, pp. 101–105. doi: 10.15439/2018F286.
- Dollár, P. *et al.* (2007) ‘Feature mining for image classification’, in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1–8.
- Drea, S. *et al.* (2005) ‘Systematic Spatial Analysis of Gene Expression during Wheat Caryopsis Development’, *The Plant Cell*, 17(8), pp. 2172–2185. doi: 10.1105/tpc.105.034058.
- Dreccer, M. F. *et al.* (2018) ‘Comparison of sensitive stages of wheat, barley, canola, chickpea and field pea to temperature and water stress across Australia’, *Agricultural and Forest Meteorology*, 248, pp. 275–294. doi: 10.1016/J.AGRFORMET.2017.10.006.
- Du, C.-J. and Sun, D.-W. (2004) ‘Recent developments in the applications of image processing techniques for food quality evaluation’, *Trends in Food Science & Technology*, 15(5), pp. 230–249. doi: 10.1016/j.tifs.2003.10.006.
- Du, C. J. and Sun, D. W. (2006) ‘Learning techniques used in computer vision for food quality evaluation: A review’, *Journal of Food Engineering*, 72(1), pp. 39–55. doi: 10.1016/j.jfoodeng.2004.11.017.
- Dutta Gupta, S. and Ibaraki, Y. (2014) *Plant Image Analysis: Fundamentals and Applications*.
- Ebrahimi, E., Mollazade, K. and Babaei, S. (2014) ‘Toward an automatic wheat purity measuring device: A machine vision-based neural networks-assisted imperialist competitive algorithm approach’, *Measurement*, 55, pp. 196–205. doi: 10.1016/j.measurement.2014.05.003.
- Edwards, M. A. (2010) ‘Morphological features of wheat grain and genotype affecting flour

yield’.

Elias, S. G. *et al.* (2012) *Seed Testing: Principles and Practices*. Michigan State University Press. Available at: <http://www.jstor.org/stable/10.14321/j.ctt7zt51m>.

EORBTA, E. O. R. B. T. A. and A.-F. C. (2011) *BBCH_STAGING_MANUAL_CEREALS_CORN | Enhanced Reader*. Available at: [moz-extension://d79df280-acb9-4a48-8a9c-b76517d4f47e/enhanced-reader.html?openApp&pdf=https%3A%2F%2Fsmavex12.espaceweb.usherbrooke.ca%2FBBCH_STAGING_MANUAL_CEREALS_CORN.pdf](https://extension.mgbr.com/extension/d79df280-acb9-4a48-8a9c-b76517d4f47e/enhanced-reader.html?openApp&pdf=https%3A%2F%2Fsmavex12.espaceweb.usherbrooke.ca%2FBBCH_STAGING_MANUAL_CEREALS_CORN.pdf) (Accessed: 30 September 2021).

Evers, T. and Millar, S. (2002) ‘Cereal Grain Structure and Development: Some Implications for Quality’, *Journal of Cereal Science*, 36(3), pp. 261–284. doi: 10.1006/jcrs.2002.0435.

Fahimirad, S. and Ghorbanpour, M. (2019) ‘Omics Approaches in Developing Abiotic Stress Tolerance in Rice (*Oryza sativa* L.)’, in *Advances in Rice Research for Abiotic Stress Tolerance*. Elsevier, pp. 767–779. doi: 10.1016/B978-0-12-814332-2.00038-1.

FAO (2019) *La situation mondiale de l'alimentation et de l'agriculture 2019. Aller plus loin dans la réduction des pertes et gaspillages de denrées alimentaires*. Available at: <http://www.fao.org/3/ca6030fr/ca6030fr.pdf>.

Ferentinos, K. P. (2018) ‘Deep learning models for plant disease detection and diagnosis’, *Computers and Electronics in Agriculture*, 145, pp. 311–318. doi: 10.1016/j.compag.2018.01.009.

Francl, L. J. and Panigrahi, S. (1997) ‘Artificial neural network models of wheat leaf wetness’, *Agricultural and Forest Meteorology*, 88(1–4), pp. 57–65. doi: 10.1016/S0168-1923(97)00051-8.

