LIÈGE université
Gembloux
Agro-Bio Tech

Wheat Yield Assessment Using In-Field Organ-Scale Phenotyping and Deep Learning Methods

Alexis Carlier

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE - GEMBLOUX AGRO-BIO TECH

Wheat Yield Assessment Using In-Field Organ-Scale Phenotyping and Deep Learning Methods

Alexis CARLIER

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteurs :
Pr. Benoît MERCATORIS
Pr. Benjamin DUMONT

Année civile : 2023

Abstract

Agriculture is currently facing a series of pressing challenges such as climate change, soil degradation, and population growth, which pose a significant threat to food security and environmental sustainability. In the past, the green revolution, which relied heavily on the use of inputs and new genotypes, led to significant gains in crop productivity. Yet, the use of intensive inputs is no longer sustainable due to its negative environmental impacts and to the societal concerns. Therefore, it is crucial to develop new solutions that enable continued crop improvement while reducing inputs such as fertilizers, pesticides, and water.

To effectively address these challenges, plant breeding plays a vital role, requiring a comprehensive understanding of plant behavior from biological, physiological, and agronomic perspectives. Large-scale phenotyping is essential to this process, as it involves characterizing plants in diverse environmental and cultural situations, enabling the identification of desirable traits for selection and optimization of crop management practices. While manual phenotyping is labor-intensive and subject to human subjectivity, automated phenotyping techniques utilizing sensors are being developed to allow for faster, more precise, and objective plant characterization. However, their application under outdoor conditions remains limited due to constraints such as wind, changing light conditions and dense canopy.

In this work, a mobile phenotyping platform has been developed, equipped with a set of cameras. The platform has been used to study several winter wheat trials using two RGB cameras and one multispectral camera. The current thesis focuses mainly on the development of the image analysis pipeline, with a pronounced investigation of the use of artificial intelligence algorithms. The methods have enabled the detection of wheat ears to count the density per hectare, the detection of disease-related damage, and the estimation of biophysical variables such as above-ground biomass, leaf area index and nitrogen content. The deep learning approaches showed to be better as traditional machine learning methods and tends to better generalized. Estimated throughout the growing seasons, these traits were used as predictors of grain yield and yield components providing a deeper understanding of these complex yet highly relevant traits.

Despite significant advancements in the field of automated phenotyping, its adoption by end-users such as breeders is still very limited. Therefore, it is essential to continue raising awareness among stakeholders in the agricultural industry about the benefits of using these technologies to improve sustainable agricultural production. By embracing new technologies, we can help to develop crops that are better adapted to diverse conditions, increase crop productivity, and reduce the negative impacts of agriculture on the environment.

Résumé

L'agriculture est actuellement confrontée à une série de défis pressants tels que le changement climatique, la dégradation des sols et la croissance démographique, qui constituent une menace significative pour la sécurité alimentaire et la durabilité environnementale. Dans le passé, la révolution verte a engendré des gains de productivité grâce à l'utilisation d'intrants et de nouveaux génotypes. Toutefois, l'emploi excessif d'intrants s'avère désormais insoutenable en raison de ses impacts environnementaux et des préoccupations sociétales. Il devient ainsi primordial de concevoir des solutions novatrices permettant d'améliorer les cultures tout en réduisant l'utilisation d'engrais, de pesticides et d'eau.

Face à ces enjeux, la sélection variétale joue un rôle essentiel, nécessitant une compréhension globale du comportement des plantes d'un point de vue biologique, physiologique et agronomique. La phénotypage à grande échelle est essentiel à ce processus, car il consiste à caractériser les plantes dans des situations environnementales et de culture diverses, ce qui permet d'identifier des traits d'intérêts pour la sélection et l'optimisation des itinéraires techniques. Alors que le phénotypage manuel est intensif en main-d'œuvre et sujet à la subjectivité humaine, des techniques automatisées basées sur des capteurs sont en développement, visant une caractérisation plus rapide, précise et objective des plantes. Néanmoins, leur déploiement en condition extérieur est restreint par des facteurs tels que le vent, les variations de luminosité et la densité végétale.

Durant ces quatres années de recherche, une plateforme mobile de phénotypage a été développée. Elle a été utilisée pour suivre plusieurs essais de blé d'hiver en utilisant deux caméras RVB et une caméra multispectrale. La thèse actuelle se concentre principalement sur le développement de la pipeline d'analyse d'image, avec une utilisation importante d'algorithmes d'intelligence artificielle. Ces méthodes ont permis de détecter les épis de blé pour estimer la densité à l'hectare, d'identifier les dommages liés aux maladies et de quantifier des variables biophysiques telles que la biomasse aérienne, l'indice de surface foliaire et la teneur en azote. Finalement, la temporalité de ces traits estimés a été exploitée pour mieux comprendre le rendement en grains et ses composantes de rendement.

Malgré les progrès significatifs dans le domaine du phénotypage automatisé, son adoption par les utilisateurs finaux, notamment les sélectionneurs, demeure limitée. Ainsi, il demeure impératif de sensibiliser les acteurs de l'industrie agricole aux avantages de ces technologies. En adoptant ces nouveautés, ces acteurs auront l'opportunité d'améliorer et d'accélérer leurs efforts pour développer des cultures résilientes, productives et moins impactantes sur l'environnement.

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude envers mon collègue et ami, Sébastien Dandrifosse. Il a été un guide inestimable pour moi, tant sur le plan professionnel grâce à ses vastes connaissances techniques, que sur le plan personnel en m'accueillant à Gembloux. Les moments que nous avons partagés ensemble, que ce soit au laboratoire, sur le terrain ou au bureau, ainsi que toutes les autres expériences vécues ensemble resteront gravés dans ma mémoire.

Ensuite, je voudrais remercier mon promoteur, Benoit Mercatoris, sans qui cette aventure n'aurait pas été possible, pour ses précieux conseils et son soutien constant au cours de ces années. Je tiens également à exprimer ma reconnaissance envers Benjamin Dumont pour les échanges fructueux et son intérêt passionné pour le blé. Mes remerciements vont également à Abdul Mouazen et Vincent Baeten, membres du comité de thèse, pour leur suivi annuel, ainsi qu'aux autres membres du jury de thèse. Je remercie également le service public de Wallonie agriculture ressources naturelles environnement pour le financement du projet PHENWEAT et mon doctorat.

Je tiens également à remercier toutes les personnes qui ont participé de près ou de loin à mon projet et à ma thèse, en particulier Jesse Jap, Rudy Schartz et Julien Kirstein pour le développement de la phénomobile, Damien Vincke et Elias Ennadifi pour le projet PhenWheat, ainsi que l'équipe de la ferme expérimentale pour l'implantation et le suivi des essais. Je suis également reconnaissant envers tous les stagiaires, étudiants et membres du personnel ayant contribué à ce projet. Je souhaite également remercier l'ensemble des personnes du service Biodyne, en particulier du GR, pour leur sympathie, parties de ping-pong ou festivités passées ensemble tout au long de cette expérience, notamment Françoise Thys, Justine Plum, Arnaud Bouvry, Roxane Bruhwyler, Frédéric Lebeau, Christophe Dupriez, Yousra Mejjaouy, Benoit Stalport, Esther Bustillo Vazquez, Nicolas De Cock et Jean-François Willaert.

Enfin, je voudrais exprimer de profonds remerciements envers ma compagne, mes amis et ma famille pour leur soutien indéfectible. Ils ont toujours cru en moi et sont fiers de cette réalisation. Un grand merci.

Contents

List of Figures **xii**

List of Tables **xiv**

Acronyms **xvi**

1 General Introduction **1**

1 Crop Productivity and Demands in a Changing World 3

2 Phenotyping Needs 3

3 Framework of the Thesis 5

4 Objectives and Organization of the Thesis 6

2 Literature Review **9**

1 Synopsis 11

2 Wheat and the Quest for Yield 12

3 Scales of Interest 12

4 Sensing Technologies 13

4.1 Morphological Traits 13

4.2 Physiological Traits 15

5 Unveiling Agronomic Insights: Transforming Raw Images into Knowledge 17

5.1 Synopsis 17

5.2 "Classical" Computer Vision Approaches 17

5.3 Vegetation Indices 19

5.4 Machine Learning 20

5.5 Deep Learning 21

3 Materials and Methods **25**
1 In-Field Phenotyping Platform 27
2 Data Acquisition System 28
2.1 Sensors 28
2.2 Acquisition Settings 30
3 Experimental Protocol 30
3.1 Field Trials 30
3.2 Reference Measurements 31
3.3 Collecting Data 37
4 Image Data Processing 39
4.1 Synopsis 39
4.2 Plant-Soil Segmentation 39
4.3 Height Map from Stereovision 41
4.4 Image Registration 43
4.5 Bi-directional Reflectance Factor 46

4 Detection and Segmentation of Wheat Ears **49**
1 Synopsis 51
2 Current State of the Art 51
3 Machine Learning Approach 52
4 Deep Learning Approach 54
5 Pixel-Based Segmentation Evaluation 55
6 Results and discussion 56
6.1 Comparative Evaluation of Segmentation Approaches 56
6.2 Human Annotation Analysis 57
6.3 Ear density estimation 58
7 Conclusion 59
8 Supplementary Materials 60

5 To What Extent Does Yellow Rust Infestation Affect Remotely Sensed Nitrogen Status? **61**
1 Synopsis 63
2 Current State of the Art and Objectives 63
3 Materials and Methods 65
3.1 Color Image Segmentation 65

3.2 Image Analysis Pipeline . . . 67
3.3 Statistical Analysis . . . 68
4 Results . . . 69
4.1 Disease Pressure . . . 69
4.2 Disease Effects on Bi-Directional Reflectance Factor . . . 71
4.3 Analysis and Modeling of Nitrogen Status Variables Under Fertilization and Fungicide Treatments . . . 73
5 Discussion and Conclusion . . . 76
5.1 CNN as a Promising Damage Detection Tool . . . 76
5.2 Disease Affects Reflectance in Two Ways . . . 77
6 Supplementary Materials . . . 79

6 Comparing CNNs and Traditional Machine Learning for Estimating Wheat Organs Biophysical Variables Using Proximal Sensing 85
1 Synopsis . . . 87
2 Current State of the Art and Objectives . . . 87
3 Materials and Methods . . . 89
3.1 Machine Learning Approach . . . 89
3.2 CNN Training . . . 90
4 Results . . . 92
4.1 Biophysical Vegetation Variable Models . . . 92
4.2 Temporal Wheat Dry Matter and Nitrogen Uptake Partitioning 95
5 Discussion . . . 96
6 Conclusions . . . 100
7 Supplementary Materials . . . 100

7 Yield Estimation From Organ-Scale Traits 101
1 Synopsis . . . 103
2 Materials and Methods . . . 105
2.1 Overview of Used Predicted Traits . . . 105
2.2 Analysis of Yield and Yield Components: Modeling and Correlation . . . 108
3 Results and discussion . . . 108
3.1 Yield Analysis . . . 108
3.2 Individual Traits Analyses . . . 109

3.3 Random Forest Modeling 114
4 Conclusions 115
5 Supplementary Materials 116

8 General Discussion **119**
1 An Emerging Science With Great Potential 121
2 Present Challenges 122
2.1 Technical Obstacles 122
2.2 Matching Needs and Utilization of Phenotyping Systems . . . 124
2.3 Data: The Core of Phenotyping 125
2.4 Enhancing Collaboration through Standardization 126
3 Conclusions and Perspectives 127

9 General Conclusions and Future Perspectives **131**
1 Summarizing Key Conclusions 133
2 Perspectives and Future Work 134

List of Achievements **137**

Bibliography **139**

List of Figures

1.1 Structure of the thesis. 8

2.1 Overview of the field phenotyping systems which concern this thesis. 11
2.2 Overview of image analysis methods details in this thesis. 18

3.1 Image acquisition system timeline. 28
3.2 Sensor set-up. 29
3.3 Overview of the season's precipitations in mm. 36
3.4 Overview of the data collection for each trial across the four seasons. . 38
3.5 Overview of the image processing techniques. 40
3.6 Example of a 800 nm image with its corresponding histogram. 41
3.7 Result of the segmentation on the 800 nm image. 41
3.8 Stereovision pipeline (adapted from [Dandrifosse, 2022]). 42
3.9 Comparison of filling algorithms for height maps. 44
3.10 Comparison of automatic and manual measurements of canopy height at flag leaf stage. 45
3.11 Average bi-directional reflectance factor (BRF) measured on the leaves throughout the day. 48

4.1 Illustration depicting the superpixel labeling process. 54
4.2 Schematic representation of the pixel-based evaluation approach. . . . 57
4.3 Comparison of five estimations of the ear densities of trial 20-F 59

5.1 Image analysis pipeline. 68
5.2 Examples of segmentation utilizing the SegVeg approach 70
5.3 Scaled Visual Score and damage index curves during the 2022 season. 71

5.4 Pearson correlation between the 2022 grain yield, and both $AUDPC_{sVS}$ and $AUDPC_{DI}$. 71
5.5 Boxplot of the Bi-Directional Reflectance Factor according to its source, i.e., from the entire image, only the leaves and only the green elements at different growth stages. 72
5.6 $NDRE_{green}$ according to the treatment for major growth stages. 74
5.7 Correlation matrix between NDRE and nitrogen status variables at different growth stages. 75

6.1 Proposed training pipeline. 91
6.2 Comparison between observed and estimated values of DM of the whole plant and LAI. 95
6.3 Comparison between observed and estimated values of %N and Nupt of the whole plant. 96
6.4 Estimated partitioning of dry matter and nitrogen uptake throughout the season for a micro-plot from the 22-F trial. 97

7.1 Key yield components (adapted from [Slafer et al., 1996]). 104
7.2 Wheat development stages and yield components. 105
7.3 Boxplot illustrating yield variations across different trials in response to varying levels of nitrogen input 110
7.4 Pearson correlation between each yield component. 110
7.5 Random Forest R^2 to assess yield components in a 5 folds cross-validation. 114
7.6 Pearson correlation between yield components and modeled traits. . . 117
7.7 Pearson correlation between yield components and images features. . 118

8.1 Rogers Curve illustrating the estimated positioning of phenotyping technologies in 2023. 128

List of Tables

3.1 Summary of the field trials. 32
3.2 Fertilization schemes of trials 20-F, 21-F and 22-F. 32
3.3 Fertilization inputs of trial 19-F. 32
3.4 Fertilization inputs and fungicide treatments of trials 20-FP, 21-FP and 22-FP. 33
3.5 Scale for the visual scoring of wheat fungal diseases 35
3.6 Comparison of the registration methods. 46

4.1 F1-score comparison between the two ear segmentation methods for common dates of trials 20-F and 20-FP 58
4.2 Description of the 15 vegetation indices and the normalized RGB components used as features for the machine learning algorithms aiming to separate leaves and ears. 60

5.1 IoU and accuracy of the SegVeg model. Soil-Plant results refer to the U-NET model, and the Green-Damage results refer to the pixel-wise classifier. 69
5.2 Rate of change between BRFs of image and leaves, and between leaves and green elements in % 73
5.3 P-values (ANOVA) of NNI, %N leaves, Nuptake leaves, %N total, and Nuptake total at different growth stages. 75
5.4 Vegetation indices selected in this study 79
5.5 Tukey HSD test to investigate the effect of the treatment on the $AUDPC_{DI}$ across the cropping season. 79
5.6 Comparison of variable between the whole leaf and the green leaf, using a paired t-test. 80
5.7 Comparison of variable between the whole leaf and the image, using a paired t-test. 81

5.8 Pearson's correlation between $\Delta_{\text{leaves-green}}$ and the damage index. . . . 82
5.9 Pearson's correlation between $\Delta_{\text{3F-xF}}$ and the damage index. 83
5.10 R^2 of the linear regression associating several feature combinations with agronomic data 84

6.1 Model performances for DM of the plant. 93
6.2 Model performances for LAI. 93
6.3 Model performances for Nupt of the plant. 94
6.4 Model performances for %N of the plant. 94
6.5 R^2 of the different models to estimate DM, Nupt and %N of each organ using the multi-output models. 97
6.6 R^2 of the different models to estimate DM and Nupt proportion of each organ. 100

7.1 Output traits of this chapter. 106
7.2 Input traits used in this chapter, derived from previous computer vision methodologies. 107
7.3 ANOVA analysis for each yield component. 111
7.4 Correlation between image traits and agronomic data by growing period. 113
7.5 Description of the 15 vegetation indices and the normalized RGB components used as features for the machine learning algorithms aiming to separate leaves and ears. 116

8.1 SWOT analysis on the adoption of phenotyping solutions. 129

Acronyms

AI Artificial Intelligence.

BR Brown rust.

BRF Bi-directional Reflectance Factor.

CNN Convolutional Neural Network.

DAS Days After Sowing.

ExR Excess Red vegetation index.

F Fertilization trials.

FP Fertilization and fungicide trials.

GY Grain Yield.

HFOV Horizontal Field Of View.

HSV Hue Saturation Value color space.

HTPP high-throughput plant phenotyping.

ILS Incident Light Spectrometer.

KNN K-Nearest Neighbours.

LAI Leaf Area Index.

LiDAR Light Detection And Ranging.

NDVI Normalized Difference Vegetation Index.

NIR Near-infrared.

NNI Nitrogen Nutrition Index.

PLSr Partial Least Squares Regression.

RF Random Forest.

RGB Red-Green-Blue.

SDK Software Development Kits.

STB Septoria tritici blotch.

SVM Support Vector Machine.

sVS scaled Visual Score.

SWIR Shortwave Infrared.

TKW Thousand Kernel Weight.

UAV Unmanned Aerial Vehicle.

UPVT Photo-Vernalo-Thermic Units (°C-days).

VFOV Vertical Field Of View.

VI Vegetation Index.

VIs Vegetation Indices.

VS Visual Score.

YR Yellow rust.

1

General Introduction

1. Crop Productivity and Demands in a Changing World

The challenge of feeding a growing global population is one of the most pressing issues facing society today. Despite the advances made in crop improvement over the last century, current trends are unlikely to be sufficient to ensure food security by 2050 [Ray et al., 2013]. According to projections by the Food and Agriculture Organization of the United Nations, demand for agricultural products is set to increase by more than 50% between 2012 and 2050 [Food and Agriculture Organization of the United Nations, 2017]. Moreover, in major crop production areas, yields have plateaued after the green revolution [Ray et al., 2012]. Chemical inputs has substantially increased crop production, but they have also raised concerns about sustainability and societal impact [Foley et al., 2011; Ramankutty et al., 2018].

Climate change has further exacerbated the challenges of food production, bringing both opportunities and concerns. Climate variations are responsible for a significant portion of the global yield variation of major crops, which could exceed 60% in some high-production areas [Ray et al., 2015]. Additionally, the scarcity of primary resources and energy increases the competition for resources needed to sustain the population [Food and Agriculture Organization of the United Nations, 2017].

The responsibility of finding solutions to these challenges falls heavily on farmers and their subsidiaries, including breeders, scientists, advisers, agrochemical industries, and policy makers. Crop breeding has been identified as a key factor in the agricultural ecosystem that have the ability to address these issues [Voss-Fels et al., 2019]. Breeding is the essence of crop improvement, but it now requires a new impetus to move towards *speed breeding*, which can accelerate the process of creating new crop varieties [Hickey et al., 2019]. This approach to crop breeding involves using a combination of technologies, such as high-throughput phenotyping and genomic selection, to develop crop varieties that can withstand environmental stress and have higher yields.

2. Phenotyping Needs

Phenotyping is the process of quantifying the traits of an individual and understanding the interactions between their genotype (G), environment (E), and management practices (M) that shape those traits. For a cropping system, a trait refers to a specific characteristic or feature of a plant, such as its height, biomass, yield, flowering time, or disease resistance. In other words, a trait is a specific aspect of a plant that can be observed or measured. Humans have been practicing phenotyping for centuries by selecting the best animals and plants for food, which forms the basis of modern plant breeding.

As advances in genetics have exploded, phenotyping methods are now identified as the bottleneck in the breeding pipeline [Araus and Cairns, 2014; Furbank and Tester, 2011]. In fact, to ensure a stable genotype, breeders need to evaluate their crops in

various locations and farming practices. However, traditional methods of phenotyping often relies on plant sampling or grading that are laborious, expensive, prone to human subjectivity and errors. Consequently, breeders tend to focus on a limited set of major traits related to crop yield without fully understanding the underlying changes.

To meet the growing demand for food, breeders must redirect their focus towards a wider diversity of advanced traits [Reynolds et al., 2020]. Conventional phenotyping methods are often insufficient to capture the complex interactions between G, E and M, and are also subject to biases introduced by human observation. To overcome these limitations, sensing technologies mainly based on different kind of sensors have been developed to allow for the rapid and accurate measurement of plant traits [Araus and Cairns, 2014; Furbank and Tester, 2011].

Coupled with advances in engineering and computer power, the emergence of novel sensing technologies has enabled the collection of vast amounts of data on the plant responses to their environment and the identification of new specific traits [Weiss et al., 2020]. In addition, these technologies allow breeders to obtain objective, non-destructive, and high-throughput data on the characteristics of crops in multiple environments and management practices throughout the crop growth cycle. Such an approach, now well known as high-throughput plant phenotyping (HTPP), has and will significantly improve the breeder's ability to select for desirable traits of a crop [Reynolds et al., 2020].

The utilization of HTPP has far-reaching implications beyond plant breeding. It offers valuable opportunities for farmers, agronomists, physiologists, and scientists to gain insights into plant characteristics. For instance, HTPP could plays a crucial role in crop growth modeling, where researchers can enhance model accuracy and gain a deeper understanding of crop growth processes by determining crop growth model parameters using HTPP data [Saint Cast et al., 2022]. Precision agriculture is another field that has greatly benefited from HTPP methods. For example, satellite imagery and field phenotyping are two related areas that can complement each other. By utilizing these technologies, farmers can optimize input usage by placing them in the right areas at the right times, leading to increased yields, profitability, and reduced environmental impact [Atzberger, 2013; Zhang and Kovacs, 2012]. Moreover, HTPP methods are crucial for large-scale yield forecasting modeling, which is essential for many sectors, including traders, governments, and grain storage companies. These players can create maps of the agro-ecological landscape, land use, agricultural practices, or forest monitoring using HTPP methods, allowing them to make informed decisions and define policies.

Overall, the implementation of HTPP methods has the potential to enhance our understanding of plant growth and development, assisting in addressing global food security issues, which are of paramount importance in today's world.

3. Framework of the Thesis

Despite the promising advances in HTPP, there are still numerous challenges, particularly when dealing with outdoor conditions [Machwitz et al., 2021; Sun et al., 2022]. One of the primary concerns is the collection of high-quality data, which is essential for accurate phenotyping. Field conditions can be unpredictable and variable, making data collection difficult due to factors such as weather, light conditions, soil conditions, and crop variability. Some crops, such as cereals, may present challenges in imaging or measuring specific traits. For instance, wheat canopy is often scattered, overlapping, and messy, making lower leaves not visible. These difficulties can often lead to a reduction of data quality and consistency. Furthermore, in natural conditions, crops are usually subject to multiple biotic and abiotic stresses which are complex to dissociate. This makes it more challenging to understand the individual effects of a specific stress on the crop.