Gegas, V. C. *et al.* (2010) ‘A Genetic Framework for Grain Size and Shape Variation in Wheat’, 22(April), pp. 1046–1056. doi: 10.1105/tpc.110.074153.

gnis-pedagogie (no date) *L'ADN - Semence Pédagogie*. Available at: <https://www.gnis-pedagogie.org/sujet/adn/> (Accessed: 6 October 2021).

Godefroy, J. E. *et al.* (2012) ‘Elliptical Fourier descriptors for contours in three dimensions: A new tool for morphometrical analysis in biology’, *Comptes Rendus Biologies*, 335(3), pp. 205–213. doi: 10.1016/J.CRVI.2011.12.004.

Gong, Z., Zhong, P. and Hu, W. (2019) ‘Diversity in Machine Learning’, *IEEE Access*, 7, pp. 64323–64350. doi: 10.1109/ACCESS.2019.2917620.

Gooding, Michael J (2009) ‘CHAPTER 2: The Wheat Crop’, in *WHEAT: Chemistry and*

Technology. 3340 Pilot Knob Road, St. Paul, Minnesota 55121, U.S.A.: AACC International, Inc. (Grain Science References), pp. 19–49. doi: 10.1094/9781891127557.002.

Gooding, Michael J. (2009) *WHEAT: Chemistry and Technology*, *WHEAT: Chemistry and Technology*. Edited by K. Khan and P. R. Shewry. 3340 Pilot Knob Road, St. Paul, Minnesota 55121, U.S.A.: AACC International, Inc. doi: 10.1094/9781891127557.

Guevara-Hernandez, F. and Gomez-Gil, J. (2011) ‘A machine vision system for classification of wheat and barley grain kernels’, *Spanish Journal of Agricultural Research*, 9(3), p. 672. doi: 10.5424/sjar/20110903-140-10.

Gürsoy, S. and Güzel, E. (2010) ‘Determination of physical properties of some agricultural grains’, *Research journal of applied sciences, engineering and technology*, 2(5), pp. 492–498.

Gustafson, P. *et al.* (2009) ‘Wheat Evolution, Domestication, and Improvement’, in Carver, B. F. (ed.) *Wheat Science and Trade*. Oxford, UK: Wiley-Blackwell, pp. 3–30. doi: 10.1002/9780813818832.ch1.

Hall, M. A. and Smith, L. A. (1998) ‘Practical feature subset selection for machine learning’.

Haun, J. R. (1973) ‘Visual Quantification of Wheat Development 1’, *Agronomy Journal*, 65(1), pp. 116–119. doi: 10.2134/agronj1973.00021962006500010035x.

Hornberg, A. (2017) *Handbook of Machine and Computer Vision*. 2nd edn. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. doi: 10.1002/9783527413409.

Howard, A. G. *et al.* (2017) ‘MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications’. Available at: <http://arxiv.org/abs/1704.04861>.

Howitt, C. A. and Miskelly, D. (2017) ‘Identification of Grain Variety and Quality Type’, in *Cereal Grains*. Elsevier, pp. 453–492. doi: 10.1016/B978-0-08-100719-8.00017-6.

Hyles, J. *et al.* (2020) ‘Phenology and related traits for wheat adaptation’, *Heredity* 2020 125:6, 125(6), pp. 417–430. doi: 10.1038/s41437-020-0320-1.

Ian Witten, WAIKATO, T. U. of (2016) ‘Advanced Data Mining with WEKA’, p. 36.

ISTA (2016) ‘International Rules for Seed Testing’, *International Rules for Seed Testing*. Edited by I. S. T. Association, 2016(1), pp. 1–384. doi: 10.15258/istarules.2016.F.

ISTA (2018a) ‘Chapter 1: ISTA Certificates’, *International Rules for Seed Testing*, 2018(1).

ISTA (2018b) ‘Chapter 3: The purity analysis’, *International Rules for Seed Testing*, 2018(1).

ISTA (2018c) ‘Chapter 8: Species and variety testing’, in *International Rules for Seed Testing*. The International Seed Testing Association (ISTA) Zürichstr. 50, CH-8303

Bassersdorf, Switzerland.

ISTA (2018d) ‘Introduction to the ISTA Rules’, in *International Rules for Seed Testing*. The International Seed Testing Association (ISTA) Zürichstr. 50, CH-8303 Bassersdorf, Switzerland.