One other challenge of using HTPP is to deal with the large volumes of data it generates, which can be difficult to manage and analyze effectively. To ensure accuracy and data quality, it's crucial to have effective data management strategies for storing, processing, and analyzing the data. The use of machine vision, machine learning algorithms, and advanced deep learning algorithms may be necessary to extract meaningful information from the complex data generated by HTPP. Many methods exist for physiological and morphological measurements but they still require many improvements [Berger et al., 2022]. One of the best path to follow nowadays are the recent advances in neural networks. Coupled with organ scale images, they offer exciting opportunities for addressing some of the challenges. These deep statistical processes can detect even minor differences in complex images without human intervention. Although their use in plant phenotyping is gaining popularity, there are still many aspects that need to be investigated before they can be considered a reliable method for HTPP.

In crop breeding programs, the primary goal of implementing HTPP is to improve breeder work efficiency by identifying and selecting plants with desirable traits. While crop yield is the most important trait, it is also necessary to understand other significant agronomic traits throughout the growing season. Actually, crop yield is a complex trait that can be influenced by various factors throughout the plant's life cycle. Indeed, HTPP can generate valuable knowledge across the growing season, aiding in predicting, understanding how different factors impact crop yield over time. Yet, integrating phenotyping data within breeding programs can be a challenging task, particularly if the programs are already complex and multifaceted. Overcoming these obstacles necessitates ongoing efforts to develop and refine HTPP methods while simultaneously addressing the technical, logistical, and analytical challenges associated with outdoor field conditions. To enable comparisons across different experiments and environments, it is essential to standardize phenotyping protocols and data collection methods. However, the phenotyping science and community are continuously evolv-

ing, and these standards have not yet been fully integrated. Therefore, there is a need for ongoing collaboration and communication within the community to establish and refine these standards, ultimately enabling the full potential of HTPP to be realized in crop breeding programs.

4. Objectives and Organization of the Thesis

This thesis was part of the PhenWheat project, funded by the Walloon Region. The project's goal was to study the growth dynamics of winter wheat varieties resistant to various stresses using a ground-based phenotyping platform. Starting in 2018 for an initial three-year span, the project was extended for three more years in 2021. The research, alongside Sebastien Dandrifosse's PhD thesis [Dandrifosse, 2022], has enabled the acquisition of high-quality data at the organ scale even in challenging outdoor conditions.

It becomes apparent that harnessing this dataset within a comprehensive agronomic framework is not only logical but imperative. Such utilization serves as a compelling demonstration of the feasibility of implementing this cutting-edge methodology at the organ level within the phenotyping community. Moreover, the recent advancements in artificial intelligence (Artificial Intelligence (AI)) have unveiled unprecedented opportunities for overcoming the challenges faced by phenotyping community.

In light of these developments, this thesis focused on exploring the potential synergy between an organ-scale phenotyping system and AI models to retrieve essential wheat yield components. The core objectives encompassed:

- Employing and validating computer vision techniques to extract pertinent traits from the collected data.
- Investigating and evaluating the effectiveness of cutting-edge deep learning algorithms in facilitating data analysis. A specific emphasis was placed on the following domains:
  - Accurate segmentation of wheat organs.
  - Unraveling the interaction between yellow rust and remotely sensed nitrogen.
  - Estimation of biophysical variables.
- Bridging the gap between the derived trait estimated and agronomy, thereby augmenting our comprehension of wheat yields within the context of their components.

To meet these objectives, the thesis is organized into the following chapters (Figure 1.1):

Chapter 2 provides a literature review on phenotyping platforms and sensors, with a focus on proximal systems for field trial conditions. This chapter also explains how to convert raw sensor data into agronomic knowledge and provides an introduction to

wheat crop.

Chapter 3 presents the field trials conducted from 2019 to 2022, including the description of all reference manual measurements and observations, the collection of environmental data, and the protocols of image acquisition. This chapter also introduces the phenotyping platforms and the sensors used, as well as some pre-processing image analysis steps developed with Dandrifosse [2022].

Chapter 4 compares two methods for wheat ear segmentation and presents the method to retrieve wheat ear density.

Chapter 5 investigates the effect of Yellow Rust disease on the measurement of reflectance and explores the impacts on the estimation of wheat nitrogen status using vegetation indices.

Chapter 6 evaluates the efficacy of Convolutional Neural Networks (CNNs) for estimating wheat vegetation biophysical variables. This chapter compares the results with traditional machine learning and explores multi-task models for dry matter and nitrogen uptake partitioning among the organs.

Chapter 7 brings together all the predicted traits from Chapters 3, 4, 5, and 6 for a study on wheat yield and yield components. This chapter explores the best periods and traits for modeling.

Chapter 8 takes a step back and discusses the issues surrounding the adoption of these new technologies beyond research.

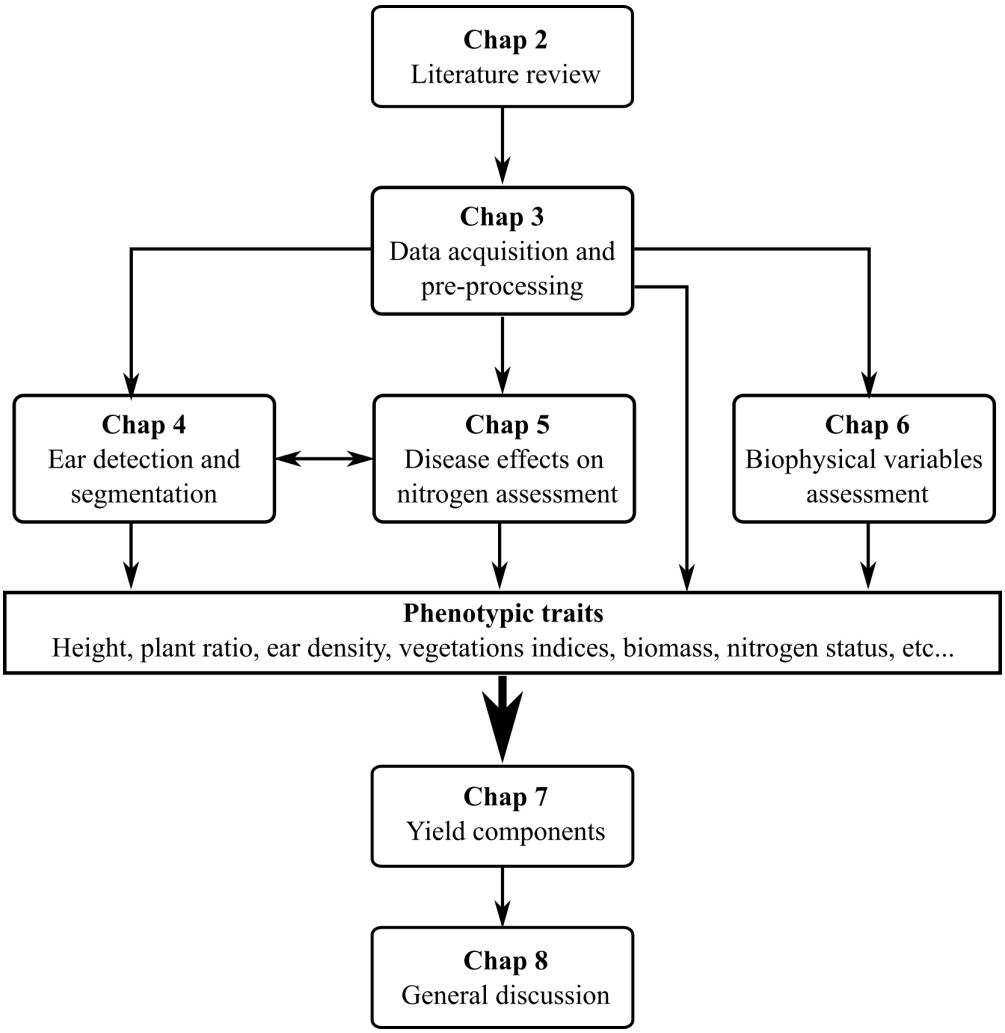

Figure 1.1: Structure of the thesis.

2

Literature Review

1. Synopsis

Every living organism, whether it be an animal or a plant, possesses a distinctive set of traits that define its identity and compose its phenotype. These traits arise from the intricate interplay between the organism's genotype, its surrounding environment, and the management practices it undergoes. Phenotypes constitute a pivotal factor considered in the field of crop enhancement. This chapter offers a comprehensive survey of the latest advancements in phenotyping systems, sensors, and data analysis methodologies pertinent to the central theme of this thesis: proximal sensing through camera technology. The objective is to provide a holistic comprehension of the ongoing cutting-edge progress in this domain, accentuating the most recent breakthroughs and their potential applications (Figure 2.1). The synthesis of this chapter draws upon a multitude of reviews that have also delved into various sensors and techniques not mentioned here [Araus et al., 2022b; Deery and Jones, 2021; Reynolds et al., 2020; Sun et al., 2022]. This literature review also underlines the new and original nature of the proposed research work.

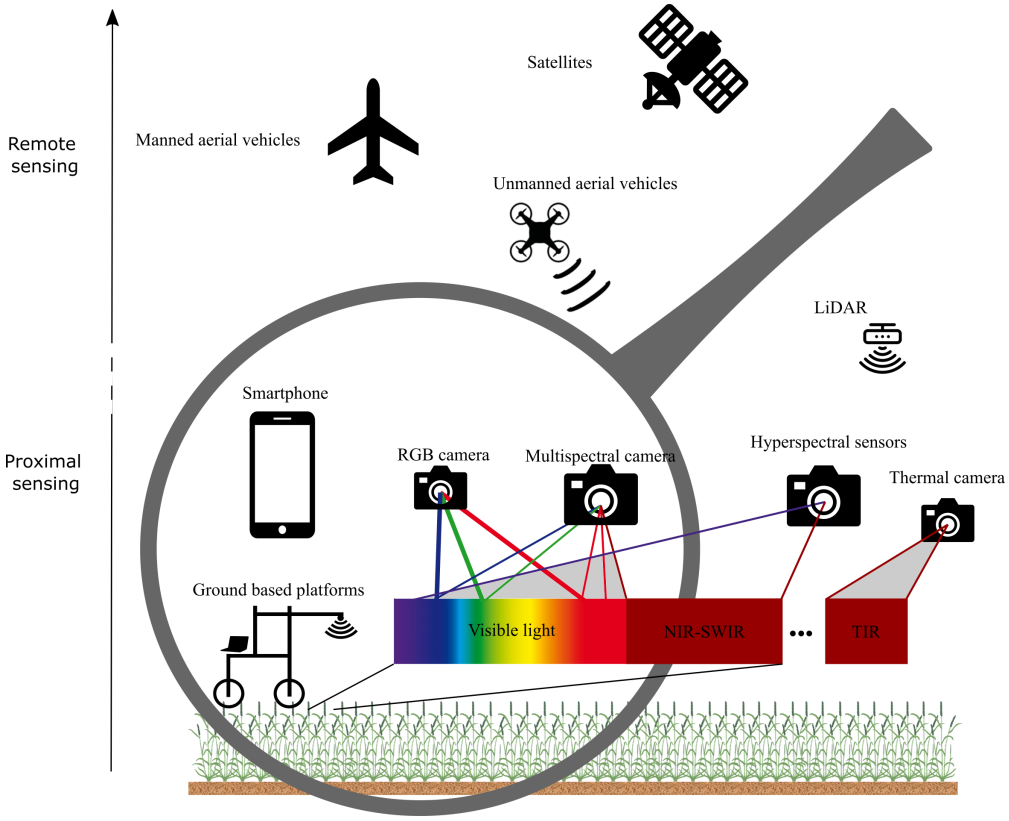

Figure 2.1: Overview of the field phenotyping systems which concern this thesis.

2. Wheat and the Quest for Yield

Before entering into the current state of the art of phenotyping systems, let's put some words about the crop of interest in this thesis. Wheat is an ancient crop that has played a significant role in human history. It is a staple food that provides around one-fifth of the world's food calories and protein [Shiferaw et al., 2013]. With an estimated global cultivation area of 217 million hectares in 2018, wheat is the most widely grown crop in the world [Erenstein et al., 2022]. The yield shows high variability. While the high-potential areas like the Hesbaye region in Belgium can produce above 9 tons per hectare [CRA-W, 2023], the global yield average is about 3.5 tons per hectare [Erenstein et al., 2022].

Yield increases have remained relatively stagnant over the years. The current yield increase is less than 1% per year globally, with the European region showing a stagnating trend [Foulkes et al., 2022; Ray et al., 2013]. These observations are consistent with high heterogeneity in wheat production systems worldwide, making it challenging to develop a unified approach to address the yield gap.

Grain yield in wheat is determined by various genetic and environmental factors, including plant growth and development, nutrient availability, water supply, temperature, light conditions and pest and disease pressure. It is a story of an entire cropping season during which several events can occur. The crop growth is characterized by various growth stages, which are defined conventionally by the BBCH scale [Lancashire et al., 1991]. This is based on phenological stages that range from 0 to 99 and represents specific events in the life of the plant (Figure 7.2). This scale covers all stages of plant growth, from seed germination and emergence to flowering, fruiting, and senescence. BBCH scale of wheat includes ten main growth stages, such as germination, emergence, tillering, stem elongation, booting, heading, anthesis, grain filling, ripening, and senescence. Each growth stage has several sub-stages that refine the plant development description.

3. Scales of Interest

The development and health of crops are influenced by a variety of factors encompassing the plant's genetic makeup, environmental conditions, and management strategies [Hawkesford and Riche, 2020]. However, understanding and tracking these factors in real-world agricultural settings can be quite challenging due to their complexity and variability. When we observe changes in images or signals from the fields, these changes may not solely be attributed to the crops themselves. Other elements, including soil quality, weed presence, pests, and lighting conditions, can also contribute to these observations. Consequently, it can be demanding to differentiate the individual impact of each factor on crop growth, especially when we are looking at crops from a high viewpoint. At elevated positions, a multitude of these factors combine for each

data point, potentially overshadowing the actual changes in crop health. On the contrary, utilizing sensing systems that are closer to the crops provides more detailed data, enabling a more comprehensive analysis of crop health and individual plant components.

The advent of HTPP equipped with embedded sensors has substantially enhanced data collection techniques. There exist two primary categories of these platforms: one involves the collection of data in close proximity to the plants, often referred as proximal sensing. The other category entails gathering data from a distance, known as remote sensing [Jin et al., 2021]. Remote sensing platforms operating from a distance can capture a large amount of data, but their frequency is limited by factors like weather conditions. On the other hand, the proximal-based sensing platforms offer a more detailed view, albeit potentially with a slower throughput.

Within the realm of remote sensing in the frame of field trial characterization, Unmanned Aerial Vehicle (UAV) are the prevalent choice for observing plant growth due to their ability to capture detailed information and ease of use [Guo et al., 2021]. Ground platform can be categorized as handheld devices, stationary setups, and mobile vehicles or robotic systems, each with distinct advantages and disadvantages [Xu and Li, 2022]. Handheld devices, encompassing devices like smartphones, leaf-clipping tools, and specialized poles, are easy to handle and transport. Stationary platforms, requiring significant initial investment due to their fixed setup, can amass substantial data with precise georeferencing and high temporal resolution within a single study [Virlet et al., 2017]. Ground vehicles or robots, including tractor-based platforms, phenomobiles, carts, and robots, can be fully or semi-autonomous and equipped with multiple and heavy sensors [Xu and Li, 2022].

Overall, close-up sensing systems provide more detailed and versatile data collection possibilities, leading to enhanced analysis of crop attributes compared to distant sensing approaches. The choice of platform should be guided by specific study requirements, encompassing considerations like the need for detailed data, data collection frequency, budget constraints, and operational ease.

4. Sensing Technologies

4.1. Morphological Traits

Crops exhibit a range of morphological traits that include plant height, canopy architecture, or the presence of awn. Thanks to recent advances in phenotyping technology, these characteristics can now be retrieved in a high-throughput manner, as many studies did [Araus et al., 2022b; Deery and Jones, 2021; Reynolds et al., 2020; Sun et al., 2022]. Each one of these measured traits are additional bricks that can help to explain the final yield.

A variety of sensors, such as Red-Green-Blue (RGB), hyperspectral, and multispectral imaging, as well as Light Detection And Ranging (LiDAR) and other devices, are capable of extracting morphological traits. Among these, RGB cameras are currently the most popular and cost-effective modality. In fact, RGB sensors have become the most popular data acquisition system due to their level of technological maturity, ease of use, and affordability. In daily life, they are widely used in consumer cameras or embedded in smartphones since this well reflects the color rendering of human vision. Such sensors consist of an array of millions of small photosensitive cells that convert the light intensity in an electrical signal which is directly digitized by the acquisition system. The three-color channels are generally obtained by means of a grid of optical filters directly placed on the sensors in order to restraint each cell to a specific spectral band. The most common RGB filter grid is the Bayer filter which combined red, green and blue cells in a proportion of 25%, 50% and 25%, respectively. The associated wavelength bands of such a filter grid are rather large with 380-520 nm for the blue, 480-620 nm for the green and 580-700 nm for the red (Figure 2.1).

Apart from being capable of reproducing visually acceptable images, RGB cameras are also an excellent source of relevant information for phenotyping tasks, for instance to compute the canopy cover, i.e. the proportion of vegetation in the image used as a good indicator of early vigor [Prey et al., 2018]. It can also be used for shoot biomass estimation [Roth and Streit, 2018], as well as for Leaf Area Index (LAI) estimation [Nielsen et al., 2012]. Tri-dimensional morphological information such as height can be computed from a combination of multi-view RGB images as well. One of this approach, called stereovision, works like the human vision by exploiting the difference in image location of an object seen by two slightly different points of views. This technique has been used by [Jay et al., 2014; Leemans et al., 2013] for plant height and leaf area estimation. Photogrammetry, which consist in acquiring images by means of a moving camera, is also widely used to estimate plant height, LAI, and biomass, especially using UAV [Li et al., 2016]. Niu et al. [2019] compared different approaches based on Vegetation Indices (VIs) and plant height information to estimate above-ground biomass from RGB imagery. Another type of technology that might prove useful for plant phenotyping is active sensors such as LiDAR, which is a laser-based sensor that measures distances by computing the time it takes for an emitted signal to return to an object after it is emitted. This technology provides a 3D point cloud and is mainly used for physiological traits too [Lin, 2015]. LiDAR is especially accurate in estimating plant height, as demonstrated by Zhu et al. [2021].

The future potential of RGB images is closely linked with the dazzling progress recently made in AI. The accuracy and robustness of AI algorithms are continually improving, providing immense potential for plant phenotyping such as fruit counting, weed detection, disease detection and various other applications [Ferentinos, 2018; Kamilaris and Prenafeta-Boldú, 2018; Nabwire et al., 2021]. With the help of AI, many of these tasks have already seen significant improvements in the last years. As

AI continues to evolve and mature, it is expected that RGB images will become an increasingly valuable resource for researchers and practitioners alike, enabling them to perform a wide range of analyses with greater efficiency and accuracy.

4.2. Physiological Traits

In addition to morphological traits, plants possess other characteristics that can help explain their final yield. These traits include the measurement of pigment content or canopy temperature, which reflect the health status of the plant and can be measured using a wide range of sensors. While RGB imaging sensors are also reliable for this purpose, spectroscopy sensors have even more potential.

In fact, RGB images remain excellent tool for highlighting contrasting colors and shapes, which makes it easier to distinguish objects in an image. For example, researchers have effectively used RGB imagery to identify specific patterns of plant diseases in multiple studies [Mahlein, 2016]. UAV-based classification of cercospora leaf spot [Görlich et al., 2021] and wheat yellow-rust detection [Alves et al., 2022; Zhou et al., 2015] are two examples of such studies. RGB imagery is also useful for assessing senescence or stay-green traits [Cao et al., 2021].

Spectroscopy, defined as the study of the electromagnetic radiation absorbed or emitted by matter, has proven to be a powerful tool to measure, in a non-destructive way, plant traits related to bio-chemical processes. The light reflected by vegetation is the result of a complex combination of spectral signatures of its components, such as chlorophyll pigments in the visible, water in the Near-infrared (NIR) and in the Shortwave Infrared (SWIR), and proteins and carbon compounds in the SWIR [Pauli et al., 2016] (Figure 2.1). Additionally, it is also influenced by the plant structure, architecture, and surrounding environment [Chakraborty et al., 2015].

The spectrum of light reflected by vegetation can be measured by spectroscopy sensors which allow to disperse the electromagnetic radiation into its component wavelengths captured by photodiodes. Such sensors can be single-spot or imaging devices and are mainly classified according to their spectral sensitive range and their spectral resolution, as described in [Cavaco et al., 2022].

Spectroradiometers are passive single-spot sensors which acquire spectrum data in a specific range depending on the sensor technology with a rather fine spectral resolution of few nanometers and with a narrow field of view, generally inferior to 25°. Such sensors have been used to derive nitrogen content [Song et al., 2022], assess diseases in wheat in field conditions [Anderegg et al., 2019; Koc et al., 2022], and measure many other traits [Schiefer et al., 2021]. Spectroradiometers are an effective proximal sensing tool that can provide comprehensive data of a mixed area, including crops and their surrounding environment, such as the soil or the incident light. While there are solutions available that allow for data acquisition solely on the crop using a clipping device, this approach significantly reduces throughput and may lack representativeness

of the plant.

In addition, imaging spectroscopy sensors can be seen as arrays of spectroradiometers. For plant phenotyping, hyperspectral and multispectral imaging systems are the main considered instruments in this category. Hyperspectral imaging systems allow discretizing the full spectrum in a specific range with a fine resolution. Different technologies have been emerging to acquire hyperspectral data cubes gathering one spectral dimension with two spatial dimensions [Cavaco et al., 2022; Sarić et al., 2022], and have the potential to touch very specific and advanced traits for a large variety of use cases. Different measuring systems are available such as point or line scanning describe by Sarić et al. [2022]. However, such sensing technologies remain technically difficult to implement outdoor and the raw data are complex to analyze due to the huge amount of generated data. On the other hand, multispectral imaging systems, which are more operational and affordable, are specifically designed to capture the reflected light intensity in a limited number of wavelength bands of interest. A multispectral device is composed of an array of monochromatic cameras, each of them mounted with an optical filter passing the light in a specific wavelength. Such a device is fast and easy to operate and can benefit from the high spatial resolution of simple consumer or industrial cameras. It also exists some other multi/hyperspectral technologies such as wheel filters [Bebronne et al., 2020] or snapshots cameras [Xie et al., 2023]. Both instruments can measure up to twenty-five wavebands offering a trade-off between multi and hyperspectral sensor.

Multi and hyperspectral sensors have proven to be suitable for plant phenotyping [Cavaco et al., 2022; Verrelst et al., 2019], particularly for biotic and abiotic stress detection and, to a lesser extent, for yield quality and plant morphology. Indeed, they are particularly valuable for plant disease detection, which is a hot topic in plant phenotyping [Cheshkova, 2022; Mahlein, 2016; Wan et al., 2022]. Currently, disease detection is possible in the field, such as with Fusarium Head Blight in wheat [Mustafa et al., 2022; Tanner et al., 2022], but it still remains complicated to predict diseases before the appearance of symptoms. Control conditions are currently more suitable. For instance, Peng et al. [2022] detected Cassava brown streak disease "well before symptoms emerge" using a multispectral imaging prototype in the laboratory. Crop nitrogen status has also been accurately assessed using this technologies for decades [Jiang et al., 2022], as well as more specific nutrient such as calcium and magnesium [Andrade et al., 2021].