ISTA (2021) ‘Chapter 2: Sampling’, *International Rules for Seed Testing*, 2021(1), pp. i-2–44. doi: 10.15258/istarules.2021.02.

Jähne, B. (1993) ‘Image Formation and Digitization’, in *Digital Image Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 19–52. doi: 10.1007/978-3-662-21817-4_2.

Jähne, B. (2005) *Digital Image Processing, Digital Image Processing: 6th Revised and Extended Edition*. Berlin/Heidelberg: Springer-Verlag. doi: 10.1007/3-540-27563-0.

Jayas, D. S., Paliwal, J. and Visen, N. S. (2000) ‘Multi-layer neural networks for image analysis of agricultural products’, *Journal of Agricultural and Engineering Research*, pp. 119–128. doi: 10.1006/jaer.2000.0559.

Jia, X., Kuo, B. C. and Crawford, M. M. (2013) ‘Feature mining for hyperspectral image classification’, *Proceedings of the IEEE*, 101(3), pp. 676–697. doi: 10.1109/JPROC.2012.2229082.

Kamilaris, A. and Prenafeta-Boldú, F. X. (2018) ‘Deep learning in agriculture: A survey’, *Computers and Electronics in Agriculture*. Elsevier, pp. 70–90. doi: 10.1016/j.compag.2018.02.016.

Katiyar, S. K. and Arun, P. V (2014) ‘Comparative analysis of common edge detection techniques in context of object extraction’, *IEEE TGRS*, 50(11b). Available at: <http://arxiv.org/abs/1405.6132> (Accessed: 3 February 2022).

Khoshroo, A. and Arefi, A. (2014) ‘Classification of Wheat Cultivars Using Image Processing and Artificial Neural Networks .’, 2(1), pp. 17–22.

Kong, L., Guo, H. and Sun, M. (2015) ‘Signal transduction during wheat grain development’, *Planta*, 241(4), pp. 789–801. doi: 10.1007/s00425-015-2260-1.

Korir, N. K. *et al.* (2013) ‘Plant variety and cultivar identification: advances and prospects’, *Critical Reviews in Biotechnology*, 33(2), pp. 111–125. doi: 10.3109/07388551.2012.675314.

Kozłowski, M., Górecki, P. and Szczypiński, P. M. (2019) ‘Varietal classification of barley by convolutional neural networks’, *Biosystems Engineering*, 184, pp. 155–165. doi: 10.1016/j.biosystemseng.2019.06.012.

Krzyk, K. (2018) ‘Coding Deep Learning for Beginners — Linear Regression (Part 2): Cost

Function | by Kamil Krzyk | Towards Data Science'. Available at: <https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d> (Accessed: 16 January 2022).

Kuhlman, T. and Farrington, J. (2010) 'What is Sustainability?', *Sustainability 2010, Vol. 2, Pages 3436-3448*, 2(11), pp. 3436–3448. doi: 10.3390/SU2113436.

Kumar, A. *et al.* (2016) 'Dissection of Genetic Factors underlying Wheat Kernel Shape and Size in an Elite × Nonadapted Cross using a High Density SNP Linkage Map', *The Plant Genome*, 9(1), p. plantgenome2015.09.0081. doi: 10.3835/plantgenome2015.09.0081.

Kumar, G. and Bhatia, P. K. (2014) 'A Detailed Review of Feature Extraction in Image Processing Systems', in *2014 Fourth International Conference on Advanced Computing & Communication Technologies*. IEEE, pp. 5–12. doi: 10.1109/ACCT.2014.74.

L'OCDE (2009) 'LA CERTIFICATION VARIÉTALE DES SEMENCES DE MAÏS ET DE SORGHO DESTINÉES AU COMMERCE INTERNATIONAL', pp. 213–248.

Laabassi, K. *et al.* (2021) 'Wheat varieties identification based on a Deep Learning approach', *Journal of the Saudi Society of Agricultural Sciences*. doi: 10.1016/j.jssas.2021.02.008.

LANCASHIRE, P. D. *et al.* (1991) 'A uniform decimal code for growth stages of crops and weeds', *Annals of Applied Biology*, 119(3), pp. 561–601. doi: 10.1111/j.1744-7348.1991.tb04895.x.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521(7553), pp. 436–444. doi: 10.1038/nature14539.

Lee, S. H. *et al.* (2017) 'How deep learning extracts and learns leaf features for plant classification', *Pattern Recognition*, 71, pp. 1–13. doi: 10.1016/j.patcog.2017.05.015.

Les, Z. and Les, M. (2020) 'Machine Perception—Machine Perception MU', in *Studies in Computational Intelligence*. Springer, Cham, pp. 9–44. doi: 10.1007/978-3-030-24070-7_2.

Li, C. *et al.* (2019) 'Wheat VRN1, FUL2 and FUL3 play critical and redundant roles in spikelet development and spike determinacy', *Development*, 146(14). doi: 10.1242/dev.175398.

Li, F.-F., Karpathy, A. and Johnson, J. (2016) *Lecture 8-Lecture 8: Spatial Localization and Detection*. Available at: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf (Accessed: 23 January 2022).

Li, L., Zhang, Q. and Huang, D. (2014) 'A Review of Imaging Techniques for Plant Phenotyping', *Sensors*, 14(11), pp. 20078–20111. doi: 10.3390/s141120078.

- Li, Y. *et al.* (2018) ‘A Genome-wide View of Transcriptome Dynamics During Early Spike Development in Bread Wheat’, *Scientific Reports*, 8(1), p. 15338. doi: 10.1038/s41598-018-33718-y.
- Liakos, K. G. *et al.* (2018) ‘Machine learning in agriculture: A review’, *Sensors (Switzerland)*, p. 2674. doi: 10.3390/s18082674.
- Lu, Y. and Young, S. (2020) ‘A survey of public datasets for computer vision tasks in precision agriculture’, *Computers and Electronics in Agriculture*. doi: 10.1016/j.compag.2020.105760.
- MacDonald, A. A. (1994) ‘Classification and grading’, in *Wheat*. Boston, MA: Springer US, pp. 37–46. doi: 10.1007/978-1-4615-2672-8_4.
- Magnin-Gonze, J. (2009) *Histoire de la botanique*. Delachaux et Niestlé (Les références du naturaliste). Available at: <https://books.google.dz/books?id=IMsqPwAACAAJ>.
- Mahdianpari, M. *et al.* (2018) ‘Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery’, *Remote Sensing 2018, Vol. 10, Page 1119*, 10(7), p. 1119. doi: 10.3390/RS10071119.
- Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018) ‘Statistical and Machine Learning forecasting methods: Concerns and ways forward’, *PLoS ONE*, 13(3), pp. 1–26. doi: 10.1371/journal.pone.0194889.
- Mcdonald, M. B. (1993) ‘The history of seed vigor testing’’, *JOURNAL OF SEED TECHNOLOGY*, 17(2).
- McMaster, G. S. (2009) ‘Development of the Wheat Plant’, in *Wheat Science and Trade*. Oxford, UK: Wiley-Blackwell, pp. 31–55. doi: 10.1002/9780813818832.ch2.
- Mebatsion, H. K., Paliwal, J. and Jayas, D. S. (2013) ‘Automatic classification of non-touching cereal grains in digital images using limited morphological and color features’, *Computers and Electronics in Agriculture*, 90, pp. 99–105. doi: 10.1016/j.compag.2012.09.007.
- Meier, U. *et al.* (2009) *Journal für Kulturpflanzen*, 61 (2). S. 41–52, 2009, ISSN 0027-7479, DOI: 10.5073/JfK.2009.02.01, Verlag Eugen Ulmer KG, Stuttgart. Available at: https://ojs.openagrar.de/volltexte/Kulturpflanzenjournal/2009/Heft02/Web-Daten/01_meier_et_al/meier_et_al.html#Meier_1985 (Accessed: 27 September 2021).
- Meier, U. (2018) ‘Stades phénologiques des mono-et dicotylédones cultivées: BBCH Monographie’. Quedlinburg: Open Agrar Repository. doi: 10.5073/20180906-075455.
- Meyer, D. J. L. and Wiersema, J. H. (2016) *AOSA Rules for Testing Seeds: Principles and*

procedures. Association of Official Seed Analysts (AOSA Rules for Testing Seeds). Available at: <https://books.google.dz/books?id=N53ctAEACAAJ>.

Mutlag, W. K. *et al.* (2020) ‘Feature Extraction Methods: A Review’, *Journal of Physics: Conference Series*, 1591(1), p. 012028. doi: 10.1088/1742-6596/1591/1/012028.