Both, biotic and abiotic stresses can display similar symptoms, or/and their interaction can be very complicated [Pandey et al., 2017]. The assessment of unique stress is currently an advanced field of research, but the evaluation of simultaneous stresses is still in its infancy [Chowhan and Chakraborty, 2022; Khanna et al., 2019]. Many studies have focused on a single stress, while reality is that crops in the field are subject to multiple stresses. To enhance our understanding of individual stress factors and

their impact on agricultural yield, it is crucial to employ more accurate and precise assessments. In such situation, hyperspectral sensing can detect subtle spectral stress signatures that are difficult to observe with the naked eye or traditional sensors [Devadas et al., 2015]. Another approach to deals with this challenge involves combining multiple sensors to capture complementary information. For instance, while pathogen-induced symptoms can be similar to those caused by water stress, Zarco-Tejada et al. [2021] found that using a combination of hyperspectral and thermal sensors enabled them to differentiate between the two types of stress. By leveraging the advantages of these advanced sensing techniques, we can better identify and manage stress factors to reach higher yields.

In addition to RGB and multi-hyperspectral sensors, thermal imaging, can be used to capture physiological features related to surface temperature [Pineda et al., 2020]. This technology has relatively low spatial resolution and reproducibility and can be strongly influenced by environmental factors such as humidity and wind. No well-established standardization solution has yet been found for proximal sensing. Nonetheless, it is particularly useful for the assessment of biotic stress [Kuska et al., 2022] and abiotic stress, particularly water stress [Gonzalez-Dugo et al., 2015].

5. Unveiling Agronomic Insights: Transforming Raw Images into Knowledge

5.1. Synopsis

This section provides an in-depth exploration of the intricate process involved in extracting agronomic traits from raw images, a pivotal component within the field of plant phenotyping. The contemporary landscape is enriched with a diverse array of sensors, offering the capacity to precisely analyze both morphological and physiological traits of plants. Moreover, HTPP not only facilitates the comprehensive monitoring of agronomical traits but also uncovers novel plant features that hold promise for integration by breeders, as underscored by [Reynolds et al., 2020]. To accomplish this, an array of image processing techniques can be harnessed, ranging from fundamental image manipulation to sophisticated machine learning and AI-driven methodologies. These methodologies collectively fall under the umbrella of computer vision. The following subsections outline the most used methods in plant phenotyping.

5.2. "Classical" Computer Vision Approaches

The term "classical" computer vision encompasses methodologies grounded in mathematical manipulations of images viewed as matrices of pixels. These techniques operate without the need for data-driven model training. They find utility both as preprocessing steps and as direct sources of valuable information (Figure 2.2). Image pre-processing stands as a pivotal initial stage within classical computer vision, aim-

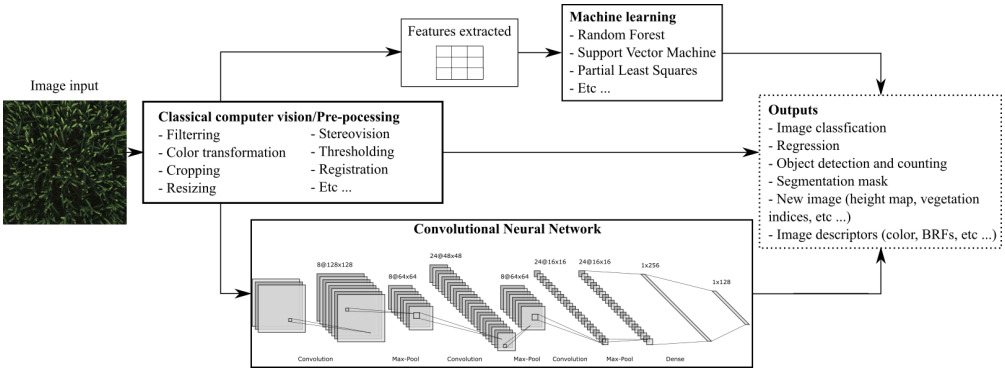

Figure 2.2: Overview of image analysis methods details in this thesis.

ing to rectify and refine raw data due to innate camera characteristics that can, to some extent, influence image quality. A prime instance is the occurrence of vignetting; a detrimental effect caused by lens peripheries leading to diminished light intensity in image margins. Dedicated algorithms can correct and eliminate such effects. Distortions, where straight lines in images bend, can also arise. While many cameras feature distortion corrections, post-processing can also address this through calibration parameters. Noisy images, containing stochastic artifacts, can undergo rectification or smoothing via convolution operations employing specific kernels. Convolution, fundamental to various computer vision methods such as edge detection and blurring, constitutes the core element of Convolutional Neural Network (CNN), elucidated in Section 5.5.

Subsequent to pre-processing, image analysis typically involves extracting relevant features associated to the considered application. Expert practitioners, drawing from their knowledge and expertise, frequently undertake this task. Examples of such features encompass color intensity, edge presence, and more sophisticated attributes like pixel or pixel-group textural characteristics.

Furthermore, a plethora of automated algorithms is available to expedite image analysis. An illustrative case is automated image segmentation, encompassing the division of an image into distinct regions of interest. For instance, the Otsu threshold [Otsu, 1979] yields satisfactory outcomes in segmenting canopies from soil, with machine learning algorithms further enhancing accuracy [Yu et al., 2017].

Incorporating multiple images taken from various viewpoints or sensors can remarkably amplify the wealth of information captured within a scene. Employing stereovision techniques, distances can be extrapolated by analyzing the relative object positions across images, thereby generating intricate 3D point clouds [Dandrifosse et al., 2020; Jay et al., 2014]. Similarly, aerial photographs can undergo geometric correction to amalgamate into an orthomosaic, a prevalent approach within UAV research.

Another illustration involves the utilization of image registration algorithms to align images stemming from disparate imaging systems. This process facilitates the creation of a holistic representation, allowing individual pixels to assimilate precise information from diverse sensors [Henke et al., 2019]. Originating primarily from the medical domain, image registration proves indispensable for analyzing proximal based images. These methodologies, devoid of the necessity for a training phase, exemplify alternative image analysis techniques that yield consistent outcomes. Nevertheless, it is imperative to acknowledge that in certain scenarios, these approaches might fall short of achieving specific objectives.

5.3. Vegetation Indices

Vegetation Indices (VIs) are fundamental numerical metrics derived from spectral data, serving as essential tools for assessing the health, density, and/or vigor of plants, which often elude simple RGB images. They are rooted in mathematical formulations that leverage the reflectance patterns across diverse wavelengths of light from a surface. Their inception dates back to the 1970s, a period marked by the launch of pioneering satellites. These early attempts were constrained by limited spectral coverage and computational resources. Through subsequent decades, the realm of VIs has evolved, with researchers in the field of remote sensing continually refining these measures. Early efforts produced rudimentary two-band indices, occasionally expanding to three or four bands. A comprehensive review of these indices was recently undertaken by [Tayade et al., 2022]. Today, over 500 VIs have been gathered in an Index Database (https://www.indexdatabase.de/) [Henrich et al., 2009]. Notably, these indices find widespread utility within parametric regression models and stand as the most time-tested and commonly employed algorithms for retrieving biophysical parameters via remote sensing [Berger et al., 2020].

The adoption of VIs spans a broad spectrum of applications, particularly within the phenotyping community. Preeminent among these is the Normalized Difference Vegetation Index (NDVI), renowned for its ability to assess the greenness of crops. NDVI's versatility extends to its relevance in assessing early vigor, biomass, Leaf Area Index (LAI), senescence, grain yield, and nitrogen content [Rouse et al., 1974]. Beyond that, inventive fusions of spectral bands, including RGB cameras, have yielded new VIs suitable for unraveling biophysical characteristics encompassing both biotic and abiotic stresses, structural traits, and the ability to limit interference from the soil.

Notwithstanding their undoubted merits, VIs are not immune to limitations. These encompass challenges such as limited generalizability, vulnerability to saturation effects, and underexploitation of the spectral information inherent in the data [Thenkabail and Lyon, 2011]. Unfortunately, some studies have condensed hyperspectral datasets, which included hundreds of bands, into a mere merging of a few selected bands [Tayade et al., 2022]. The efficacy of a specific VI hinges heavily upon the

underlying objective and the correlation with the agronomic data. Consequently, interpreting a solitary value from an NDVI may not confer a universally consistent interpretation. However, even amid the rise of machine learning methodologies, the enduring importance of VIs persists due to their ability to provide comparable performance [Tayade et al., 2022]. Their importance goes beyond the simple reliance on regression models, justifying recognition as stand-alone indicators of botanical traits.

5.4. Machine Learning

The integration of advanced machine learning algorithms, such as Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbours (KNN), within scientific domains, particularly complex fields like remote sensing, presents profound advantages over conventional methods such as linear regression or fundamental image processing techniques. By harnessing machine learning's innate abilities, it has made significant strides in managing vast and heterogeneous data collections. A notable example is Partial Least Squares Regression (PLSr), a renowned algorithm for handling hyperspectral data [Sarić et al., 2022]. This aptitude not only uncovers insights that may evade traditional methodologies but also facilitates the amalgamation of multimodal data, adeptly handles noise-induced variability, and fosters contextual comprehension. While simpler methods have their place in specific scenarios, the complexity of modern scientific challenges and the wealth of available data highlights the central role of machine learning in advancing our understanding of the world around us.

Machine learning is an innovative field with the capacity to acquire knowledge from data. It involves deploying advanced algorithms and statistical models to enable these systems to autonomously identify patterns, extract insights, and make informed predictions or decisions, all without the need for explicit and rigid programming instructions. It is a subfield of artificial intelligence and has gained immense popularity in recent years due to its ability to automate tasks that were previously thought to be impossible for machines to perform. The spectrum of applications spans classification, regression, prediction, clustering, and object detection (Figure 2.2). Successful implementation hinges on relevant feature selection and substantial examples with corresponding labels for training purposes. Broadly categorized as supervised, unsupervised, and reinforcement learning, the former involves instructing models such as RF, SVM, or KNN using labeled data, enabling the model to extrapolate correct outputs for new inputs. Unsupervised learning entails model training without labels, necessitating independent identification of patterns and relationships, often applied in clustering and anomaly detection. Reinforcement learning, pivotal in robotics, gaming, and interactive scenarios, guides models to maximize cumulative rewards through learned actions. The choice of learning method rests on the problem and available data, each bearing distinct merits.

Supervised learning has played a crucial role in phenotyping tasks, and is integrated

into most phenotyping studies, as highlighted by [Araus et al., 2022b; Singh et al., 2016]. Its application spans plant disease detection, yield prediction, and/or growth monitoring [Mahlein, 2016; Sun et al., 2017; Wang et al., 2022a].

The efficacy of machine learning algorithms pivots on feature quality for training. Conventionally, these features, derived through computer vision techniques, necessitate expert domain knowledge and manual crafting. However, this process, prone to errors and time constraints, has been alleviated by robust libraries such as OpenCV [Bradski, 2000] or scikit-learn in Python [Pedregosa et al., 2011]. Coupled with computational advancements, algorithm training times have notably diminished. Despite these improvements, the feature engineering process remains a critical step in the machine learning workflow and requires careful consideration to ensure the best possible performance.

5.5. Deep Learning

5.5.1. Introduction and Evolution

Deep Learning, a prominent subfield of machine learning, has revolutionized artificial intelligence by enabling the training of intricate artificial neural networks to decipher intricate tasks. In contrast to conventional machine learning methods, deep learning networks possess the capacity to autonomously extract relevant features from raw data, eliminating the need for manual feature engineering. Drawing inspiration from the neural architecture of the human brain, the inception of deep learning dates back to 1958 when the pioneering perceptron model was proposed by Rosenblatt [Rosenblatt, 1958]. However, between the 1960s and 2010s, deep learning remained in the shadows of other machine learning algorithms, failing to gather significant momentum. A turning point for deep learning materialized with the emergence of AlexNet [Krizhevsky et al., 2017], a groundbreaking convolutional neural network. This pivotal moment occurred during the ImageNet competition [Deng et al., 2009], where AlexNet not only outperformed existing machine learning methods but also achieved an impressive reduction in error rates, from 26% to 16%. This resounding victory unequivocally positioned deep learning at the forefront of computer vision for the subsequent decade.

Deep learning networks consist of multiple layers of interconnected nodes, or neurons, where each layer processes input data at varying levels of abstraction. The initial layer receives raw input, while subsequent layers receive output from preceding ones. Learning occurs by adjusting weights between nodes in each layer to minimize a loss function, comparing predicted and actual outputs. This iterative adjustment process, called backpropagation, continues until the network generates accurate predictions.

Deep learning's impact is profound, advancing diverse fields like image and speech recognition, natural language processing, and image analysis. Consequently, it plays a pivotal role in our daily lives, from facial recognition on social platforms to natural language-driven automated conversations. However, a significant challenge lies in

the substantial data demands to achieve impressive outcomes. Unfortunately, certain domains such as HTPP processing face limited data access, which can be both time-consuming and costly to acquire. While free-access databases exist in some domains, specific applications lack such resources, exacerbating the application challenge for deep learning techniques. Nonetheless, deep learning's potential merits are extensive, and ongoing research strives to surmount these hurdles, rendering it more accessible across diverse applications.

5.5.2. Convolutional Neural Networks

One of the most used types of neural networks in deep learning is the Convolutional Neural Networks (CNN). CNNs are specifically designed to recognize patterns in visual data, making them ideal for image recognition tasks. Composed of a series of layers combined to form a single architecture, CNNs are highly versatile and can be tailored to fit specific needs (Figure 2.2). In recent years, an impressive number of CNN models have emerged, each aimed at maximizing performance while minimizing the number of parameters required for training. Moreover, researchers are continually pushing the boundaries of CNNs to improve their computational efficiency, ensuring that they remain at the forefront of image recognition technology.

The key element of a CNN is the convolutional layers, which apply filters or kernels to the input image. Each filter produces a feature map that highlights where those features appear in the image, and multiple filters are applied in each convolutional layer to extract a range of features. These features are then passed on to subsequent layers, which can extract increasingly semantically rich features. Another essential element of a CNN is the pooling layer, which reduces the size of the feature maps while retaining the most significant information, helping to prevent overfitting and decrease the number of parameters in the network. The convolution and pooling layers essentially do the job of the engineer looking for best features in an image. The final layers of a CNN typically include fully connected layers, which use the flattened output from the previous layers to make predictions about the input image. The number of neurons in these layers depends on the specific task that the network is being trained for. To determine the best filters or neurons weight for each layer, the model update each weight iteratively through the process of training. This can require a high level of computing power, especially for large datasets. However, once the network has been trained, it can process new images with a high degree of accuracy very fast.

CNNs have emerged as a powerful tool in image classification tasks and have achieved state-of-the-art results on benchmark datasets like ImageNet [Deng et al., 2009]. However, their applications extend beyond image classification and have been employed in a variety of other domains, including regression, object detection, and segmentation.

Thanks to the advent of accessible computing resources and libraries, CNNs have become increasingly accessible to a broader audience. For instance, the release of CUDA by NVIDIA and the availability of web-based platforms such as Google Colab

have made it easier to train and run models. Furthermore, numerous frameworks, including Tensorflow and Pytorch, offer accessible APIs that enable even novice users to develop their own CNN models. Free access databases, such as ImageNet, and COCO, provide a wealth of labeled data that can be used for people training and testing these models, further contributing to their widespread use.

While CNNs have proven to be effective in a broad range of applications, it is important to note that there are still domains where their effectiveness may be limited. For example, in more specific topics such as phenotyping, the story could be different, especially regarding the access to large databases.

5.5.3. CNN for phenotyping

Plant phenotyping is an area where deep learning algorithms have recently shown great promise [Jiang and Changying Li, 2020; Kattenborn et al., 2021; Singh et al., 2018]. Early applications of CNNs in outdoor conditions were focused on plant classification and weed recognition, with the popular PlantNet web platform being a key example [Affouard et al., 2017; Mortensen et al., 2016]. Nowadays, CNNs have demonstrated impressive performances in a range of other tasks, such as wheat head detection, canopy segmentation, and disease classification [Dandrifosse et al., 2022b; Fan et al., 2022; Ferentinos, 2018; Görlich et al., 2021; Serouart et al., 2022]. By finding the most appropriate features, CNNs have shown better abilities to generalize than conventional approaches. For instance, Ma et al. [2019] have shown that a CNN was less sensitive to plant density when estimating wheat biomass at early growth stages than a Random Forest. Furthermore, these novel methods can help to better estimate traits and unlock the extraction of more advanced parameters such as crop growth rate when applied to time-series data [Buxbaum et al., 2022]. Yet field conditions exhibit much more image variability due to lighting conditions, backgrounds, and overlapping plants. This variability could be better managed by these deep learning approaches provided there is a sufficient dataset with large variability.

Indeed, large annotated datasets are essential for developing models that can be generalized and accurate across a broad range of cultivars and conditions. In the plant phenotyping community, it is currently challenging to obtain a large annotated dataset under field conditions. A training dataset is usually composed of images pairs with their corresponding targets, such as classification category, segmentation masks, or object detection bounding boxes, all of which require human annotation or labeling that can be time-consuming to provide accurately. Although several tools are being developed to assist with this task, and alleviate the need of data, model accuracy still depends heavily on massive datasets with labels.

To address the issue of limited labeled data, several methods have been proposed to train robust models. The two most used approach are transfer learning and data augmentation. The first use pre-trained models with defined meaningful weights trained on very large dataset like the ImageNet dataset [Russakovsky et al., 2015]. It has

been successful in estimating forage biomass [Castro et al., 2020; de Oliveira et al., 2021]. However, when dealing with multichannel images, pre-trained models that are typically trained on RGB images are not suitable. Another possibility is to perform data augmentation. It artificially increases the dataset size by applying transformations to the images [Zhang et al., 2022]. In a similar way, it might also be possible to enlarge the dataset by creating synthetic data using a functional–structural plant model (FSPM) [Li et al., 2023]. Other methods are also gaining popularity such as self-supervised learning or reinforcement learning which train without fully labeled dataset.

Recently, open-source datasets have emerged as a solution to this problem in response to high demand of certain traits. The Global Wheat Head Detection dataset is probably one of the best example for the phenotyping community [David et al., 2021]. It is dedicated to wheat head detection and has gained considerable popularity by unlocking the issue of wheat head detection. PlantNet has also released a dataset comprising 300,000 images of plant species [Garcin et al., 2021]. The VegAnn dataset, which was created by [Madec et al., 2023] for crop segmentation, is particularly useful for retrieving canopy cover. Other web platforms, such as PlantVillage [Lobet et al., 2013], also contain crop-related datasets, but with a focus on indoor conditions.

Deep learning is rapidly advancing and are quickly becoming the standard for image analysis. These techniques have demonstrated excellent performances in various tasks. Still, there are numerous challenges that need to be addressed regarding the phenotyping needs, for instance for regression purpose. By addressing these challenges, we can further improve the effectiveness of CNNs for plant image analysis, and continue to push the boundaries of what is possible in this exciting field.

3

Materials and Methods

1. In-Field Phenotyping Platform

In contrast to the indoor phenotyping systems, which involve moving plants to the measuring instruments, the demands of in-field phenotyping require the sensors themselves to be transported to the plants. Within the scope of the PhenWheat project, in conjunction with both this doctoral study and the contributions of Sebastien Dandrifosse, an innovative motorized platform has been conceived and engineered. The platform and the sensor system used in this study have been supported and validated in the thesis of Sébastien Dandrifosse, and are therefore outside the scope of this current work.

Designed to navigate and carry cameras within trial fields divided into micro-plots, the platform initially employed over the initial three-year period was a stationary, manually operated structure. This setup necessitated the involvement of two individuals for operation and was used to capture vegetation data, as illustrated in Figure 3.1. This static system was eventually replaced by the development of a dynamic mobile platform. Extensive testing of this mobile platform was undertaken in 2021 to verify its ability to operate smoothly throughout the 2022 growing season.

The structural design of the mobile platform was based on a four-wheel motorized system. The mechanical construction of the chassis was performed in collaboration with the CRA-W. The implementation of the motorization and the embedded electronics to control the navigation and the data acquisition were carried out by the mechatronics laboratory of Gembloux Agro-Bio Tech. The platform is modular and can be easily dismantled into four parts for transportation and micro-plots adaptability. In order to avoid any interference caused by shadows, the sensors were mounted on a cantilever beam always oriented towards the sun. The maximum height of it was 2.25m. A tensioners were added to limit any vibration during data acquisition. Regarding the electronic design, each wheel is driven in propulsion and steering by a control unit connected via a CAN BUS network to a central unit. This assembly enables the regulation of the rotation speed and the orientation of the four wheels for linear trajectories, i.e. forward, backward and lateral displacements, and circular trajectories with different curvature radii. The navigation is remotely controlled by only one operator which is a significant improvement over the manned platform, particularly in terms of reducing the physical demands of the work. It also contributed to increase the data acquisition flow.

To further improve the navigation capabilities of the platform, a semi-autonomous navigation system is currently being develop using two complementary approaches. Firstly, a low-cost geo-referencing system such as RTK GPS will be utilized to accurately position the platform at specific points. This will ensure that the platform is always precisely located, even in challenging environments. Secondly, for simple trajectories like micro-plots in a straight line, cameras installed at the corners of the platform and positioned above the wheels will be used to capture the crop rows to fol-

low. This would allow for more efficient and accurate navigation, even in situations where the GPS signal may be weak or unreliable. By combining these two approaches, the platform could achieve a higher level of autonomy and precision in its navigation, leading to improved performance and efficiency.

In addition to the proposed navigation improvements, enhancing the mechanical structure of the platform could help it better withstand the rugged terrain it operates in. To achieve this, a stabilizer could be implemented to eliminate vibrations that the platform encounters during operation. This would improve the stability of the platform and prevent any potential damage to the sensors. By addressing these mechanical concerns, the platform would be better equipped to operate under challenging field conditions and collect accurate data without any disruptions.

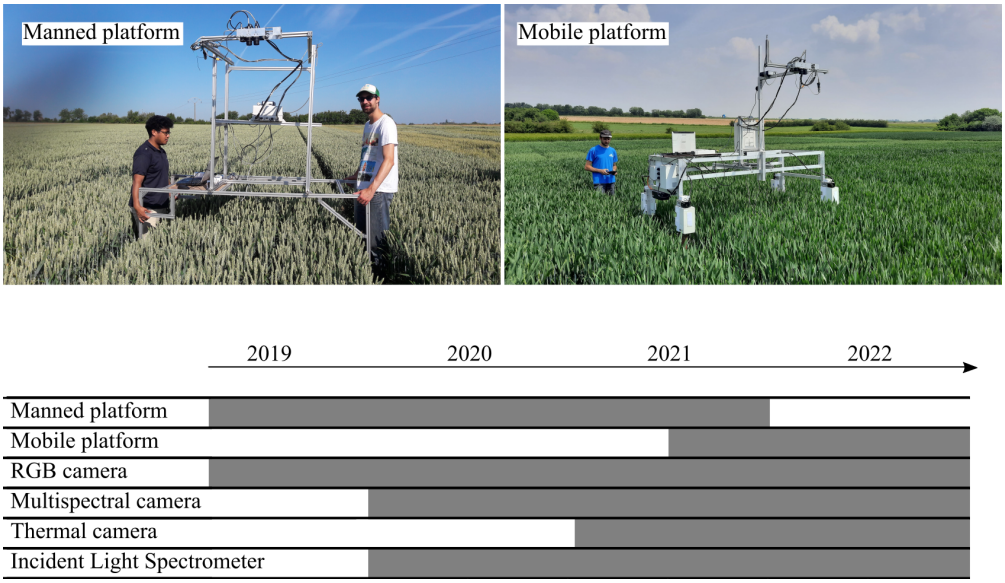

Figure 3.1: Image acquisition system timeline.