Nevatia, R. (1982) ‘Machine perception.’, *PRENTICE-HALL, INC., ENGLEWOOD CLIFFS, NJ 07632, 1982, 209*.

Nuttall, J. G. *et al.* (2017) ‘Models of grain quality in wheat—A review’, *Field Crops Research*, 202, pp. 136–145. doi: 10.1016/j.fcr.2015.12.011.

O’Grady, M. J. and O’Hare, G. M. P. (2017) ‘Modelling the smart farm’, *Information Processing in Agriculture*, 4(3), pp. 179–187. doi: 10.1016/j.inpa.2017.05.001.

OCDE (2019) *Oecd Seed Schemes for the Varietal Certification or the Control of Seed Moving in International Trade*. Available at: <https://www.oecd.org/agriculture/seeds/documents/oecd-seed-schemes-rules-and-regulations.pdf>.

Paliwal, J. *et al.* (2003) ‘Cereal Grain and Dockage Identification using Machine Vision’, *Biosystems Engineering*, 85(1), pp. 51–57. doi: 10.1016/S1537-5110(03)00034-5.

Paliwal, J., Visen, N. S. and Jayas, D. S. (2001) ‘Evaluation of neural network architectures for cereal grain classification using morphological features’, *Journal of Agricultural and Engineering Research*. Academic Press, pp. 361–370. doi: 10.1006/jaer.2001.0724.

Pantazi, X. E., Moshou, D. and Bochtis, D. (2020) ‘Artificial intelligence in agriculture’, in *Intelligent Data Mining and Fusion Systems in Agriculture*. Elsevier, pp. 17–101. doi: 10.1016/B978-0-12-814391-9.00002-9.

Patrício, D. I. and Rieder, R. (2018) ‘Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review’, *Computers and Electronics in Agriculture*, 153, pp. 69–81. doi: <https://doi.org/10.1016/j.compag.2018.08.001>.

Pitt, H. (2014) ‘Lesley Head, Jennifer Atchison and Alison Gates: Ingrained: a human biogeography of wheat’, *Agriculture and Human Values*, 31(2), pp. 327–328. doi: 10.1007/s10460-014-9493-7.

Ponce-García, N. *et al.* (2017) ‘Grading Factors of Wheat Kernels Based on Their Physical Properties’, in *Wheat Improvement, Management and Utilization*. InTech. doi: 10.5772/67246.

Portères, R. (1966) ‘Aspects de l’Ethnobotanique comme discipline scientifique affirmée’, *Journal d’agriculture tropicale et de botanique appliquée*, 13(12), pp. 701–704. doi:

10.3406/jatba.1966.2909.

Posner, E. S. (2009) 'CHAPTER 5: Wheat Flour Milling', in *WHEAT: Chemistry and Technology*. AACC International, Inc. (Grain Science References), pp. 119–152. doi: doi:10.1094/9781891127557.005.

Powell, A. A. (2009) 'What is seed quality and how to measure it?', in *PROC. 2nd World Seed Conf. Responding to the challenges of a changing world: The role of new plant varieties and high quality seed in agriculture*, FAO Headquarters, Rome, pp. 142–149. Available at: <http://www.fao.org/docrep/014/am490e/am490e03.pdf>.

Pridmore, T. (2015) 'Imaging methods for phenotyping of plant traits', in Kumar, J., Pratap, A., and Kumar, S. (eds) *Phenomics in Crop Plants: Trends, Options and Limitations*. New Delhi: Springer India, pp. 61–74. doi: 10.1007/978-81-322-2226-2_5.

Radhakrishna, D. and Junsong, Y. (2020) 'Computer Vision Basics'. Available at: <https://www.coursera.org/lecture/computer-vision-basics/high-level-vision-hzN9p> (Accessed: 6 February 2022).

Rahaman, M. M. *et al.* (2015) 'Advanced phenotyping and phenotype data analysis for the study of plant growth and development', *Frontiers in Plant Science*, 6. doi: 10.3389/fpls.2015.00619.

Remagnino, P. *et al.* (2017) *Computational Botany*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-53745-9.

Rosentr, K. A. and Evers, A. A. D. (2018) *Kent's Technology of Cereals*. Elsevier. doi: 10.1016/C2014-0-02847-X.