2. Data Acquisition System

2.1. Sensors

A multi-sensor system was developed during the course of the project PhenWheat and validate during the thesis of Sébastien Dandrifosse funded by a FRIA grant (Figures 3.1 and 3.2) [Dandrifosse, 2022]. In 2018, the data acquisition system consisted of two RGB GO-5000C-USB cameras (JAI A/S, Copenhagen, Denmark), equipped with a 2560 x 2048 pixels CMOS sensor and a LM16HC objective (Kowa GmbH, Düsseldorf, Germany), with a focal length of 16 mm. The cameras had a Horizontal

Field Of View (HFOV) of 44.3° and a Vertical Field Of View (VFOV) of 33.6°, and the aperture was set to f/4.0.

From the 2020 season, a multispectral camera Micro-MCA (Tetracam Inc., Gainesville, FL, USA), consisting of six monochrome cameras with 1280 x 1024 pixel CMOS sensors, complemented the RGB ones. Narrow band-pass optical filters centered at 490, 550, 680, 720, 800, and 900 nm were mounted on each camera, with a width of 10 nm except for the 900 nm band, which had a width of 20 nm. The lenses had a focal length of 9.6 mm and an aperture of f/3.2. The HFOV was 38.26° and the VFOV was 30.97°. However, the device was not yet fully mastered, and the 2019 data were not exploited in this thesis.

In 2021, a PI640 thermal camera (Optris GmbH, Berlin, Germany) was added to the system, equipped with a sensor of 640 x 480 pixels and covering a spectral range from 7.5 to 13 μm, with a focal length of 18.7 mm. The HFOV and VFOV were 33° and 25°, respectively. However, this sensor will not be used in the rest of this thesis.

Along with these cameras, an Incident Light Spectrometer (ILS) was also added to the set-up. It was an AvaSpec-ULS2048 (Avantes, Apeldoorn, The Netherlands) equipped with a cosine corrector. It had a bandwidth from 200 to 1100 nm with a resolution of 0.5 nm. The irradiance calibration was carried out at the manufacturer on March 23, 2020, and in winter 2021-2022.

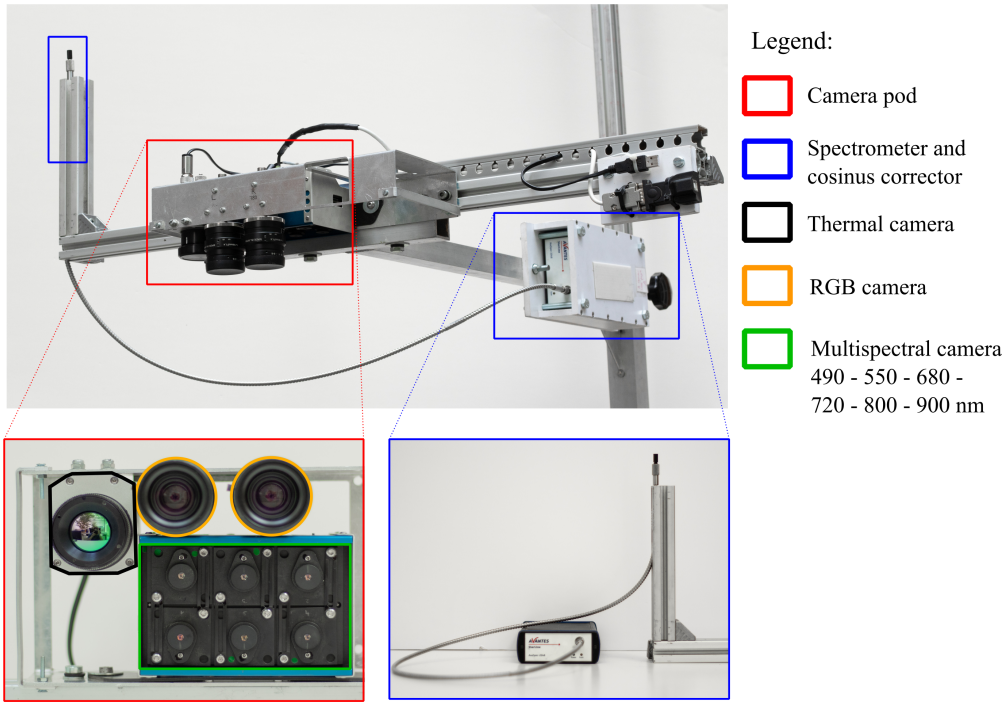

Figure 3.2: Sensor set-up.

2.2. Acquisition Settings

Through all the measurement campaigns from 2019 to 2022, from two to four scenes were acquired within each micro-plot depending on the trial. Nadir frames with the HFOV perpendicular to the wheat rows were taken by the cameras at a constant height 1.6 m above the canopy, except in 2019 where the height was 1 m. A homemade software was developed by the lab to control the system and enabled the setting-up and the triggering of the sensors. To ensure perfect synchronization, an external trigger from an Arduino micro-controller was used for the two RGB cameras. The spectrometer and the multispectral camera were triggered within one second due to technical compatibility.

From 2019 to 2021, the RGB cameras recorded 12-bit and 8-bit color resolution images, while in 2022, only 8-bit images were recorded. The 12-bit images were reduced to 8-bit images due to the requirements of existing image processing libraries by dividing pixel values by 16. A dichotomous algorithm was designed to find the highest exposure time with less than 0.1% of pixels saturated in the image. It was based on the manufacturer exposure. No change in ISO value was made. The images were saved in TIFF format on a laptop to avoid information loss and radial distortion was corrected using a chessboard and the OpenCV-Python library.

The exposure time of each channel of the multispectral array was determined using a master-slaves method. An auto-exposure mode was adjusted on the channel centered at 800 nm, considered as the master. The exposure time of the other channels were deduced by means of correction factors, which needed to be adjusted throughout the season. Four combinations of coefficients were used across the season, from tillering, which had a large amount of soil in the image, through greener canopy to maturity with yellow plants. The multispectral images were saved on SD cards and processed in the PixelWrench software to obtain 8-bit grayscale images.

The ILS recorded an average of three spectra at each image acquisition, which took less than one seconds. The spectral data underwent correction for dark noise and pixel non-linearity associated with exposure time, utilizing calibration data to compensate for these sensor and electronic-induced noise factors. A proper exposure time was manually set through the user interface to avoid saturation in the spectrum on each date of acquisition. Digital values were converted to irradiance data thanks to factory calibration.

3. Experimental Protocol

3.1. Field Trials

Field trials were conducted in Lonzée, Belgium (50°33'50" N and 4°42'00" E) on homogeneous deep silt loamy soil under a temperate climate over a period of four

cropping seasons (2019-2022). To introduce a distinctive contrast in both the canopy architecture and the fungal infection, aimed at rigorously testing the proposed phenotyping system, the experimental trials were meticulously structured. On one hand, variations were introduced solely in the nitrogen fertilization strategy, while on the other hand, a combination of fungal foliar disease protection levels and varying fertilization amounts was employed. A summary of the trials is provided in Table 3.1. The trials were set up with a seed drill after plowing. The rest of the crop itinerary was executed under optimal conditions to avoid additional stressors. All these tasks were carried out by the experimental farm of Gembloux Agro-Bio Tech. Thus, the first trial focused on nitrogen fertilization (F), with different levels and fractions described in detail in Table 3.2 and Table 3.3. The aim was to produce a range of wheat canopies to identify significant differences in wheat growth and yield. That was based on the nitrogen fertilization practices recommended in Belgium that suggest a total of 180kg/ha of nitrogen in three fractions of 60kg/ha at tillering (BBCH 25), stem elongation (BBCH30), and flag leaf (BBCH39), which may be adjusted based on the previous crop and soil nitrogen residues. The fertilizer used in this trial was 27% ammonium nitrate. The trials consisted of four complete randomization blocks for images and reference measurements, and four other replicates for final grain yield.

The second type of trial involved a bi-factor approach (FP), combining fertilization and fungicide treatments as described in Table 3.4. This trial was designed to investigate the interactions between plant nitrogen status and incidences of fungal diseases. Nitrogen input scenarios were similar to those in the fertilization trials, with four fungicide scenarios. The 0F scenario involved no fungicide application, while 1F had a single fungicide application at the flag leaf stage (BBCH 39). The 2F scenario involved fungicide application at the second node stage (BBCH 32) and at the heading stage (BBCH 55). The 3F scenario involved fungicide application at BBCH 32, 39 and at flowering (BBCH 65). Fungicide mixtures were composed of Triazole (Kerstrel, $1.25\,\mathrm{L\,ha^{-1}}$) at BBCH 30, Triazoles-Pyrazoles-carboxamides (Librax, $1.5\,\mathrm{L\,ha^{-1}}$) at BBCH 39 and 55, and Triazole (Prosaro, $1\,\mathrm{L\,ha^{-1}}$) at BBCH 65. No inoculation was performed, and the disease was allowed to occur naturally on chosen disease-sensitive cultivars. It was actually very annoying because some diseases may or may not occur out of control (see Chapter 5). The trials comprised four complete randomization blocks for images and reference measurements and four additional replicates for final grain yield for 21 and 22-FP. In 2019, no replicates were sown for grain yield.

3.2. Reference Measurements

Parallel to the image acquisitions, reference measurements were conducted in order to identify correlations between the numerical data and the conventional agronomic knowledge. These reference measurements are described in this section.

Table 3.1: Summary of the field trials.

Trial ID	Type	Variety	Sowing date	Sowing (grains.m^{-2})	Previous crop
19-FP	Fertilization × Protection	Smart	23/10/2018	250	Potato
20-F	Fertilization	Mentor	07/11/2019	250	Spinach
20-FP	Fertilization × Protection	Vertikal	05/11/2019	250	Spinach
21-F	Fertilization	Mentor	20/10/2020	275	Potato
21-FP	Fertilization × Protection	Vertikal	27/10/2020	300	Potato
22-F	Fertilization	Mentor	28/10/2021	300	Potato
22-FP	Fertilization × Protection	Bennington	28/10/2021	300	Potato

Table 3.2: Fertilization schemes of trials 20-F, 21-F and 22-F. The last column mentioned for which modality destructive reference measurements were performed (see Section 3.2.1).

Nitrogen inputs (kg N.ha^{-1}) at:			Total nitrogen	Destructive
Tillering	Stem elongation	Flag leaf	inputs(kg N.ha^{-1})	measurements
0	0	0	0	✓
30	30	30	90	✓
40	40	40	120	
50	40	55	145	
60	60	60	180	✓
80	40	60	180	✓
90	30	60	180	
105	105	105	315	✓

Table 3.3: Fertilization inputs of trial 19-F. The last column mentioned for which modality destructive reference measurements were performed (see Section 3.2.1).

Nitrogen inputs (kg N.ha^{-1}) at:			Total nitrogen	Destructive
Tillering	Stem elongation	Flag leaf	inputs(kg N.ha^{-1})	measurements
0	0	0	0	✓
30	30	30	90	✓
0	60	60	120	
60	60	0	120	
90	30	60	180	
60	60	60	180	✓
80	40	60	180	✓
105	105	105	315	✓

Table 3.4: Fertilization inputs and fungicide treatments of trials 20-FP, 21-FP and 22-FP. The last column mentioned for which modality destructive reference measurements were performed (see Section 3.2.1).

Nitrogen inputs (kg N.ha^{-1}) at:			Total nitrogen	Fungicide at	Destructive
Tillering	Stem elongation	Flag leaf	inputs(kg N.ha^{-1})	BBCH stages:	measurements
40	40	40	120	/	✓
60	60	60	180	/	
80	60	60	200	/	
100	80	80	260	/	✓
40	40	40	120	39	✓
60	60	60	180	39	✓
80	60	60	200	39	
100	80	80	260	39	✓
40	40	40	120	32 - 55	
60	60	60	180	32 - 55	
100	80	80	260	32 - 55	
40	40	40	120	32 - 39 - 65	✓
60	60	60	180	32 - 39 - 65	
80	60	60	200	32 - 39 - 65	
100	80	80	260	32 - 39 - 65	✓

3.2.1. In-season Destructive Measurements

Destructive samples of wheat biomass were taken at major phenological stages including stem elongation (BBCH 30), second node (BBCH 32), flag leaf (BBCH 39), flowering (BBCH 65), medium milk (BBCH 75), and maturity (BBCH 89) stages as shown in Figure 3.4. The growing season for each year is described in terms of Photo-Vernalo-Thermic Units (°C-days) (UPVT) detailed in Section 3.2.4. The focus was primarily made on the Fertilization trials (F) trials from 2019 to 2021, with destructive measurements at all those growth stages, and fewer measurements for the Fertilization and fungicide trials (FP) trials. Due to the COVID-19 pandemic, the first sample was not made in 2020. In contrast, in 2022, the FP trial was sampled at all growth stages, whereas the F trial had only three sample dates.

The sampling consisted in cutting fresh wheat plants at ground level to a length of 50 cm in three central contiguous rows. The biomass was manually divided into the stems, the flag leaves, the other leaves, and the ears. Each category was weighed and dried in an oven at 65°C until a constant weight was achieved. The dry matter (DM) was then weighed and sent to an external laboratory for nitrogen concentration determination using the Dumas method. Nitrogen uptake (Nupt) of each organ was obtained by multiplying the corresponding nitrogen concentration (%N) and DM. Nitrogen Nutrition Index (NNI) was also calculated based on the equation from [Justes, 1994] as

shown below:

$$NNI = \frac{\%N}{Nc}; Nc = 5.35(DM)^{-0.422} \tag{3.1}$$

where %N represents the actual nitrogen concentration and Nc is the critical nitrogen concentration corresponding to the crop DM (t/ha).

To determine LAI, plants were sampled by taking one row measuring 0.50 m in length. The leaves were separated from the stems, weighed, spread on a white paper using a transparent adhesive sheet, and scanned. An Otsu segmentation method was employed to isolate the leaves from the white background [Otsu, 1979]. The leaf surface area was calculated by summing the areas of the scanned paper sheets multiplied by the proportion of pixels segmented as leaf. Since this protocol was time-consuming, only five microplots with contrasting fertilization were selected for manual LAI measurements at each collection date. These LAI values were correlated with the associated fresh masses by means of a linear regression to predict the LAI of the other microplots. Each correlation had a really high correlation above 0.9, thus validate this method as a reference.

3.2.2. Grain Yield and Yield Components

The dried ear samples collected at maturity were split into grain and chaff using a hand harvesting machine. They were then weighed and sent to an external laboratory for nitrogen concentration determination, which allowed for the calculation of the grain nitrogen content (g/m2) and concentration. The grain and chaff nitrogen content were combined using a weighted average to retrieve the ear concentration at maturity.

Grain Yield (GY) was harvested using a combine harvester in four dedicated replicates, except for the 20-FP trial in which micro-plots harvested where also those with samples holes. Thus, the 20-FP yield was corrected by accounting for the surface that contained sample holes. Unfortunately, violent storms at the end of the 2021 season lodged most of the two trials for several weeks before harvest. Therefore, the GY obtained in 2021 was the result of manual samples that were taken earlier, which limited the effect of lodging but induce a measurement bias.

Thousand Kernel Weight (TKW) was determined on combine harvester samples using four replicates by counting 100 seeds with a seed counter and then weighting them. It was measured on a bare seed basis.

GY is defined by the multiplication of the three yield components: the TKW, the number of kernels per ear, and the ear density (see Chapter 4). While calculating the exact kernel count per ear can be complex, we can easily calculate it from the other factors we have measured.

Table 3.5: Scale for the visual scoring of wheat fungal diseases. The scale is based on three criteria: i) the affected foliar floor (L1 refers to the flag leaf, L2 to the second upper leaf, ...), ii) the average intensity of the infection on a leaf (Lo = low, M = medium and Hi = high), and iii) the repartition of the disease in the micro-plot or within the plants leaves.

		Affected foliar floor											
		L4 and <			L3			L2			L1		
	Average intensity	Lo	M	Hi	Lo	M	Hi	Lo	M	Hi	Lo	M	Hi
Repartition	<3 plants	1	1	1	1	1	1	1	1	1	1	1	1
	<10 plants	2	2	2	2	2	2	2	2	2	2	2	2
	10 <x <50 plants	2	2	2	2	2	2	3	3	3	4	4	4
	1 in 10 plants	2	2	3	3	3	4	4	5	6	6	7	8
	1 in 2 plants	2	2	3	3	4	4	5	6	6	7	8	8
	All the plants	2	2	3	3	4	5	5	6	7	7	8	9
	All the leaves	3	3	4	4	5	5	6	7	7	8	9	9

3.2.3. Disease Assessment

Trials were inspected at each image acquisition date, but disease grading was only performed when necessary. The three main diseases, namely Septoria tritici blotch (STB), Yellow rust (YR), and Brown rust (BR), were graded using a Visual Score (VS) for all treatments, described in Table 3.5. The grade is based on the average intensity of the disease on the highest affected foliar floor and for a certain number of plants. It is a fast scoring method that represents the global incidence of the disease of the micro-plot. It was re-scaled following the equation:

$$sVS = \frac{9 - VS}{8} \tag{3.2}$$

This equation scales the VS to scaled Visual Score (sVS) where a value of 0 signifies no disease, and a value of 1 corresponds to a very high level of disease pressure. For the trial 21-FP, scoring was performed four times on 16th, 25th of June, and 2nd, 9th of July due to low disease pressure. However, for the trial 22-FP, diseases appeared early in the season, so the trials were scored ten times on 19th, 25th of April, 2nd, 9th, 17th, 23th, 30th of May, 2nd, 13th, and 21st of June. It is worth noting that other trials showed symptoms of STB in the lower leaves but were considered negligible in terms of the date or intensity.

3.2.4. Field Observations and Environmental Data

During each data acquisition session, a variety of information was meticulously noted in a field notebook. This included the date, time, trial name, operators involved, relative exposure settings of the multispectral camera, sensor height, plant height, growth stage, reference measurement type (biomass sample or disease notation), weather conditions, and overall remarks. The global remarks section provided space to highlight any noteworthy field observations or encountered issues, such as pest problems or software bug.

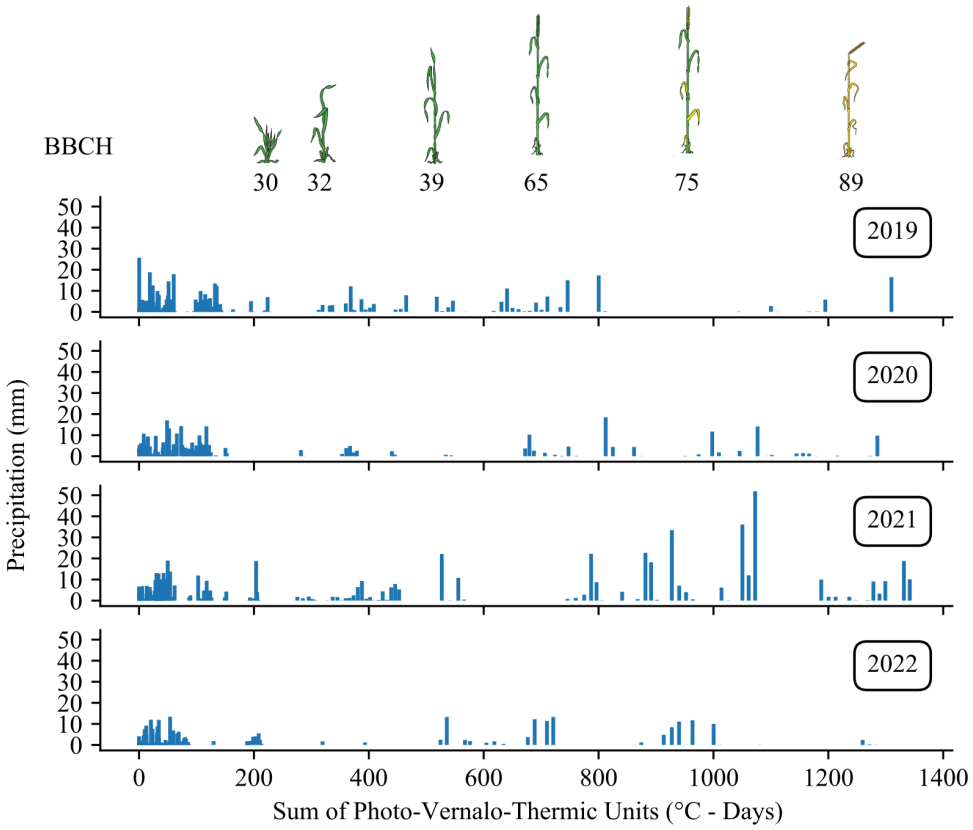

Figure 3.3: Overview of the season's precipitations in mm.

In addition to the simple weather comments, more advanced meteorological data were provided by CRA-W/Agromet.be from a measurement station located in Sombreffe at five kilometers from the trials. This station recorded hourly data on air temperature and precipitation. As shown in Figure 3.3, the seasons vary in terms of precipitation that have induced different level of diseases. 2019 and 2021 were quite dried during the growing season. 2021 was very wet and included violent storms late in sea-

son. 2022 was a bit wet during spring to bring high level of diseases, but in a steady way that did not disturb data acquisition at all. The thermal time, UPVT, used in this thesis was based on the research of [Duchene et al., 2021] which took into account a slowdown factor for photoperiod and vernalization.

3.3. Collecting Data

Starting in 2019, a substantial volume of images was procured for analysis. The dates of image acquisition and the corresponding destructive measurements conducted within those years are outlined in Figure 3.4. In the initial year, 2019, the image count was relatively limited due to project initiation and the need for a gradual launch. However, beginning from the subsequent years, the objective was to acquire data every 10 to 15 days, with careful attention to weather forecasts. Nonetheless, in both 2021 and 2022, there were extended intervals without data collection. The latter year witnessed the impact of severe storms that caused lodging in the two trials, thereby disrupting data collection for several weeks leading up to the harvest. Additionally, the COVID-19 pandemic hindered data acquisition at the beginning of the 2020 season.