Rust, I. W., Davis, J. H. and Goldberg, R. A. (1957) 'A Concept of Agribusiness', *Journal of Farm Economics*, 39(4), p. 1042. doi: 10.2307/1234228.

Sapirstein, H. D. and Kohler, J. M. (1999) 'Effects of sampling and wheat grade on precision and accuracy of kernel features determined by digital image analysis', *Cereal Chemistry*, 76(1), pp. 110–115. doi: 10.1094/CCHEM.1999.76.1.110.

Schaeffel, F. (2017) 'Processing of Information in the Human Visual System', in *Handbook of Machine and Computer Vision*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, pp. 1–29. doi: 10.1002/9783527413409.ch1.

Selvam, V. and Babu, R. (2015) 'An Overview of Machine Learning and its Applications', *International Journal of Electrical Sciences & Engineering (IJESE)*, 1(1), pp. 22–24.

Serna-Saldivar, S. O. (2010) *Cereal Grains*. CRC Press. doi: 10.1201/9781439882092.

- Serranti, S., Cesare, D. and Bonifazi, G. (2013) ‘The development of a hyperspectral imaging method for the detection of Fusarium-damaged, yellow berry and vitreous Italian durum wheat kernels’, *Biosystems Engineering*, 115(1), pp. 20–30. doi: 10.1016/j.biosystemseng.2013.01.011.
- Shahin, M. A., Symons, S. J. and Hatcher, D. W. (2014) ‘Quantification of mildew damage in soft red winter wheat based on spectral characteristics of bulk samples: a comparison of visible-near-infrared imaging and near-infrared spectroscopy’, *Food and Bioprocess Technology*, 7(1), pp. 224–234.
- Shen, Y. *et al.* (2018) ‘Detection of stored-grain insects using deep learning’, *Computers and Electronics in Agriculture*, 145, pp. 319–325. doi: 10.1016/j.compag.2017.11.039.
- Shewry, P. R. (1996) *Cereal Grain Quality, Cereal Grain Quality*. Edited by R. J. Henry and P. S. Kettlewell. Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-1513-8.
- Shewry, P. R. *et al.* (2009) ‘CHAPTER 8: Wheat Grain Proteins’, in *WHEAT: Chemistry and Technology*. AACC International, Inc. (Grain Science References), pp. 223–298. doi: doi:10.1094/9781891127557.008.
- Shorten, C. and Khoshgoftaar, T. M. (2019) ‘A survey on Image Data Augmentation for Deep Learning’, *Journal of Big Data*, 6(1), p. 60. doi: 10.1186/s40537-019-0197-0.
- Shouche, S. . *et al.* (2001) ‘Shape analysis of grains of Indian wheat varieties’, *Computers and Electronics in Agriculture*, 33(1), pp. 55–76. doi: 10.1016/S0168-1699(01)00174-0.
- Singh, A. *et al.* (2016) ‘Machine Learning for High-Throughput Stress Phenotyping in Plants’, *Trends in Plant Science*, 21(2), pp. 110–124. doi: 10.1016/j.tplants.2015.10.015.
- Sladojevic, S. *et al.* (2016) ‘Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification’. doi: 10.1155/2016/3289801.
- Stone, B. and Morell, M. K. (2009) ‘CHAPTER 9: Carbohydrates’, in *WHEAT: Chemistry and Technology*. AACC International, Inc. (Grain Science References), pp. 299–362. doi: doi:10.1094/9781891127557.009.
- Sun, D.-W. (2016) *Computer Vision Technology for Food Quality Evaluation*.
- Surget, A. and Barron, C. (2005) ‘Histologie du grain de blé’, in.
- Szczypiński, P. M., Klepaczko, A. and Zapotoczny, P. (2015) ‘Identifying barley varieties by computer vision’, *Computers and Electronics in Agriculture*, 110, pp. 1–8. doi: 10.1016/j.compag.2014.09.016.
- Szczypiński, P. M. and Zapotoczny, P. (2012) ‘Computer vision algorithm for barley kernel

identification, orientation estimation and surface structure assessment’, *Computers and Electronics in Agriculture*, 87, pp. 32–38. doi: 10.1016/j.compag.2012.05.014.