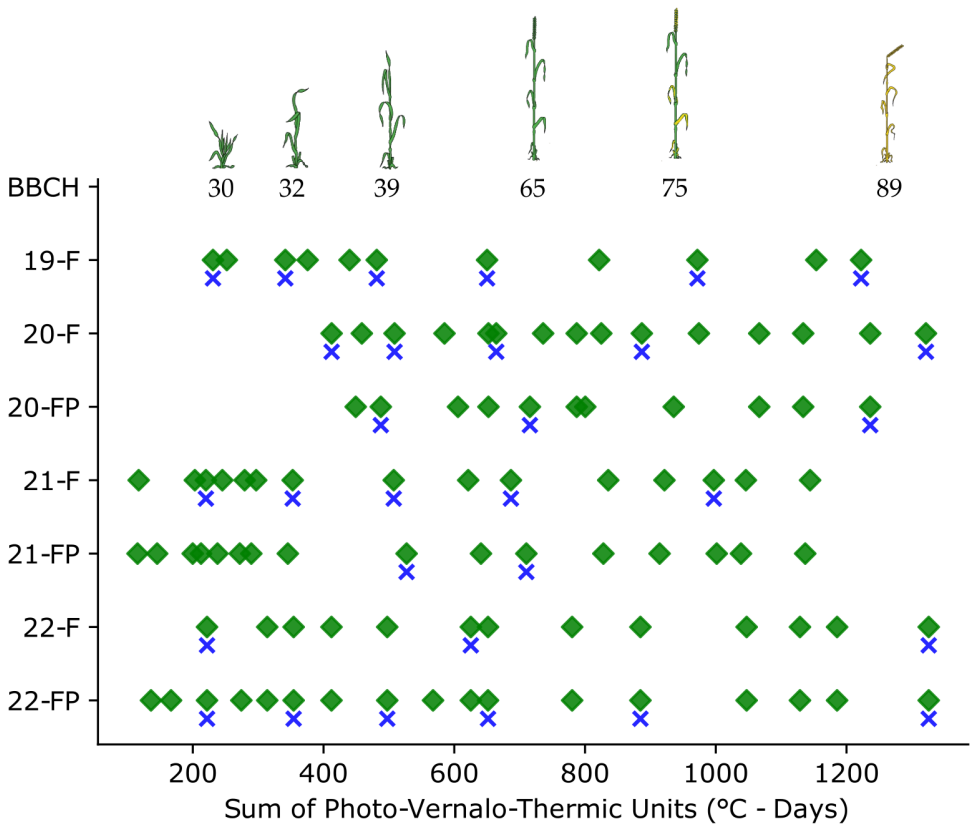

Figure 3.4: Overview of the data collection for each trial across the four seasons: images (green diamonds) and destructive measurements (blue crosses).

4. Image Data Processing

4.1. Synopsis

This section provides a comprehensive overview of the raw image processing techniques employed in this study. These methodologies were primarily developed within the framework of Sébastien Dandrifosse's doctoral project, carried out in collaboration with the PhenWheat project and the proposed thesis. The central objective was to integrate data captured by the multi-sensor system, detailed in Section 3.3, to derive both depth and spectral information at the scale of individual plant organs, subsequently extracting pertinent agronomic traits.

The image processing pipeline, depicted in Figure 3.5, encompassed several key steps, including plant-soil segmentation, canopy height mapping via stereovision [Dandrifosse et al., 2020], computation of bi-directional reflectance factors (BRF) [Dandrifosse et al., 2021], and pixel-wise registration of all images [Dandrifosse et al., 2021]. While this section furnishes a solid foundational understanding of each procedure, for more exhaustive insights, readers are encouraged to delve into the respective research papers.

4.2. Plant-Soil Segmentation

The presence of soil within images can introduce noise to the accurate representation of the crop signal [Daughtry et al., 2000; Wang et al., 2022b]. A spectrum of image segmentation techniques has been proposed to mitigate or eliminate this effect, encompassing both elementary methods like grayscale image thresholding [Meyer and Neto, 2008], to more sophisticated approaches such as the use of SVM [Hamuda et al., 2017] or deep learning methodologies (refer to Chapter 5 for details).

The devised approach hinges upon utilizing the 800 nm image for two principal reasons. Firstly, it served as the reference image during the registration process (Section 4.4), thus remaining unaffected by any deformation. The inclusion of a vegetation index like NDVI might not be suitable in this context, as it necessitates aligning two images, thereby introducing potential deformations. Secondly, owing to the relatively high reflectance of plants in the NIR region compared to soil, segmentation of plants is more feasible even during advanced growth stages. As depicted in Figure 3.6, plant pixels exhibit higher values compared to soil pixels. Therefore, a simplistic threshold-based method was adopted. The threshold demarcating the two classes was determined by identifying the first minimum in the pixel value histogram (Figure 3.6). Computation of this threshold was facilitated by the scikit-image library [Van der Walt et al., 2014], which locates the local minimum within pixel values ranging from 25 to 50. In scenarios where no minimum was discernible, a Gaussian mixture model with three components was employed to ascertain the threshold, found at the intersection of the initial two components.

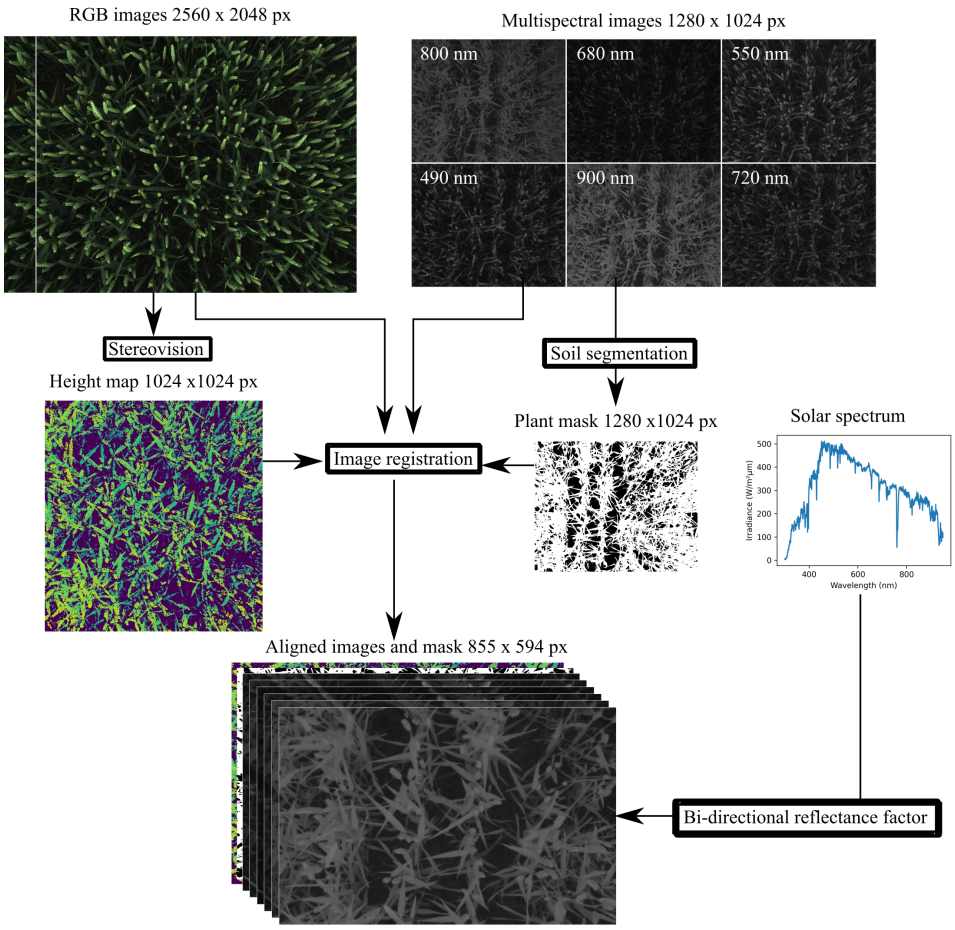

Figure 3.5: Overview of the image processing techniques.

In instances characterized by intense direct sunlight, soil pixels might exhibit higher values than plant pixels, potentially leading to misclassification. To mitigate this, the 490 nm image was enlisted to identify strongly illuminated soil pixels, subsequently excluded based on a 95th percentile threshold. This supplementary step was exclusively executed when the cloudiness index surpassed 0.90, a derived from ILS data through the subsequent equation:

$$C_t = 1 - \frac{E}{E_0 cos(z)} \tag{3.3}$$

where E represents solar irradiance (W/m²) within the ILS spectral measurement range, E_0 corresponds to the solar constant (1360 W/m²), and z signifies the sun's zenith angle.

Though no dedicated study has been conducted to formally evaluate this method,

its efficacy is evident from the visual assessment of the resultant masks, showcased in Figure 3.7. These masks exhibited satisfactory visual accuracy across various scenarios, substantiating their capability to derive canopy cover by computing the ratio of plant pixels to total pixels. As part of a broader investigation (Chapter 5), a novel technique based on RGB imagery and employing deep learning was also explored.

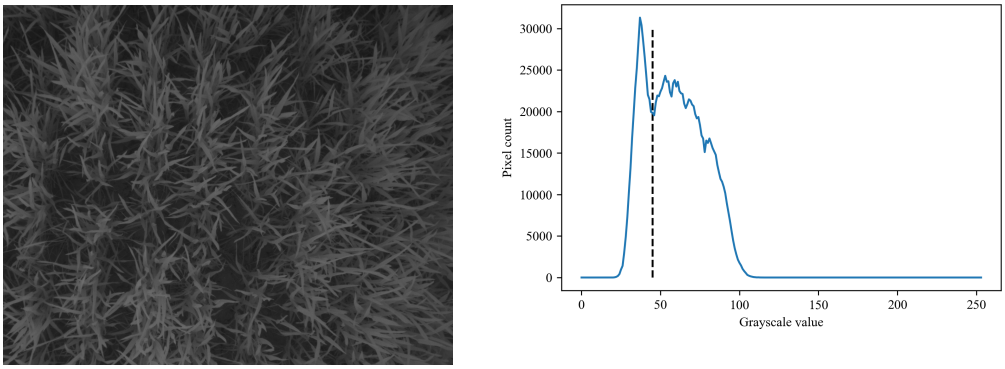

Figure 3.6: Example of a 800 nm image with its corresponding histogram. The black dashed line represents the minimum found by the proposed algorithm.

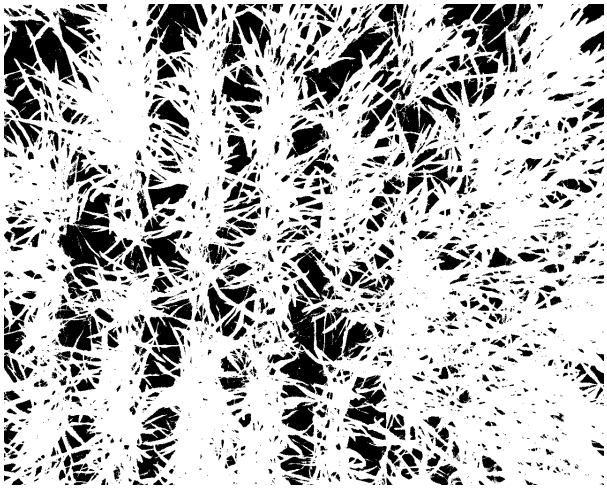

Figure 3.7: Result of the segmentation on the 800 nm image.

4.3. Height Map from Stereovision

This section relies on the following published paper [Dandrifosse et al., 2020]:

Dandrifosse S., Bouvry A., Leemans V., Dumont B. & Mercatoris B., 2020. Imaging wheat canopy through stereo vision : overcoming the challenges of the laboratory to

field transition for morphological features extraction. Front. Plant Sci. 11(February), 1–15. DOI:10.3389/fpls.2020.00096.

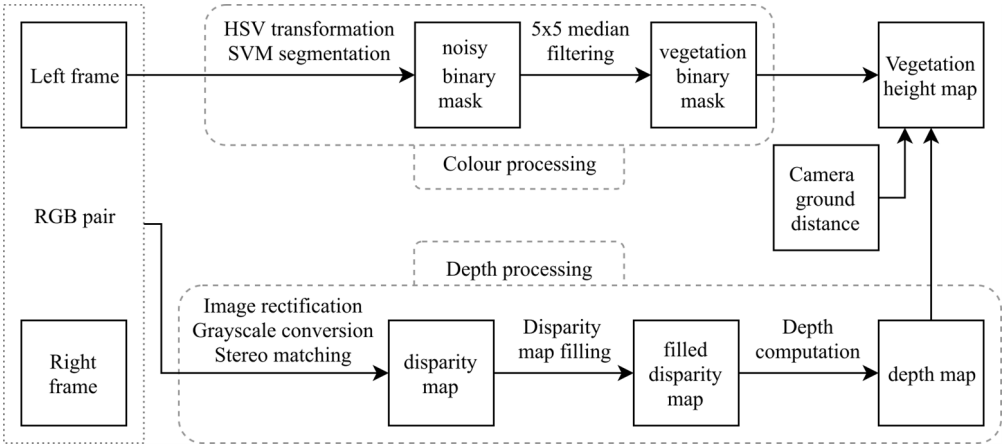

Figure 3.8: Stereovision pipeline (adapted from [Dandrifosse, 2022]).

Stereovision is a process that enables the acquisition of 3D information using two closely placed cameras with different viewpoints, similar to human vision. This method provides a depth map that assigns a height value to each pixel, enabling the extraction of height statistical descriptors, computing leaf angle distribution, or other geometrical traits. Initially developed in the laboratory [Tilneac et al., 2012], stereovision has recently been applied in field conditions [Müller-Linow et al., 2015].

The stereovision process relied on the use of the two RGB cameras of the multi-sensor setup. Before each campaign, the two-camera system was calibrated using a black and white chessboard divided in 40 x 40 mm². This allowed the determination of the intrinsic and extrinsic camera properties used to correct image distortion and in the stereovision process. The entire process was carried out using the OpenCV-Python library (version 4.5.3.56). The 12-bit images acquired in 2019, 2020, and 2021 were converted to 8-bits to be compatible with this library.

Following the pipeline shown in Figure 3.8, the left image was used to segment the green canopy using an Excess Red vegetation index (ExR) [Meyer and Neto, 2008] with a threshold of 0.05, which was determined empirically. This yielded a satisfactory segmentation that was adequate for deriving plant height statistics.

The first step in the stereovision process is the rectification, which aims at projecting both images into a common y-coordinate system using the extrinsic parameters. The images are then converted to grayscale and reduced to 1280 x 1024 pixels using 2 x 2 averaging pooling. The second step is the stereo matching, which involves finding corresponding pixels in both rectified images using the Semi-Global Block matching algorithm. To optimize pixel matching, the algorithm considers the image from mul-

tiple directions to find the best matches. The disparity map is obtained by calculating the difference between the x-coordinates of both matched pixels. Since both RGB images may not match at every pixel, a couple of filtering processes can be applied. In this case, the Weighted Least Squares filter was chosen to smooth and fill the gaps.

Depth is inversely proportional to disparity considering the focal length and the baseline. This results in a depth map of the scene. From there, a height map is computed by subtracting the height of the camera from the depth map. The green canopy mask can be applied to the height map to extract the height histogram of the crop.

The height map is the primary output of the stereovision process, with the option of using a filling filter to enhance the results (Figure 3.9). This filtered image is a valuable source of information and will be used in various specific tasks throughout this thesis. Additionally, [Dandrifosse et al., 2020] proposed using the height map as a 3D point cloud to estimate the mean tilt angle of leaves, the leaf angle distribution and the leaf area index.

Besides the height map, statistical descriptors of the observed scene can be computed using stereovision. For instance, in Figure 3.10, percentiles 75 and 95 were compared to manual measurements. For the trials 18-F (similar to the 19-F trial) and 19-F, the manual reference point was set at the flag leaf insertion, and measurements were taken using a tape measure. Pearson's correlation coefficients between the manual measures and the median, percentile 75, and percentile 95 of automatic measures were 0.66, 0.58, and 0.39, respectively. These results demonstrate that this method can provide useful height descriptors of a canopy. It also questions the relevance of comparing these electronic measures with traditional measures.

4.4. Image Registration

This section relies on the following published paper [Dandrifosse et al., 2021]:

Dandrifosse S., Carlier A., Dumont B. & Mercatoris B., 2021. Registration and Fusion of Close-Range Multimodal Wheat Images in Field Conditions. Remote Sens. 13(7), 1380. DOI:10.3390/rs13071380

The employed multi-sensor system comprised three distinct cameras. Due to their proximity to the canopy and varying optical axis positions, directly combining their individual images posed a challenge in extracting intricate traits at the pixel level. While data fusion techniques enable merging of resultant traits, the fusion of source data, the images, presents a distinct endeavor. Such fusion can amplify knowledge and comprehension of the plant system by augmenting the data at each pixel [Jiang et al., 2018]. To tackle this complexity, an automated registration approach was devised for close-range multimodal wheat canopy images captured under field conditions. This method harnessed local deformations, drawing inspiration from analogous challenges encountered in the medical domain [Sotiras et al., 2013].

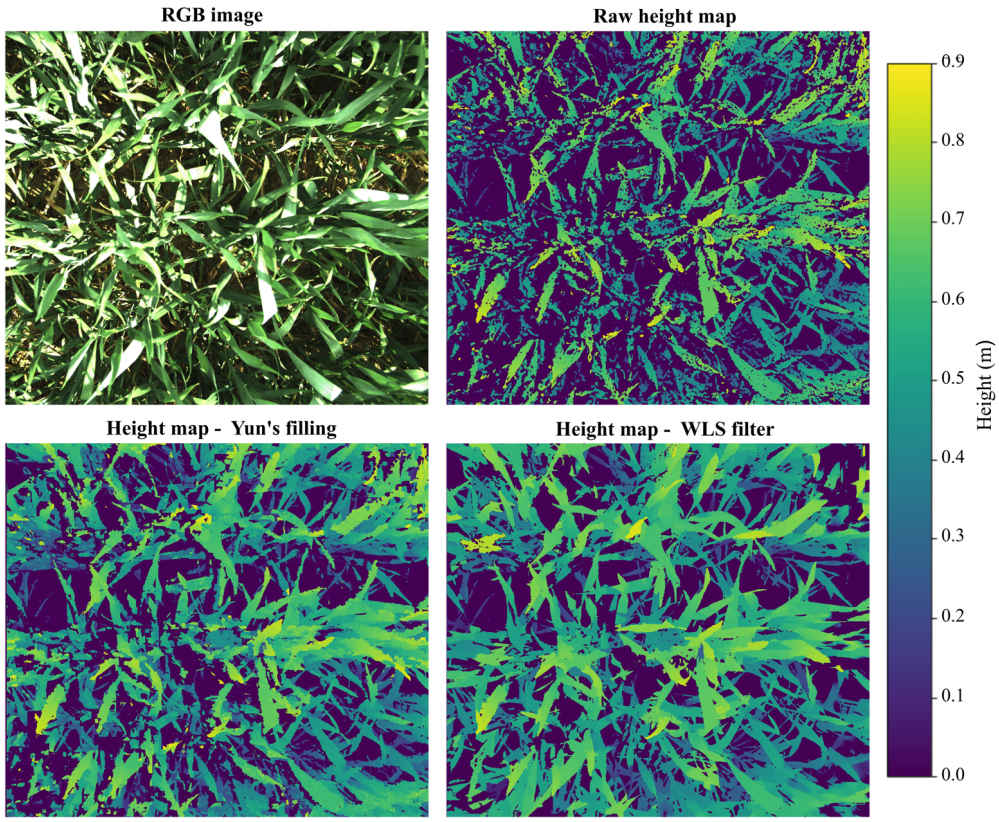

Figure 3.9: This figure shows the RGB image cropped to the zone commonly observed by the two cameras, the raw height map, the height map filled with the algorithm of Yun, and the height map filled with the WLS method. The dark blue zones, at 0 on the scale, corresponded either to soil or to plant pixels for which the height couldn't be computed (from [Dandrifosse, 2022]).

This section encompasses the entire multi-sensor system depicted in Figure 3.2. For registration purposes, the 800 nm lens of the multispectral camera was designated as the master, selected due to its central position within the sensor setup and its proficiency in plant identification. Furthermore, the nearest RGB camera, i.e., the left one, was chosen.

The primary step of the registration process involved applying a global transformation based on the canopy distance. Affirmative results from laboratory experiments validated the application of an affine transformation, with the distance obtained through stereovision serving as input. Following a meticulous laboratory calibration procedure, images were registered using this method, yielding precise alignment. However, the global transformation alone was inadequate in accurately fusing all pix-

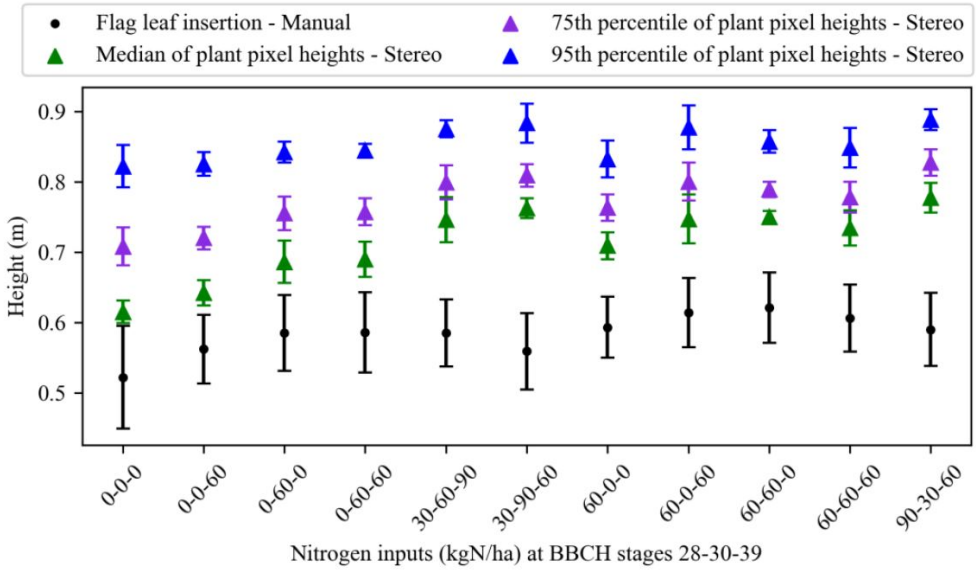

Figure 3.10: Comparison of automatic and manual measurements of canopy height at flag leaf stage. The measurements were performed on May 24th, 2018 on eleven micro-plots from the 18-F trial, each in a different fertilization scenario (from [Dandrifosse, 2022]).

els due to challenges arising from occlusion and parallax effects.

Consequently, subsequent to the initial pre-alignment, a secondary alignment was performed using an image-based method capitalizing on the inherent resemblances between the content of the slave and master images. Among the methodologies evaluated by [Dandrifosse et al., 2021], the B-SPLINE approach was adopted. This technique maximized a similarity metric between the master and slave images, employing the normalized mutual information metric. It subsequently executed a localized transformation of slave images based on a 3rd-order (cubic) B-spline model, employing a grid spacing of 16 units. The implementation of this technique leveraged the Elastix library, encapsulated within the pyelastix (version 1.2) Python library.

Ultimately, all images were meticulously cropped to dimensions of 855 x 594 pixels, capturing a common region that minimized potential distortions occurring at the image peripheries. Furthermore, a prudent recommendation included the erosion of the plant mask to ensure retention of solely the object of interest.

The registration results of the B-SPLINE and other tested methods are presented in Table 3.6. The B-SPLINE outperformed all other methods with a very low error of about 1.9 mm and 2.0 mm for the 900 nm and RGB images, respectively, in terms of the error on manual control point. It also produced the best results for plant mask error, but it had a high computation time. Despite the good accuracy achieved, some distortions occurred in the images, especially in situations with moderate to high wind.

Table 3.6: Comparison of the registration methods. The comparison is based on three criteria: the average computation time to register one image of all the cameras, the control point error and the plant mask error. For the RGB images, some methods were discarded (NA values) because they did not yield a plausible alignment for all the test images (from [Dandrifosse, 2022]).

Method	Average time (s)	Control point error (mm)		Plant mask error (%)
		900 nm	RGB	900 nm
SIFT	4	3.7	NA	9.7
SURF	6.2	3.6	NA	9.5
ORB	1	5.5	NA	10.6
A-KAZE	2.7	3.4	NA	10.1
DDTM	2.6	5.2	NA	11.7
DFT	41.3	3.9	NA	9.7
ECC	21.9	3.2	3	9.8
B-SPLINE	176.7	1.9	2	7

This was due to the asynchronous acquisition of scenes, which could result in relative displacement of the wheat leaves. To address this issue, a 5x5 matrix was used to erode the plant mask. Each image registration vectors were saved that allow the registration of other related image such as the height map or the soil segmentation mask.

4.5. Bi-directional Reflectance Factor

This section relies on the following published paper [Dandrifosse et al., 2021]:

Dandrifosse S., Carlier A., Dumont B. & Mercatoris B., 2022. In-Field Wheat Reflectance: How to Reach the Organ Scale? Sensors 22(9), 3342, DOI:10.3390/s22093342.

The ability to acquire crop health information from reflectance data has been introduced in Chapter 2. However, the acquisition, analysis, and correlation of such data with agronomic factors present challenges outdoors due to the dynamic interplay of canopy anatomy, physiology, and incident light [Comar et al., 2012]. These frequent variations during in-field data collection pose hurdles in securing accurate measurements. While the utilization of vegetation indices, like the NDVI, is a straightforward approach to address these fluctuations, it has its limitations in signal exploitation and insight generation. To surmount these constraints and enhance reflectance data analysis, an ILS was employed to normalize the data with respect to varying illumination conditions.

Enabling this approach necessitated the creation of camera response curves as the first step. These curves establish a connection between digital numbers and the corresponding exposure ($H(\lambda)$), accounting for the intricacies of the camera's electronics and optics. To formulate these curves, a known reflectance panel was captured by the camera at varying exposure times. Furthermore, the ILS recorded the solar spectrum, integrated over relevant wavelength bands, and amalgamated with exposure values to formulate the camera response curves.

Consequently, each individual camera had its unique relation, facilitating the retrieval of $H(\lambda)$ from digital numbers, thereby enabling the computation of crop reflectance according to the formula:

$$\rho\left(\lambda\right) = \frac{\pi * H(\lambda)}{E(\lambda) * t(\lambda)} \tag{3.4}$$

where $H(\lambda)$ signifies exposure at the camera lens (J/m² sr), E represents incident irradiance (W/m²), and $t(\lambda)$ denotes exposure time (s). Notably, the term "reflectance" commonly used actually refers to Bi-directional Reflectance Factor (BRF), which accommodates the directional nature of light.

An experiment conducted from 9 a.m. to 5 p.m. on six different dates assessed the influence of the sun zenith angle on BRF measurements using the multi-sensor system. This assessment aimed to isolate the impact of the sun zenith angle, assuming no alterations in plant physiology or material dynamics throughout the day. The methodologies delineated in this chapter, encompassing soil segmentation, image registration, and ears segmentation (Chapter 4), were employed to exclusively derive wheat leaf BRF.

As depicted in Figure 3.11, the average evolution of each BRF across the day unveils distinct patterns. In the majority of instances, BRF exhibited a morning surge followed by a decline and eventual stabilization. This trend was particularly pronounced for the initial April date, which had a lower zenith angle. Two dates demonstrated consistent measurements corresponding to overcast days. However, this does not unequivocally imply that the other dates enjoyed clear blue skies; cloud passages might have occurred without impacting measurements. Overall, data collection around noon is prudent, with observations indicating that a zenith angle surpassing 55° ensures precision in measurements. Consequently, this approach was consistently applied to all multispectral images throughout the thesis, typically acquired between 11 a.m. and 3 p.m.

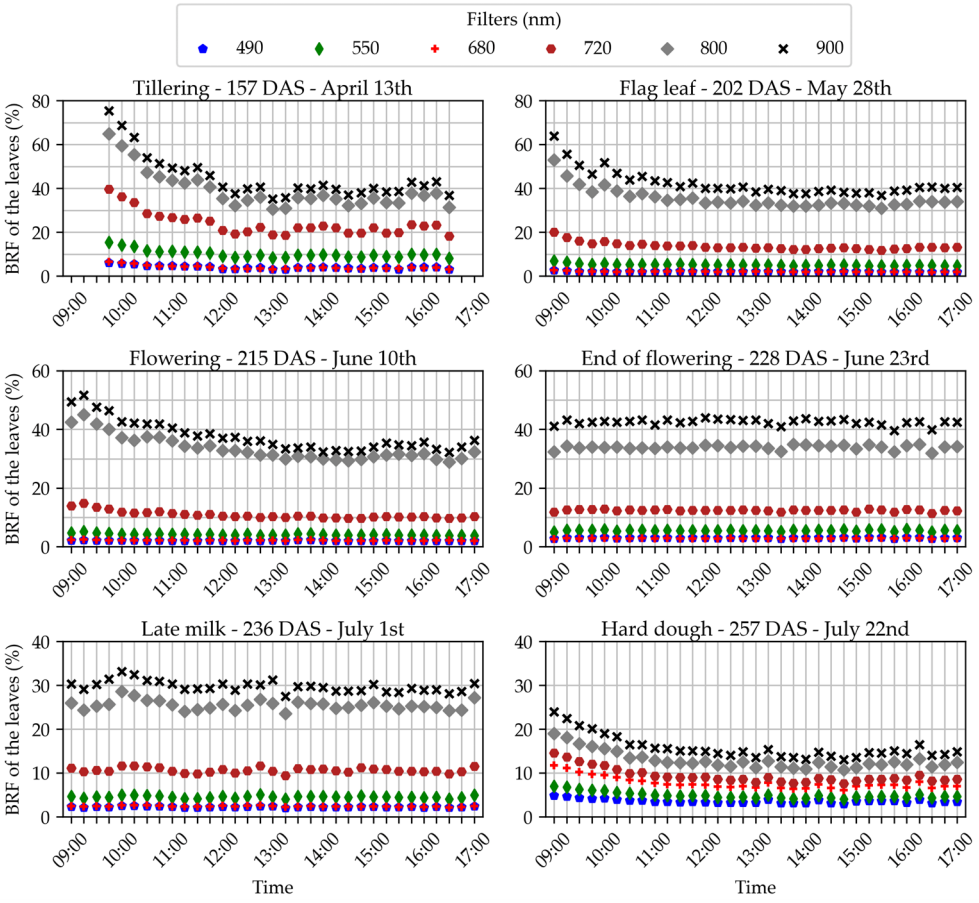

Figure 3.11: Average bi-directional reflectance factor (BRF) measured on the leaves throughout the day. Each subplot is dedicated to an acquisition date, designated by the development stage of the crop and the number of days after sowing (DAS). The evolution of the measured BRF is represented for the six different optical filters of the cameras, indicated by a color and symbol code on each subplot ([Dandrifosse et al., 2022a]).

4

Detection and Segmentation of Wheat Ears

1. Synopsis

Recent advancements in sensing technologies, coupled with significant enhancements in computing power, have assist in a new era of non-destructive data extraction from crop canopies throughout their growth stages. Proximal and remote sensing techniques now leverage high-resolution imagery obtained from cameras mounted on terrestrial vehicles or low-altitude UAVs. These technologies have enabled the discrimination of various wheat plant components within the field, such as leaves, stems, and ears. Recognizing the distinct developmental patterns and functional roles of these components is vital for accurately interpreting sensor data in high-throughput phenotyping strategies.

This chapter focuses on the automated segmentation of wheat ears and introduces two sequential methodologies that were developed for this purpose. The initial approach involves the classification of simple linear iterative clustering (SLIC) superpixels through feature extraction from fused RGB and multispectral images [Carlier et al., 2022]. In contrast, the second approach employs deep learning techniques to detect and segment ears within RGB images [Dandrifosse et al., 2022b]. The core aim of this study was to comprehensively compare the performance of both methods using a custom-designed evaluation framework. Furthermore, this investigation sought to underscore the remarkable capabilities of the deep learning approach in estimating ear density, thereby presenting a departure from conventional manual counting techniques.

2. Current State of the Art

The existing knowledge in this field underscores the significant influence of ear presence within the canopy on UAV-scale image signals, as revealed by studies conducted by Anderegg et al. [2020]; Li et al. [2021a]; Prey and Schmidhalter [2019b]. An additional insight, emphasized by Liu et al. [2021], pertains to the scarcity of research concerning light interception patterns during the post-flowering phase; a period wherein ears assume a greater role in light interception. This knowledge gap is particularly significant due to the pivotal role of canopy light interception in determining final yield, with the contribution of ear photosynthesis to grain filling remaining a subject not fully comprehended [Maydup et al., 2010]. Furthermore, the density of ears serves as a reliable proxy for one of the primary yield components. Although traditionally assessed manually using a wooden stick in the field, the adoption of automated methods for head density measurement promises enhanced accuracy and alleviates the complexities of physical measurement. Notably, a more focused analysis on ears offers improved insights into specific wheat diseases such as Fusarium head blight, which exclusively impact reproductive organs. Consequently, the automated segmentation of ear images emerges as a pivotal step within the domain of computer vision. This task, however, is laden with challenges, ranging from organ overlap to the intricate variations in color

and architecture induced by diverse growth stages, cultivars, and lighting conditions.

In the realm of computer vision, the current cutting-edge techniques for ear detection predominantly revolve around deep learning methods, specifically leveraging CNN. This paradigm shift is largely attributed to the advent of the Global Wheat Head Detection dataset V2 (GWHD) [David et al., 2021], which triggered a surge in research endeavors. Notably, the community witnessed nearly forty research papers published by April 2023, reflecting the dynamic response to this dataset's availability. While these studies yield commendable results in ear detection, they frequently limit themselves to bounding box delineations around the ears, omitting the provision of segmentation masks or the computation of ear density.

Before the advent of the GWHD dataset, deep learning approaches for ear detection were comparatively underdeveloped and accompanied by certain drawbacks. These methods necessitated copious amounts of training data, which translated into arduous and costly endeavors. Moreover, the underlying intricacies of deep learning processes remain partially veiled Singh et al. [2018], rendering them challenging to configure and execute. In light of these challenges, alternative strategies for ear detection and segmentation have surfaced. For instance, Fernandez-Fernandez-Gallego et al. [2019] explored thermal imagery for ear counting, though the method exhibited limitations in robustness . Meanwhile, Sadeghi-Tehran et al. [2019] achieved favorable counting outcomes through a hybrid model integrating superpixels and deep neural networks. Textural feature-based machine vision approaches have also demonstrated efficacy in ear counting and segmentation, particularly under diffuse lighting conditions [Cointault et al., 2008; Fernandez-Gallego et al., 2018]. Additionally, wheat ear recognition from multisensor data has been pursued using hand-crafted machine vision techniques [Zhou et al., 2018]. However, these prior investigations were often constrained by limited acquisition dates, and the assessment of segmentation performance remained infrequent.

3. Machine Learning Approach

This method furnishes sufficient information to comprehend the process, which is more detailed in the following published paper [Carlier et al., 2022]:

Carlier, A., Dandrifosse, S., Dumont, B. & Mercatoris, B., 2022. Wheat Ear Segmentation Based on a Multisensor System and Superpixel Classification. Plant Phenomics 2022, https://doi.org/10.34133/2022/9841985.

Following the preprocessing steps detailed in Section 3.4 Chapter 3, the ear segmentation process was executed through the application of a superpixel classification technique. Here, a superpixel denotes a collection of pixels sharing common attributes such as color and spatial characteristics. The concept of utilizing superpixels for ear

segmentation has been explored in previous research by Ma et al. [2020]; Sadeghi-Tehran et al. [2019]. To generate these superpixels from RGB images, we employed the Simple Linear Iterative Clustering (SLIC) algorithm, a variant of k-means clustering. This algorithm harnessed five features from the CIELAB color space in addition to pixel coordinates. The use of CIELAB is advantageous for detecting subtle color differences. The implementation was executed using the Python 3 scikit-image library (version 0.17.2). Key parameters of the SLIC algorithm were optimized to accommodate all crop development stages, encompassing the approximate superpixel count, compactness, and maximum k-means iterations. Through experimentation, values of 1500, 10, and 30 were determined for these parameters, striking a balance between overly large or overly small superpixels.

Each generated superpixel was subsequently classified as either an ear or non-ear. Three distinct machine learning algorithms were tested for this purpose: Random Forest (RF), Multilayer Perceptron (MLP), and Support Vector Machine (SVM). These algorithms were fed a comprehensive array of features, including average values of superpixels across the six monochrome channels from the multispectral camera, normalized RGB channels, Hue Saturation Value color space (HSV) values, and fifteen vegetation indices (Table 4.2). In addition, two non-image-related features, Days After Sowing (DAS) and the cloudiness index (refer to Equation (3.3)), were incorporated to account for wheat growth stage and lighting conditions, respectively. Prior to classification, these features were standardized using a standard scaler technique involving mean removal and standard deviation scaling. Classifier hyperparameters were then fine-tuned based on these 29 features, with accuracy serving as the evaluation metric.

To refine feature selection and minimize noise from redundant attributes, a sequential backward feature selection procedure was performed. In each iteration, the algorithm generated all possible feature subsets of size n-1 (where n denotes the total feature count) and subjected each subset to 5-fold cross-validation on the training set. The feature with the least contribution to the highest-accuracy subset was removed. This iterative process continued until performance stabilizes. Subsequently, a second round of hyperparameter optimization was conducted utilizing the selected features.

Constructing the classification dataset involved labeling superpixels into two categories: leaves and ears. The labeling procedure employed RGB images from the 20-F and 21-FP trials and was executed using an online machine learning platform available at https://www.apeer.com. Labeled regions were manually delineated using a brush-like pointer, as illustrated in Figure 4.1. To ensure dataset diversity and class balance, a comparable amount of data was allocated to both leaves and ears (46% earmarked as ears). These labeled regions were then translated into labeled superpixels, with superpixels containing at least 10% labeled pixels of a class being assigned to that class. Superpixels housing labeled pixels for both classes or having fewer than two pixels were excluded. Rigorous attention was given to selecting images representative of di-

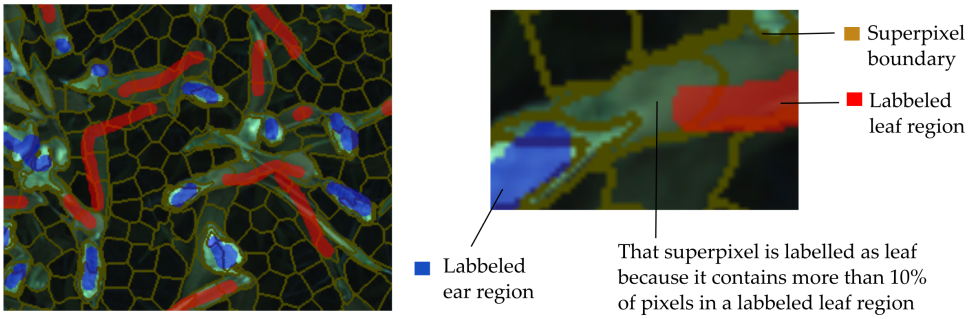

Figure 4.1: Illustration depicting the superpixel labeling process. Manual labeling was performed using a brush-like pointer.

verse acquisitions, spanning nitrogen input, growth stages, and lighting conditions. A minimum of two images from each date were chosen for labeling, yielding a dataset of 43 images. This dataset comprised 15,765 superpixels, randomly partitioned into an 80% training set and a 20% validation set.

4. Deep Learning Approach

This method furnishes sufficient information to comprehend the process, which is more detailed in the following published paper [Dandrifosse et al., 2022b]:

Dandrifosse S., Ennadifi E., Carlier A., Gosselin B., Dumont B. & Mercatoris B., 2022. Deep learning for wheat ear segmentation and ear density measurement: From heading to maturity. Comput. Electron. Agric. 199(June). DOI:10.1016/j.compag.2022.107161.

The second approach utilized a deep learning method to detect and segment the wheat ears. The state-of-the-art object detection algorithm YOLOv5 was employed along with images from the Global Wheat Head Detection dataset V2 [David et al., 2021] containing more than 275 000 labeled wheat heads, supplemented with additional images from the 20-F and 20-FP trials. To accommodate the size of the images, supplemented images were divided into four sub-images of 1280 × 1024 pixels, and each sub-image was resized to 1024 × 1024 pixels. The LabelImg tool (https://github.com/tzutalin/labelImg) was used to label the sub-images, resulting in two labeled datasets: a validation dataset and a dataset to augment the GWHD V2 for model training. The validation dataset consisted of around forty randomly selected sub-images from each acquisition date and trial, resulting in 64,091 labeled ears across 566 sub-images. The training set included images from growth stages not sufficiently represented in the GWHD V2, specifically heading and maturity. One hundred and

three sub-images were randomly selected from two dates at the heading stage (June 3rd for trial 20-F, June 2nd for trial 20-FP), and thirty-two sub-images were randomly selected from three dates at the maturity stage (July 29th for trial 20-F, July 13th and July 22nd for trial 20-FP). The training set contained a total of 13,566 labeled ears across 135 sub-images.

The training process involved four stages. In the first stage, the GWHD V2 was used to train the model. In the second stage, the trained model was employed to predict ear bounding boxes for all images taken in 2020, and the predicted boxes were recorded as pseudo-labels. In the third stage, the GWHD V2 labels and pseudo-labels were utilized together to retrain the model. Lastly, transfer learning was applied to enhance the model using the labeled training set obtained from our 2020 images. Consequently, the most recent model weights were based on reliable data, ensuring its credibility.

The DeepMAC neural network model was then used to segment wheat ears within the bounding boxes without the need for manual training masks [Birodkar et al., 2021]. The pre-trained DeepMAC model had strong generalization abilities, and ear masks were generated for each sub-image. The sub-masks were then transformed back to the original sub-image size to form a mask of the same format as the original RGB image. However, parts of the ears cut at the sub-image junctions did not match to the pixel, so a filling algorithm was used to consider pixels between two ear pixels from either side of a junction as belonging to the ear mask. While there was a risk of regrouping ear parts that did not belong to the same ear, the filling algorithm proved to be useful.

Finally, the ear density, expressed in ears per square meter, was calculated by dividing the number of ears in the image by the area of the image, i.e., the image footprint at ear height. To generate a map of ear depths, the ear mask obtained from the segmentation step was applied to the depth map obtained through stereovision. Since the ears were mostly vertical, the depth points were considered to be located at the tops of the ears. To estimate the image footprint in the middle of the ear layer, the median of ear depths was increased by 0.05 m empirically to account for the size of the ears.

5. Pixel-Based Segmentation Evaluation

The assessment of segmentation methods at the pixel level necessitated an approach that extended beyond the evaluation of superpixel classification or DeepMAC model performance. A supplementary evaluation framework was designed to encompass the complete segmentation process. To this end, a protocol was devised, culminating in the creation of a custom annotation tool. This tool aimed to facilitate rapid and straightforward evaluation, enabling a comprehensive comparison of method performances across different dates and fertilization treatments. The design employed involved annotating 18 pixels, strategically distributed in three rows within each RGB image (Figure 4.2). This configuration was judiciously chosen to strike a balance be-

tween representing image heterogeneity and expediting the evaluation process across a substantial number of images. The tool systematically zoomed in on each pixel, allowing the operator to assign one of three classes: class 1 for background elements (soil, leaves, stems, etc.), class 2 for ears, and class 3 for cases where distinguishing between class 1 and class 2 proved challenging. This scenario often occurred when pixels were positioned at the boundary between an ear and the background.

This procedure was implemented across images for half of the dates spanning from heading to maturity in both 2020 trials. The maturity stage was reached approximately 62 days after heading. For robust statistical comparison between human annotations and predicted segmentations, the F1-score was adopted. Particularly suited for addressing unbalanced classes, which characterize this problem, the F1-score is defined as:

$$F1score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4.1}$$

$$Precision = \frac{TP}{TP + FP}; Recall = \frac{TP}{TP + FN} \tag{4.2}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively, derived from the confusion matrix.

Furthermore, to gauge potential biases arising from the rapid human annotation process, annotations from three different operators were compared using Cohen's kappa coefficient. Widely acknowledged for its robustness compared to accuracy, Cohen's kappa accounts for the likelihood of true values arising by chance. It is defined as:

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{4.3}$$

where p_0 denotes the empirical probability of agreement on the label assigned to any sample (the observed agreement ratio), and p_e represents the anticipated agreement when both operators assign labels randomly, in accordance with scikit-learn documentation.

6. Results and discussion

6.1. Comparative Evaluation of Segmentation Approaches

The comprehensive evaluation of superpixel classification across RF, MLP, and SVM algorithms yielded overall accuracies of 0.93, 0.94, and 0.94 on the test set, respectively. Consequently, the SVM algorithm was selected for superpixel classification. Fine-tuning led to the configuration of the SVM model with a C regularization parameter of 100, a radial basis function kernel, and a kernel gamma coefficient set at 0.1. However, it is prudent to scrutinize the model's robustness considering the validation

Figure 4.2: Schematic representation of the pixel-based evaluation approach. Annotators select 1, 2, or 3 to classify pixels into background, ears, or uncertain categories.

dataset's construction. Notably, the possibility arises that superpixels in the validation set might be proximate to superpixels present in the training dataset.

A comparative analysis based on the pixel-level evaluation unequivocally demonstrates the superior performance of the deep learning approach in contrast to the superpixel method across all dates (Table 4.1). The deep learning technique exhibited an improvement of over 16%, showcasing consistent high performance across diverse dates. Conversely, the machine learning approach exhibited subpar results during the initial dates around flowering. With this compelling outcome in mind, and acknowledging the broad generalization capacity inherent to the deep learning model, we proceeded to adopt the method that integrates both YOLOv5 and DeepMac for the subsequent phases of this thesis.

6.2. Human Annotation Analysis

Human annotation has rarely been evaluated in previous studies, yet it is also a source of error in the calculation of the final metric. The class 3, "uncertain", was added to the pixel-based segmentation evaluation tool to build the cleanest dataset possible in a quick way. In fact, this label was attributed to 3.2% of the total amount of pixels, ranging from 1.3 to 4.5% depending on the considered date. It mainly concerned pixels at the edges between the ears and the background or pixels difficult to identify, for example, in shaded zones. Depending on the nitrogen treatment and the date, 62 to 77% of the pixels labeled "uncertain" have been predicted as background which shows

Table 4.1: F1-score comparison between the two ear segmentation methods for common dates of trials 20-F and 20-FP. Both trials commenced on May 28th.

Date	Trial	Superpixel	Deep learning
06-02	20-FP	0.38	0.76
06-03	20-F	0.47	0.75
06-16	20-FP	0.65	0.88
06-18	20-F	0.81	0.90
07-07	20-FP	0.82	0.91
07-13	20-F	0.68	0.74
07-22	20-FP	0.79	0.84
07-29	20-F	0.68	0.86
Average		0.71	0.83

that this class has fulfilled its mission well and is therefore recommended.

Three annotations from different human operators were compared on the best date, i.e., on June 18th. Cohen's kappa coefficients of 0.79, 0.75, and 0.78 between the different operators were computed, which can be interpreted as "good to strong" agreement and thus validate the evaluation method used. Although algorithms are capable of detecting objects that might be missed by human annotators in images with direct illumination, such as those where the lower part of the canopy is darker, there are often doubts regarding the reliability of the ground truth determined by humans when using AI.