Tanabata, T. *et al.* (2012) ‘SmartGrain : High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis 1 [C][W][OA]’, 160(December), pp. 1871–1880. doi: 10.1104/pp.112.205120.

Tanno, K. -i. and Willcox, G. (2006) ‘How Fast Was Wild Wheat Domesticated?’, *Science*, 311(5769), pp. 1886–1886. doi: 10.1126/science.1124635.

Thenmozhi, K. and Srinivasulu Reddy, U. (2019) ‘Crop pest classification based on deep convolutional neural network and transfer learning’, *Computers and Electronics in Agriculture*, 164(June), p. 104906. doi: 10.1016/j.compag.2019.104906.

Too, E. C. *et al.* (2019) ‘A comparative study of fine-tuning deep learning models for plant disease identification’, *Computers and Electronics in Agriculture*, 161, pp. 272–279. doi: 10.1016/j.compag.2018.03.032.

Tripodi, P. *et al.* (2018) ‘Sensing Technologies for Precision Phenotyping in Vegetable Crops: Current Status and Future Challenges’, *Agronomy*, 8(4), p. 57. doi: 10.3390/agronomy8040057.

Turnbull, K. M. and Rahman, S. (2002) ‘Endosperm Texture in Wheat’, *Journal of Cereal Science*, 36(3), pp. 327–337. doi: 10.1006/JCRS.2002.0468.

Tzanis, G. *et al.* (2006) ‘Modern applications of machine learning’, in *Proceedings of the 1st Annual SEERC Doctoral Student Conference–DSC*, pp. 1–10.

Ubbens, J. R. and Stavness, I. (2017) ‘Corrigendum: Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks’, *Frontiers in Plant Science*, 8, p. 1190. doi: 10.3389/fpls.2017.02245.

UPOV (1990) ‘Guidelines for the conduct of tests for distinctness, homogeneity and stability: Kentucky bluegrass : (apomictic varieties)’. (Union Internationale Pour La Protection Des Obtentions Vegetales), p. 55. Available at: http://www.upov.int/about/en/list_p.

UPOV (2010) *NOTES EXPLICATIVES SUR LA DÉFINITION DE LA VARIÉTÉ SELON L’ACTE DE 1991 DE LA CONVENTION UPOV, UPOV/EXN/VAR/1*.

Uzal, L. C. *et al.* (2018) ‘Seed-per-pod estimation for plant breeding using deep learning’, *Computers and Electronics in Agriculture*, 150, pp. 196–204. doi: 10.1016/J.COMPAG.2018.04.024.

Vallade, J. (2004) ‘Endosperme ou albumen? Petite histoire d’un choix terminologique relatif à l’organisation de l’ovule et de la graine chez les Phanérogames’, *Acta Botanica Gallica*,

151(2), pp. 205–219. doi: 10.1080/12538078.2004.10516035.

Venora, G., Grillo, O. and Saccone, R. (2009) ‘Quality assessment of durum wheat storage centres in Sicily: Evaluation of vitreous, starchy and shrunken kernels using an image analysis system’, *Journal of Cereal Science*, 49(3), pp. 429–440. doi: 10.1016/j.jcs.2008.12.006.

Vergauwen, D. and De Smet, I. (2017) ‘From early farmers to Norman Borlaug — the making of modern wheat’, *Current Biology*, 27(17), pp. R858–R862. doi: 10.1016/j.cub.2017.06.061.

Vermeulen, P. *et al.* (2018) ‘Discrimination between durum and common wheat kernels using near infrared hyperspectral imaging’, *Journal of Cereal Science*, 84, pp. 74–82. doi: 10.1016/j.jcs.2018.10.001.

Visen, N. S. *et al.* (2002) ‘Specialist Neural Networks for Cereal Grain Classification N.’, *Biosystems Engineering*, 82(2), pp. 151–159. doi: 10.1006/bioe.2002.0064.

Vithu, P. and Moses, J. A. (2016) ‘Machine vision system for food grain quality evaluation: A review’, *Trends in Food Science & Technology*, 56, pp. 13–20. doi: 10.1016/j.tifs.2016.07.011.

WADDINGTON, S. R., CARTWRIGHT, P. M. and WALL, P. C. (1983) ‘A Quantitative Scale of Spike Initial and Pistil Development in Barley and Wheat’, *Annals of Botany*, 51(1), pp. 119–130. Available at: <http://www.jstor.org/stable/42764187>.

Wäldchen, J. and Mäder, P. (2018) *Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review*, *Archives of Computational Methods in Engineering*. Springer Netherlands. doi: 10.1007/s11831-016-9206-z.

Wolfert, S. *et al.* (2017) ‘Big Data in Smart Farming – A review’, *Agricultural Systems*, 153, pp. 69–80. doi: 10.1016/J.AGSY.2017.01.023.

Wrigley, C. (2017a) ‘Assessing and Managing Quality at all Stages of the Grain Chain’, in *Cereal Grains*. Elsevier, pp. 3–25. doi: 10.1016/B978-0-08-100719-8.00001-2.

Wrigley, C. (2017b) ‘Cereal-Grain Morphology and Composition’, in *Cereal Grains*. Elsevier, pp. 55–87. doi: 10.1016/B978-0-08-100719-8.00004-8.

Wrigley, C., Batey, I. and Miskelly, D. B. T.-C. G. (Second E. (eds) (2017) ‘Woodhead Publishing Series in Food Science, Technology and Nutrition’, in *Cereal Grains*. Elsevier, pp. xxiii–xxxviii. doi: 10.1016/B978-0-08-100719-8.00032-2.

Wrigley, C. W. (2009) ‘CHAPTER 1: Wheat: A Unique Grain for the World’, in *WHEAT: Chemistry and Technology*. 3340 Pilot Knob Road, St. Paul, Minnesota 55121, U.S.A.: AACCC International, Inc. (Grain Science References), pp. 1–17. doi: 10.1094/9781891127557.001.

- Xie, Q., Mayes, S. and Sparkes, D. L. (2015) ‘Carpel size, grain filling, and morphology determine individual grain weight in wheat’, *Journal of Experimental Botany*, 66(21), pp. 6715–6730. doi: 10.1093/jxb/erv378.
- Yan, L. *et al.* (2017) ‘Identification of QTL for Grain Size and Shape on the D Genome of Natural and Synthetic Allohexaploid Wheats with Near-Identical AABB Genomes’, *Frontiers in Plant Science*, 8. doi: 10.3389/fpls.2017.01705.
- Yin, X.-C. *et al.* (2014) ‘Shallow Classification or Deep Learning: An Experimental Study’, in *2014 22nd International Conference on Pattern Recognition*. IEEE, pp. 1904–1909. doi: 10.1109/ICPR.2014.333.
- Yinian, L. *et al.* (2008) ‘Physicochemical properties of wheat fractionated by wheat kernel thickness and separated by kernel specific density’, *Cereal chemistry*, 85(4), pp. 534–543.
- Zaitoun, N. M. and Aqel, M. J. (2015) ‘Survey on Image Segmentation Techniques’, *Procedia Computer Science*, 65, pp. 797–806. doi: 10.1016/j.procs.2015.09.027.
- Zapotoczny, P. (2011a) ‘Discrimination of wheat grain varieties using image analysis: morphological features’, *European Food Research and Technology*, 233(5), pp. 769–779. doi: 10.1007/s00217-011-1573-y.
- Zapotoczny, P. (2011b) ‘Discrimination of wheat grain varieties using image analysis and neural networks. Part I. Single kernel texture’, *Journal of Cereal Science*, 54(1), pp. 60–68. doi: 10.1016/j.jcs.2011.02.012.
- Zapotoczny, P. (2014) ‘Discrimination of Wheat Grain Varieties Using Image Analysis and Multidimensional Analysis Texture of Grain Mass’, *International Journal of Food Properties*, 17(1), pp. 139–151. doi: 10.1080/10942912.2011.615085.
- Zapotoczny, P., Zielinska, M. and Nita, Z. (2008) ‘Application of image analysis for the varietal classification of barley’, *Journal of Cereal Science*, 48(1), pp. 104–110. doi: 10.1016/j.jcs.2007.08.006.
- Zheng, C. *et al.* (2016) ‘Wheat Grain Yield Increase in Response to Pre-Anthesis Foliar Application of 6-Benzylaminopurine Is Dependent on Floret Development’, *PLOS ONE*. Edited by G. He, 11(6), p. e0156627. doi: 10.1371/journal.pone.0156627.