6.3. Ear density estimation

The ear detection using YOLOv5 consistently achieved high accuracy with an F1-score above 0.9 from heading to maturity [Dandrifosse et al., 2022b]. However, the best performance was observed between the end of flowering and early dough growth stage. Figure 4.3 presents a comparison of ear density measurements obtained through the use of the algorithms, three human operators in the field, and a reference measurement from samples counted in the laboratory, for the eight fertilization scenarios of trial 20-F. The reference method recorded the highest ear densities for all scenarios, while operator 3 consistently recorded the lowest values. A two-way variance analysis revealed an interaction between the measurement method and the fertilization scenario factor. As a result, the methods were statistically compared separately for each fertilization level. Nearly all of the corresponding one-way variance analyses showed a significant effect of the measurement method. Therefore, a post-hoc Tukey HSD tests were conducted. In most scenarios, operators 1 and 3 differed significantly from the reference, while operator 2 differed only in one scenario. These significant differences indicate the presence of an operator bias observable in Figure 4.3. This bring back into

question the previous statement about is the ground truth the real truth. Overall, the deep learning method, on the other hand, was never found to be significantly different from operator 2 or the reference. Based on this insights, deep learning was considered a more accurate estimator than operators and used for the rest of the thesis.

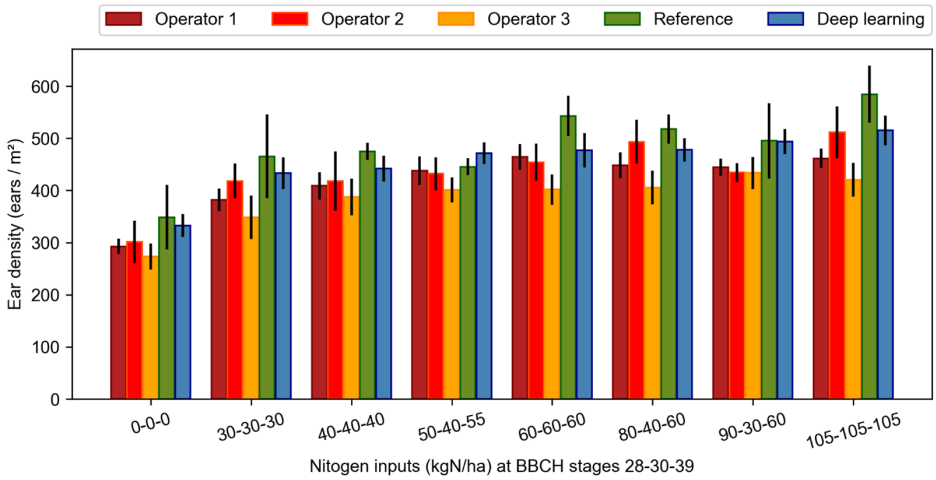

Figure 4.3: Comparison of the trial 20-F ear densities estimated by three operators in the field (July 6th), the deep learning algorithm (images from July 7th) and the reference measurement carried out on wheat samples after the harvest. The standard deviation is indicated on each bar (from [Dandrifosse et al., 2022b]).

7. Conclusion

This chapter delved into two distinctive approaches for ear segmentation. The initial method revolved around superpixel classification, yielding satisfactory outcomes albeit necessitating an additional step for pinpointing individual ears. In contrast, the second approach, a pioneering fusion of YOLOv5 bounding box detection and the DeepMAC segmentation model, offered a comprehensive solution encompassing ear counting and segmentation. Significantly, this segmentation approach demonstrated superior performance over the machine learning alternative, thereby shaping the trajectory of this thesis. Through coupling this technique with the height map derived from stereovision, the computation of ear density became feasible. Remarkably, the automatic method produced more accurate outcomes compared to the labor-intensive field-based counting procedure.

These performances also prompt inquiries into the accuracy of ground truth measurements, particularly when these measurements are resource-intensive and time-consuming. The findings underscore the potential limitations of conventional measurement methods, raising questions about their reliability and precision. As these

techniques are further developed and applied, it becomes imperative to reassess established measurement practices and consider the evolving landscape of technology-enabled alternatives. This reflection is vital for ensuring the accuracy and robustness of agricultural research and practice.

8. Supplementary Materials

Table 4.2: Description of the 15 vegetation indices and the normalized RGB components used as features for the machine learning algorithms aiming to separate leaves and ears. r, g and b are the channels of the RGB camera. blue, green, red, rededge and nir are the channels of the multi-spectral camera array, respectively at 490, 550, 680, 720 and 800 nm.

Index	Index formula	Reference
NDVI	$\frac{nir-red}{nir+red}$	[Rouse et al., 1974]
GNDVI	$\frac{nir-green}{nir+geen}$	[Gitelson et al., 1996]
NDRE	$\frac{nir-rededge}{nir+rededge}$	[Barnes et al., 2000]
GRVI	$\frac{nir}{green}$	[Sripada et al., 2006]
mNDblue	$\frac{blue-red}{blue+nir}$	[Jay et al., 2017]
SR	$\frac{nir}{rededge}$	/
RDVI	$NDVI \times (Nir - red)$	[Roujean and Breon, 1995]
OSAVI	$\frac{(1+0.16)\times(nir-red)}{nir+red+0.16}$	[Rondeaux et al., 1996]
TDVI	$1.5\frac{nir-red}{\sqrt{nir+red+0.5}}$	[Roujean and Breon, 1995]
MSAVI	$\frac{(2nir+1-\sqrt{(2nir+1)^2-8(nir-red)}}{2}$	[Qi et al., 1994]
MCARI	$((rededge - red)-0.2\times(rededge-green))\times \frac{rededge}{red}$	[Daughtry et al., 2000]
TCARI	$3 \times ((rededge - red) - 0.2 \times (rededge - green) \times rededge/red)$	[Haboudane et al., 2002]
VARI	$\frac{green-red}{green+red-blue}$	[Gitelson et al., 2002]
CIrede	$\frac{nir}{rededge} - 1$	[Roujean and Breon, 1995]
CIgreen	$\frac{nir}{green} - 1$	[Roujean and Breon, 1995]
R	$\frac{r}{r+g+b}$	/
G	$\frac{g}{r+g+b}$	/
B	$\frac{b}{r+g+b}$	/

5

To What Extent Does Yellow Rust Infestation Affect Remotely Sensed Nitrogen Status?

1. Synopsis

This chapter relies on the following published paper [Carlier et al., 2023]:

Carlier, A., Dandrifosse, S., Dumont, B. & Mercatoris, B., 2023. 'To what extent do leaf disease pressures affect remotely sensed nitrogen status? The case study of yellow rust', Plant Phenomics, Aug. 2023, doi: 10.34133/plantphenomics.0083.

In field conditions, crops are exposed to several stresses at the same time. Whether biotic, such as pests and diseases, or abiotic such as drought and nutrient deficiency, stresses result in the reduction of the quantity and/or the quality of the harvest. While agricultural inputs have historically been the primary means of mitigating these stresses on a large scale, their extensive use has recently been the subject of many societal and environmental concerns. Moreover, the emergence of concurrent or sequential stresses can exacerbate their negative impact, altering the pattern of symptoms and further hindering crop productivity and stress identification [Pandey et al., 2017; Suzuki et al., 2014]. On this basis, understanding how plants respond to multiple stresses is essential for improving crop yield and quality [Pandey et al., 2017]. In this context, this chapter investigates how wheat yellow rust infestation influences reflectance measurements and nitrogen status assessment by remote sensing. The chapter includes all the elements included in the corresponding research document with very few adaptations for this thesis.

2. Current State of the Art and Objectives

Stress identification and quantification have become common practices using remotely sensed data. Recent plant phenotyping methods offer new possibilities to screen plants in high-throughput, non-destructive and objective way [Araus and Cairns, 2014; Furbank and Tester, 2011]. They have been identified as promising tools to assist plant improvement [Deery and Jones, 2021; Reynolds et al., 2020; Weiss et al., 2020]. Nonetheless, the diversity of data acquisition systems, data management, and analyses bring many challenges to the phenotyping community [Machwitz et al., 2021; Morisse et al., 2022; Sun et al., 2022].

In this context, the investigation of abiotic stresses continues to be a prominent and ongoing subject of study [Al-Tamimi et al., 2022]. One particular area of focus is the detection of nitrogen deficiency, which plays a crucial role in agricultural practices, particularly in relation to fertilization strategies [Berger et al., 2020; Hawkesford and Riche, 2020]. Insufficient nitrogen availability in crops can result in reduced biomass growth and the manifestation of yellowing leaves, both of which are key indicators of nitrogen deficiency. Spectral data analysis, facilitated by techniques such as machine learning, radiative transfer modeling, and vegetation indices (VIs), either independently or in combination, has proven to be effective in tracking these symptoms across

entire crop areas [Berger et al., 2020; Cammarano et al., 2014; Peng et al., 2021; Song et al., 2022]. Moreover, it is important to acknowledge that the presence of background elements, such as soil, can introduce mixing and disturbances to the targeted data associated with the specific plant under study. For instance, Song et al. [2022] reported that the accuracy of nitrogen status estimation using a spectroradiometer was relatively lower during the early growth stage of crops compared to the subsequent vegetative phase. Consequently, certain researchers have effectively addressed this issue of spectral mixing at the canopy level. Wang et al. [2021] proposed a novel approach known as abundance-adjusted VIs, which mitigates spectral mixing at the canopy level and enhances the accuracy of leaf nitrogen concentration estimation. Noteworthy achievements have been attained by researchers who have focused their analysis on specific regions within the canopy. For instance, Jay et al. [2017] successfully estimated beet chlorophyll content by concentrating on the most illuminated pixels of green vegetation.

Under field conditions, the detection of diseases still comes up against many difficulties such as the similarity of symptoms, the possibility of observing them, and the diversity of plant responses [Berger et al., 2022]. Common methods include the use of spectral data [Wan et al., 2022] or RGB images [Mahlein, 2016]. For instance, Anderegg et al. [2019] has quantified Septoria Tritici Blotch using spectral and temporal features from a spectroradiometer. Other image analysis methods, such as textural analysis from proximal multispectral images, were relevant to estimate the severity of wheat main diseases [Bebronne et al., 2020]. Recently, deep learning algorithms are paving a new avenue for plant phenotyping [Jiang and Changying Li, 2020; Kattenborn et al., 2021; Singh et al., 2018]. In particular, convolutional neural networks (CNNs) have indeed demonstrated their good performances in phenotyping task related to object detection [David et al., 2021], segmentation [Fan et al., 2022], or disease classification [Ferentinos, 2018; Görlich et al., 2021].

Despite all these stress detection possibilities, very few studies have addressed the effects of multi-stress and their interactions on single trait estimation, as depicted by [Berger et al., 2022; Zhang et al., 2019]. Authors emphasised the need for further research to address this challenge. In fact, all mentioned studies related to nitrogen stress estimation have treated the question in optimal management practices, while it is known that diseases could substantially disturb the nitrogen dynamic [Simón et al., 2020]. The nutritional habit of the disease could induce a reduction in leaf nitrogen concentration [Bancal et al., 2008], and thus induce a bias in the nitrogen status estimation. Hyperspectral systems appear as the most suitable sensor to face this challenge. For instance, Devadas et al. [2015] was able to delimit stripe rust and nitrogen deficiency using specific vegetation indices with a spectrometer. Another solution might be the use of multi-sensor approach. For instance, while symptoms associated with a pathogen were similar to those of water stress, Zarco-Tejada et al. [2021] were able to distinguish between both stresses using the combination of hyperspectral and thermal

sensors. Generally speaking, many biotic stresses manifest visible symptoms that can be segmented provided sufficient spatial resolution. In this context, close-range systems such as mobile platforms or gantries are good candidates, as they can carry multiple sensors in close range and thus provide high-resolution data [Xu and Li, 2022].

This chapter presents an approach to investigate the impact of yellow rust on vegetation indices that are usually used in the frame of nitrogen status retrieval. The hypothesis is that the diseases induce a bias in the estimation of nitrogen status not only through its visible symptoms but also through its biological interaction with the plant. To test this hypothesis, proximal RGB and multispectral images were acquired on a wheat field trial spanning two cropping seasons, where different fungicide applications and nitrogen inputs were combined. The methodology involved isolating the leaves within the images, segmenting disease symptoms (i.e. leaf damages), and using the resulting mask to study the correlation between vegetation indices from healthy or diseased leaves, and nitrogen status variables of the plant.

3. Materials and Methods

3.1. Color Image Segmentation

3.1.1. SegVeg method

A robust RGB image segmentation technique, known as SegVeg, was utilized in this study, as described in the work by [Serouart et al., 2022]. To ensure accurate and reliable results, the method employed a combination of two distinct techniques: (i) a deep learning approach for soil-vegetation segmentation and (ii) a pixel-wise approach for green-yellow vegetation segmentation. This approach allowed for the generation of a robust model that effectively removes soil, enables specific classification of plant pixels and reduces the effort of annotating them. Furthermore, employing a binary classification approach twice would not result in detrimental instances of misclassification, unlike a three-class approach where a disease could potentially be misidentified as soil.

The deep learning approach was based on the popular U-NET model, which has been successfully applied in various image segmentation tasks [Ronneberger et al., 2015]. To leverage the benefits of pre-trained models and accelerate training convergence, the U-NET structure utilized different pre-defined encoders, or backbones, that were pre-trained on ImageNet [Deng et al., 2009]. The encoder, which is the down-sampling part of the U-NET, was implemented using two state-of-the-art architectures - ResNet34 proposed by [He et al., 2015], and EfficientNetB2 proposed by [Tan and Le, 2020]. The decoder, or the up-sampling part of the U-NET was implemented as a classical design.

The pixel-wise segmentation was carried out using features from the RGB images.

Different color spaces and transformations were computed, namely the normalized RGB channels, the HSV, the CIELab, and the Sobel filter. That made a total of ten features per pixel. Two models were tested; the support vector machine (SVM) widely used in phenotyping [Singh et al., 2016], and the eXtreme Gradient Boosting known as XGboost, a bagging approach known for its performances and rapidity.

The implementation of the algorithms was done using the 'Segmentation Models' package by [Iakubovskii, 2019], Tensorflow 2.4, XGBoost 1.7 and Scikit-learn 1.2.

3.1.2. Dataset Preparation and Training

The VegAnn dataset [Madec et al., 2023] has been enhanced with additional images from the current study. This dataset comprises a collection of RGB images along with corresponding binary masks for plant-soil segmentation. The VegAnn dataset encompasses 3775 multi-crop RGB images captured under diverse illumination conditions, using various systems and platforms, and representing different phenological stages.

To ensure the adequacy of the model to the present study, 30 RGB images from the 22-FP trial dataset and 8 from the 21-FP trial dataset have been selected. This two trial were refereed as 2022 and 2021 respectively in this chapter. The other trials were not used. The RGB images have been manually segmented into two classes: soil and plant parts with and without damage. The masks were generated using the plug-in Labkit [Arzt et al., 2022] from the Fiji software [Schindelin et al., 2012]. It is a user-friendly platform for manual and automated image segmentation. The segmentation process involved manually drawing the soil and plants on a few areas in the image, after which a fast Random Forest based pixel classifier was used to segment the entire image. The number of manually drawn areas varied depending on the annotator's judgement and the heterogeneity of the image and comprised several dozens of pixels. This approach allowed for improving the integrated Random Forest by adding more labels to regions that were poorly predicted by the classifier. Although less accurate than manual labelling, this tool provided a faster way to generate masks.

Then, the images and corresponding masks were partitioned into twenty non-overlapping images of 512×512 pixels, similar to the VegAnn dataset. One-third of this dataset was used as the validation dataset, comprising a total of 260 patches.

For training, a batch size of 16 was set for 100 epochs. The Adam optimizer with default parameters of Tensorflow was used, and each image was scaled according to the corresponding backbone, similar to the ImageNet pre-processing. The dice loss was chosen as the loss function as it presents better capacity to handle unbalanced dataset.

Secondly, to generate a comprehensive pixel-wise segmentation dataset, 120 RGB images from the 2022 dataset were selected for training, and 33 RGB images from the 2021 dataset were selected for validation. The manual annotations included a few pixels of both green and damaged parts to ensure a balanced distribution of each class.

This annotation were also done using the Labkit tool without applying the Random Forest classifier. The selection of images was based on the need to represent the heterogeneity of images encountered. Thus, the training dataset comprised approximately 15,000 pixels for each class, while the validation dataset contained around 2,500 pixels for each class.

3.1.3. Evaluation Metrics

Models were evaluated using the accuracy as a standard metric (Equation (5.2)). A common other metric for semantic segmentation is the Intersection over Union (IoU) (Equation (5.1)) with a threshold of 0.5. It is the ratio between the area formed by the overlap of the predicted and the labeled regions and the area formed by the set of these two regions. It ranges from 0 to 1. The lower the IoU, the worse the prediction result. The computing configuration was a NVidia Tesla V100 GPUs.

$$IoU = \frac{Intersection}{Union} = \frac{TP}{FP + TP + FN} \tag{5.1}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5.2}$$

TP, TN, FP, and FN, respectively, stand for the number of pixels of true positives, true negatives, false positives, and false negatives.

3.2. Image Analysis Pipeline

3.2.1. Processing of Images

The developed image analysis pipeline is presented in Figure 5.1. First, the RGB image was used to segment the soil, the green plant parts which were mainly leaves, and the damaged parts (Section 3.1). Then, wheat ears were detected using YoloV5 and segmented within each bounding box by the Deep Mac model (Chapter 4). Those masks were combined to build a mask with four classes: the soil, the ears, the damaged parts of the leaves and the green parts of the leaves. Note that, the leaves refer to the leaves complemented by the stem parts visible in a nadir view. Secondly, these masks were applied on the multispectral images.

The other processes details in Section 3.4 were also performed. That encompassed the image registration process and the BRF calculation. Thus, five vegetation indices (VIs) were selected for their demonstrated relationships with crop nitrogen status and plant health, disease, or senescence (Table 5.4). These VIs included the Normalized Difference Red Edge (NDRE) index [Barnes et al., 2000], the modified normalized difference blue Index (mDNb) [Jay et al., 2017], and the Chlorophyll Index Red Edge (CIred-edge) [Gitelson et al., 2006], which are known to be correlated with crop nitrogen status, and the Normalized Difference Vegetation Index (NDVI) [Rouse et al., 1974] and the Plant Senescence Reflectance Index (PSRI) [Anderegg et al., 2020; Mer-

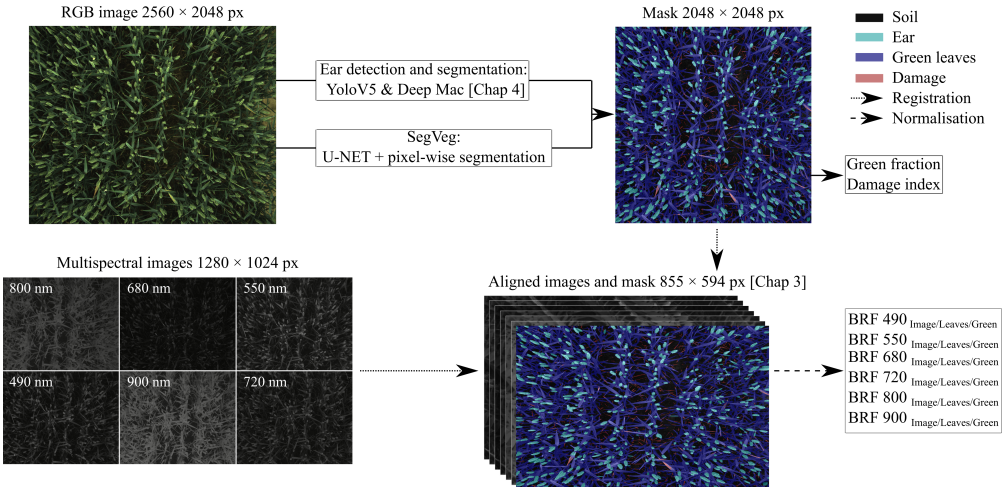

Figure 5.1: Image analysis pipeline. RGB images were used to segment the scene into soil, ear, green leaf and damage. Combined with multispectral images, the pipeline allows to extract the BRFs of each mentioned classes.

zlyak et al., 1999] for their sensitivity to plant health, disease, or senescence.

3.2.2. Foliar Damage Quantification

On the basis of the RGB segmentation, the proportion of pixels representing green and damaged parts of the plants were computed. The green fraction (GF) was defined as the number of green plant pixels divided by all pixels in the image, meanwhile, the damage index (DI) was calculated as the proportion of damage pixels relative to the sum of the green plant and damage pixels. The study made the assumption that the observed damage was primarily attributed to disease and that no other sources of damage, such as physiological or insect-related damages, were observed throughout the experiment.

To correctly assess the disease importance throughout the growing season and take into account the temporal dynamics of the severity of the disease, the area under the disease progression curve (AUDPC) was computed following the procedure proposed by [Simón et al., 2020]. Thus, $AUDPC_{sVS}$ was calculated from the sVS scores observed within plots whereas $AUDPC_{DI}$ was computed from the DI extracted from the mask.

3.3. Statistical Analysis

The study employed a comprehensive data analysis approach that integrated various statistical methods. A analysis of variance (ANOVA) was conducted to assess the impact of treatments on both image features and agronomical data. However, it was important to exclude the 180_1F treatment from the agronomic data analysis. This ex-

Table 5.1: IoU and accuracy of the SegVeg model. Soil-Plant results refer to the U-NET model, and the Green-Damage results refer to the pixel-wise classifier.

		Soil-Plant		Green-Damage	
		EfficienNetB2	ResNet34	Xgboost	SVM
Training	IoU	0.89	0.84	0.85	0.86
	Accuracy	0.94	0.91	0.92	0.92
Validation	IoU	0.76	0.67	0.78	0.79
	Accuracy	0.87	0.83	0.88	0.88

clusion was necessary because this specific nitrogen treatment had only one fungicide factor, which could introduce bias into the statistical analysis. Therefore, to ensure accurate interpretation of the results, it is crucial to consider this exclusion. Furthermore, a post-hoc Tukey HSD test was performed to identify any significant differences in the data.

In addition, Pearson correlation coefficient (r) was utilised to explore the relationship between image features and agronomical data. The study also employed a multiple linear regression to determine the added value of image features in modeling agronomical data, with the calculation of coefficient of determination R^2. The hypothesis was that GF and DI were good indicators of plant health and could improve the estimation model performances.

Finally, paired t-tests were conducted to compare BRFs and VIs obtained from the leaves with those obtained from the entire image and green elements.

4. Results

4.1. Disease Pressure

The EfficientNetB2 backbone demonstrated the best performance, achieving an IoU of 0.76 and an accuracy of 0.87 for soil segmentation on the validation set (Table 5.1). In terms of distinguishing between green elements and damaged ones, both pixel-wise classifiers showed high accuracy and IoU. Notably, the Xgboost model was selected over the SVM due to its speed, being approximately 10 times faster. Consequently, the SegVeg model was formulated by combining the EfficientNetB2 and Xgboost models to effectively segment the entire dataset. This unified model facilitated highly accurate segmentation of soil, green elements, and damaged regions, even in difficult strong direct sunlight conditions, as visually demonstrated in Figure 5.2.

When the SegVeg approach was applied throughout the entire cropping season, it became apparent that the 2021 season was characterized by a relatively low disease pressure. However, during the grain filling period, there was a slightly increased in-

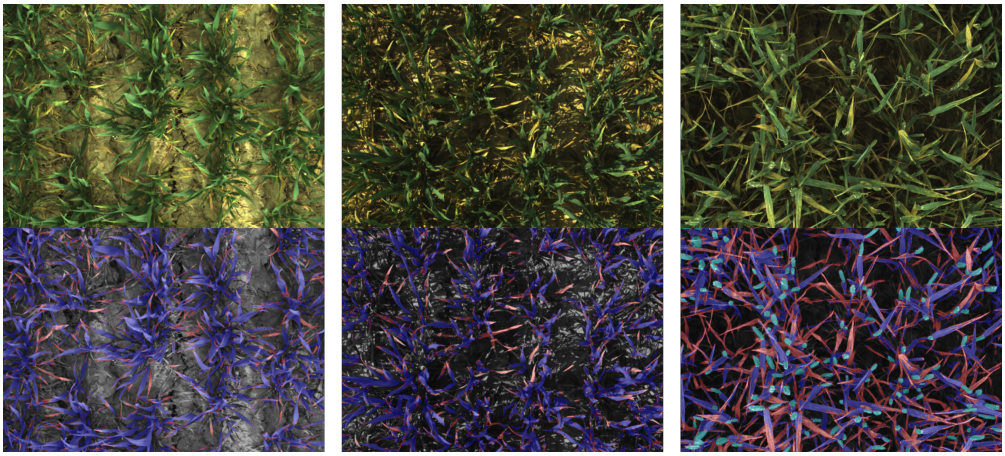

Figure 5.2: Examples of segmentation utilizing the SegVeg approach, which combines the EfficientNetB2 and XGboost models, along with ear segmentation employing Yolov5 and DeepMAc on April 25, May 5, and May 30, respectively from left to right. In the segmentation results, the soil regions are depicted in shades of grey, green plants in blue, ears in sky blue, and damages in red.

cidence of diseases, although specific data is not presented. Among the diseases affecting wheat, Septoria Tritici Blotch (STB) was the primary concern, with only a few treatments reaching an average severity value score (sVS) of 0.625 in early July. Yellow Rust (YR) was detected on less than 20% of the plots and had a maximum average sVS of 0.625.

In contrast, the climatic conditions experienced in 2022 resulted in the early onset of YR at the end of April, specifically at growth stage (GS) 30. Subsequently, the disease exhibited significant development across all experimental plots during the stem elongation period, as depicted in Figure 5.3. Notably, the 0F treatment displayed the highest sVS and damage index (DI) throughout the season, with the DI reaching a maximum of approximately 60%. Conversely, the 1F treatment effectively managed the disease pressure through the application of fungicide on May 17, resulting in a stabilization of sVS and a decrease in DI. The 2F and 3F treatments exhibited similar dynamics of disease pressure until the end of May, as illustrated in Figure 5.3 and table 5.5. However, after the flowering stage, the effectiveness and timing of application of the 3F treatment became evidently discernible.

The calculation of the Area Under the Disease Progress Curve (AUDPC) serves as a reliable indicator of the overall disease pressure throughout the cropping season, offering an advantage over single-point notations such as sVS or DI. The correlation between $AUDPC_{sVS}$ and $AUDPC_{DI}$ and the final grain yield indicates that as the season progressed, the correlations became stronger (in absolute value), as depicted in Figure 5.4. Notably, both AUDPC measurements exhibit a high correlation after the

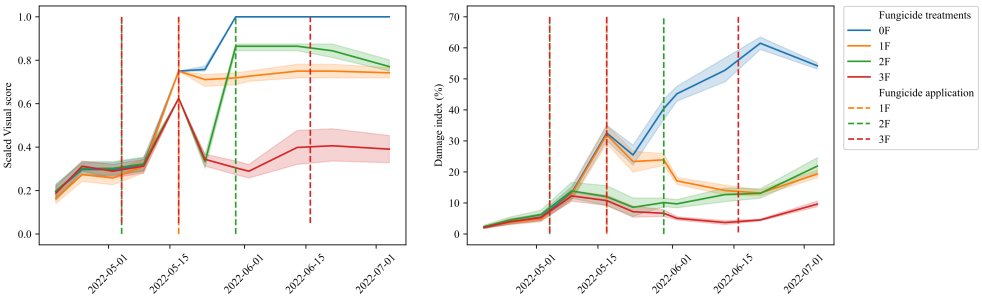

Figure 5.3: Scaled Visual Score and damage index curves during the 2022 season. Shaded bands represent standard deviation.

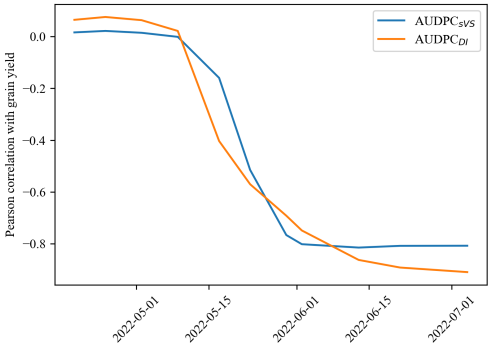

Figure 5.4: Pearson correlation between the 2022 grain yield, and both $AUDPC_{sVS}$ and $AUDPC_{DI}$.

flowering stage. However, the proposed method, $AUDPC_{DI}$, demonstrates a higher correlation compared to $AUDPC_{sVS}$, which tends to reach a plateau earlier in the season, around May 30. Furthermore, prior to any fungicide applications, the $AUDPC_{DI}$ values for the lower fertilization treatment were statistically different from those of the higher nitrogen treatments, as indicated in Table 5.5. In a broader sense, it can be observed that the Tukey HSD grouping was initially determined based on the nitrogen treatment at the beginning of the season and subsequently based on the fungicide treatment, which transitioned from May 9th to May 17th. This observation reveals that the differences initially arose from variations in nitrogen input and subsequently from variations in disease pressure.

4.2. Disease Effects on Bi-Directional Reflectance Factor

In Figure 5.5, the blue boxplot represents the entire image signal, which includes soil reflectance. Comparing the blue boxplot to the boxplots of leaf and green element reflectance, significant differences were observed across most dates for all BRFs, ex-

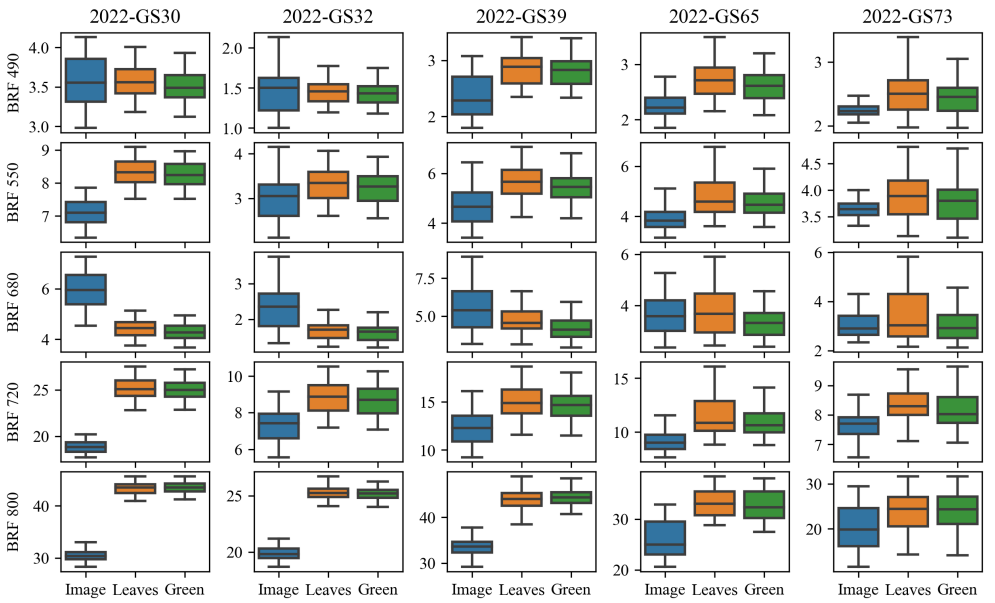

Figure 5.5: Boxplot of the Bi-Directional Reflectance Factor according to its source, i.e., from the entire image, only the leaves and only the green elements at different growth stages.

cept for the dates specified in Table 5.7. Furthermore, the differences in BRFs were more pronounced early in the season when the canopy cover was low. For BRF 800, the maximum variation rate, calculated as the percentage difference between the mean BRF 800 and the mean BRF 800 of leaves, reached up to 41.8% (Table 5.2).

The first hypothesis of this study posited that diseases or any damaging stress could impact the BRFs of the crop, primarily due to the presence of lesions and the potential signal disturbance caused by soil. The results revealed that BRFs obtained from the leaves and those derived from the green elements exhibited similar values; however, a paired t-test revealed statistically significant differences between them, with a few exceptions (refer to Table 5.6). Notably, BRF 680 demonstrated higher variation compared to BRF 800, which was minimally affected (Table 5.2). For instance, in 2022-GS65, disease led to a reduction of 12.5% in between BRF 680 of the leaves and of the green elements.

Moreover, a strong correlation was observed between the discrepancy in BRFs derived from the leaves and those from the green elements, and the damage index, representing the extent of disease (see Table 5.8). The majority of these correlation exceeded 0.5 in absolute value, with values as high as 0.9 observed during periods of heightened disease pressure. These findings were consistent when analyzing vegetation indices as well. It can be concluded that the discrepancy of BRFs was clearly influenced by the amount of damages in this study.

Table 5.2: Rate of change between BRFs of image and leaves (I-L), and between leaves and green elements (L-G) in %.

	BRF 490		BRF 550		BRF 680		BRF 720		BRF 800	
	I-L	L-G	I-L	L-G	I-L	L-G	I-L	L-G	I-L	L-G
2022-GS30	0.40	-1.70	16.28	-0.70	-23.30	-3.22	32.42	-0.49	41.81	0.35
2022-GS32	-1.08	-1.47	9.55	-2.33	-28.00	-3.60	19.87	-1.81	26.60	-0.29
2022-GS39	20.83	-1.36	20.15	-3.44	-15.37	-10.56	22.23	-2.57	31.04	1.08
2022-GS65	20.15	-3.77	23.17	-5.77	4.47	-12.53	25.08	-4.89	26.94	-0.68
2022-GS73	11.72	-2.74	6.74	-3.06	10.15	-9.63	8.72	-2.67	16.59	0.51

The second hypothesis of this study proposed that diseases not only affect visibly symptomatic plant parts but also symptomless ones, with varying impacts depending on the nutritional strategy of the disease [Simón et al., 2020]. Consequently, this phenomenon could potentially disrupt the measurement of green element BRFs. To exemplify this effect, the NDRE was selected as a well-established vegetation index associated with nitrogen status. An analysis of variance (ANOVA) revealed that the $NDRE_{green}$, computed on healthy areas, was influenced not only by fertilization but also by the fungicide treatment (refer to Table 5.3).

Furthermore, during GS39 in 2022, the $NDRE_{green}$ value for the 120_3F treatment was found to belong to the same group as the other 3F treatments, but not with the 120_0F treatment (Figure 5.6). This discrepancy indicates that the presence of disease impacted the $NDRE_{green}$ value for the 120_3F treatment differently compared to the 120_0F treatment.

Additionally, it was observed that the difference ($\Delta_{3F\text{-}xF}$), which represents the variation between the values obtained from full protection and those obtained with reduced protection under constant nitrogen input, exhibited a strong correlation with the DI for both $BRFs_{green}$ and VIs_{green}. Specifically, the correlation between $\Delta_{3F\text{-}xF}$ of the $NDRE_{green}$ and the DI exceeded 0.70 starting from May 17, as indicated in Table 5.9..

4.3. Analysis and Modeling of Nitrogen Status Variables Under Fertilization and Fungicide Treatments

The analysis of variance revealed only one significant interaction term for all nitrogen status variables (see Table 5.3) - %N leaves 2021-GS65. However, values close to 0.05 were also observed for 2022-GS65 for NNI and %N leaves. Therefore, while the two factors can be analyzed separately, caution must be exercised when drawing conclusions.

The fertilization factor significantly impacted most variables, with %N leaves and %N total showing significance at tillering (2022-GS30). However, no effect was observed on Nuptake leaves and Nuptake total for 2021-GS39, as well as on NNI, Nup-

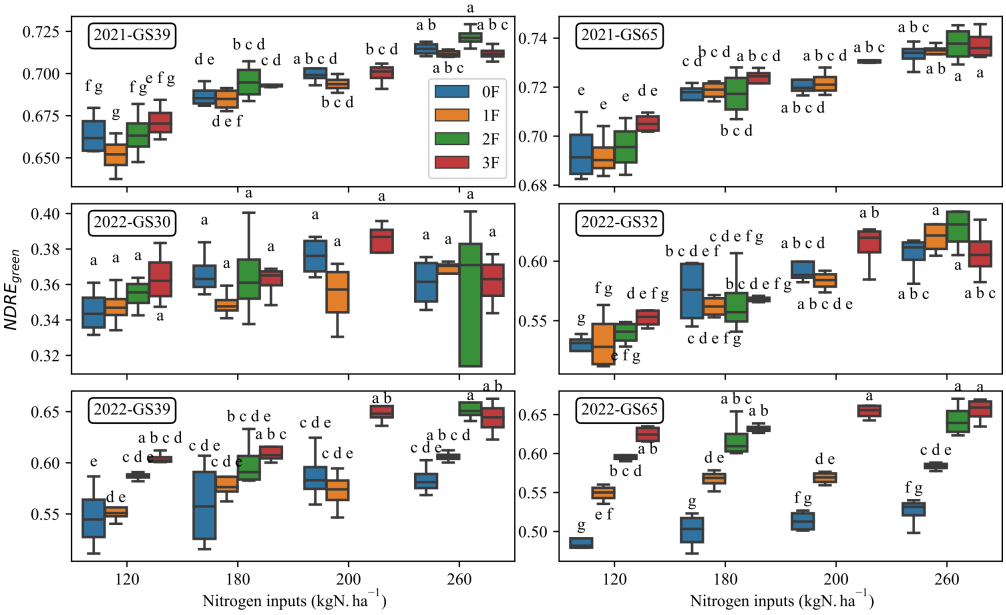

Figure 5.6: $NDRE_{green}$ according to the treatment for major growth stages. Letters represent the groups created by the post-hoc Tukey HSD test.

take total, and %N total for 2022-GS75.

Fungicide did not affect NNI and %N total, except for the maturity growth stage. However, Nuptake leaves and Nuptake total were significantly influenced by fungicide starting from 2022-GS65. %N leaves were affected earlier, at 2022-GS39.

$NDRE_{green}$ displayed a similar trend to %N leaves in 2022 regarding the impact of fungicide application. However, it was not affected by fertilization at GS75. In 2021, it was influenced by both fertilization and fungicide. Additionally, $NDRE_{green}$ exhibited a stronger correlation with leaf nitrogen status than $NDRE_{leaves}$ when disease was present, specifically from GS 39 to 75 in 2022 (Figure 5.7). Meanwhile, the correlation with NNI was much lower in 2022 compared to 2021 for GS 39 and 65. It is noteworthy that there was a decrease in the correlation between %N of leaves and plant nitrogen status variables, such as NNI and %N total, during periods of high disease pressure. Moreover, an interesting result is that the correlation between $NDRE_{leaves}$ and $NDRE_{green}$ was perfect while the values may differ (see above section).

Finally, some multiple regression analyses were conducted using features from both RGB and multispectral imagery. The inclusion of the green fraction and the damage index improved most of the model performances for dates with high disease pressure (Table 5.10). Moreover, it also improve general model which encompass all data, i.e., dates and plots.

Table 5.3: P-values (ANOVA) of NNI, %N leaves, Nuptake leaves, %N total, and Nuptake total at different growth stages, where 'Total' refers to the entire plant.

Source of variation	2021-GS39	2021-GS65	2021-GS89	2022-GS30	2022-GS32	2022-GS39	2022-GS65	2022-GS75	2022-GS89
NNI									
Fertlisation (N)	<0.01	<0.01	<0.01	0.08	<0.01	<0.01	<0.01	0.197	<0.05
Fungicide (F)	0.199	0.531	0.429	0.404	0.965	0.414	0.676	0.055	<0.05
$N \times F$	0.814	0.198	0.553	0.328	0.579	0.914	0.058	0.185	0.473
%N leaves									
Fertlisation (N)	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.05	<0.01
Fungicide (F)	0.55	0.456	<0.01	0.96	0.815	<0.05	<0.01	<0.01	<0.01
$N \times F$	0.845	<0.05	0.642	0.355	0.263	0.679	0.056	0.169	0.535
Nuptake leaves									
Fertlisation (N)	0.27	<0.01	<0.01	0.562	<0.01	<0.01	<0.01	<0.01	<0.01
Fungicide (F)	0.392	0.631	0.074	0.156	0.904	0.225	<0.05	<0.01	<0.01
$N \times F$	0.445	0.237	0.706	0.451	0.672	0.724	0.285	0.538	0.385
%N total									
Fertlisation (N)	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.269	<0.01
Fungicide (F)	0.322	0.463	0.313	0.96	0.9	0.461	0.476	0.487	0.342
$N \times F$	0.379	0.098	0.242	0.355	0.383	0.985	0.254	0.179	0.736
Nuptake total									
Fertlisation (N)	0.232	<0.01	<0.05	0.562	<0.01	<0.01	<0.01	0.151	<0.05
Fungicide (F)	0.385	0.62	0.36	0.156	0.857	0.467	<0.05	<0.01	<0.01
$N \times F$	0.384	0.349	0.769	0.451	0.722	0.867	0.09	0.238	0.346
$NDRE_{green}$									
Fertlisation (N)	<0.01	<0.01		0.409	<0.01	<0.01	<0.01	0.682	
Fungicide (F)	<0.01	<0.05		0.512	0.568	<0.01	<0.01	<0.01	
$N \times F$	0.417	0.891		0.72	0.258	0.248	0.905	0.352	

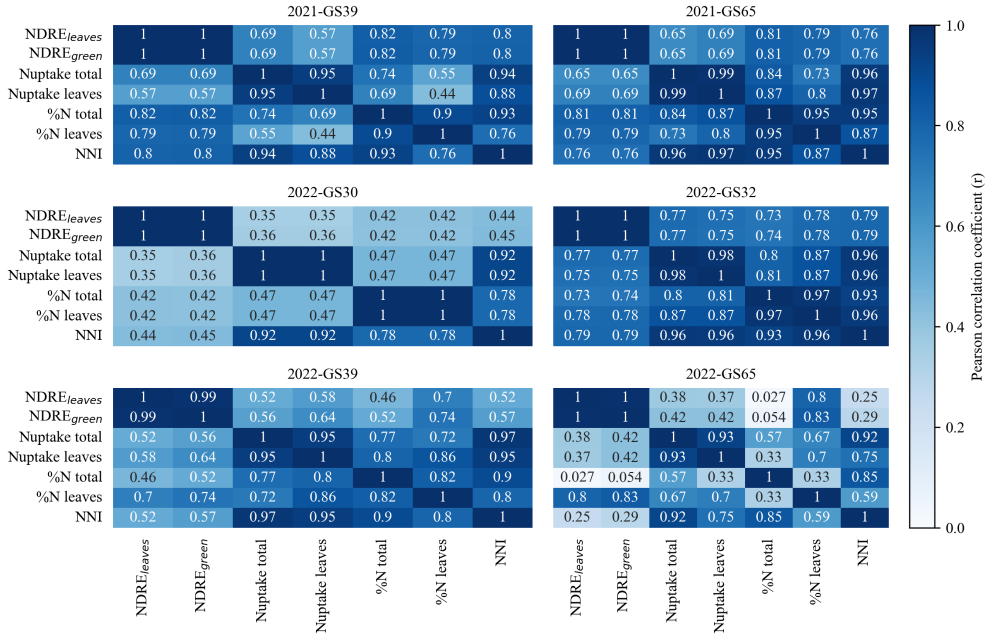

Figure 5.7: Correlation matrix between NDRE and nitrogen status variables at different growth stages.

5. Discussion and Conclusion

5.1. CNN as a Promising Damage Detection Tool

Plant diseases are commonly identified through the observation of visible symptoms, a process conducted by agronomists to assess the plant's resistance capabilities. However, this approach can be time-consuming, labor-intensive, and susceptible to subjectivity. To address these limitations, a scoring method named SegVeg was proposed by [Serouart et al., 2022] for evaluating non-green elements, predominantly characterized as disease symptoms in the present study. The SegVeg model employed a two-step methodology that involved generating a mask with three distinct categories: soil, green plant parts, and damaged parts. This approach utilized a U-NET architecture in conjunction with a pixel-wise classifier. While the model yielded satisfactory segmentation masks, it was found that misclassified pixels could arise under direct sunlight conditions [Serouart et al., 2022]. To mitigate this issue, additional data encompassing various illumination conditions were incorporated into the VegAnn dataset as part of the current study. Nevertheless, due to the inherent scattered nature of the wheat canopy, particularly in intense sunlight conditions, the creation of accurate masks became more challenging, as exemplified in the left image of Figure 5.2. Specifically, the lower regions of the canopy tended to exhibit significant darkness. The utilization of a high dynamic range (HDR) camera presents a potential solution to alleviate the aforementioned issues.

Both visual scoring and the SegVeg method have demonstrated their efficacy in characterizing the impact of fungicide treatment on wheat plants. In contrast to the study conducted by [Koc et al., 2022], these two methods are distinct and yield different outputs. However, they can be utilized for similar purposes, such as estimating the effect of disease on yield loss. The SegVeg method lead to the damage index that objectively quantifies the extent of damage in a nadir view, providing a 2D assessment. On the other hand, visual scoring considers the disease intensity on the most significant leaves. In our study, the damage index indicated very high levels of infestation, exceeding 50% in the zero protection treatment, which resulted in a completely devastated plot even under natural inoculation. Conversely, the well-protected treatment, with minimal disease observed by human assessors, yielded a damage index value of 10%, suggesting a slight overestimation by the models. The current nadir view system restricts observations to visible symptoms on the upper leaves, limiting its capacity to assess diseases such as STB that primarily develop in the lower canopy. To address this limitation, a potential solution could involve implementing a system that opens the canopy, similar to the manual manipulation performed by human assessors during visual scoring. This approach could be combined with an object detection CNN to detect and quantify disease spots, as demonstrated by [Schirrmann et al., 2021]. Furthermore, extending the pixel-wise annotation to include damage classification could offer enhanced insights into the nature of the damage. However, this would require

substantial efforts in image acquisition and annotation tasks.

The nadir view perspective can also explain the observed decrease in the damage index when new green leaves emerge. This phenomenon led to the utilization of the AUDPC as a metric to account for the negative impact of disease throughout the season, which proves to be a suitable measure for studying its influence on grain yield. As the season progresses, the treatments became more distinguishable from one another, and their correlation with the final grain yield has strengthened. Similar findings were reported by [Zhou et al., 2015] using the green fraction. Towards the end of the season, computer vision techniques surpassed the visual scoring method, likely because sVS reaches its maximum value early on and can no longer differentiate between the different treatments. Notably, foliar diseases have a detrimental effect on carbon accumulation by reducing the green leaf area until senescence occurs [Bancal et al., 2008; Simón et al., 2020]. However, the fungicide mixture, consisting of Triazoles-Pyrazoles-carboxamides, may also significantly impact the green leaf area, while the last triazole appeared to have no effect. This fungicide interaction adds complexity to the already complex relationship between nitrogen plant fluxes and rust severity [Schierenbeck et al., 2019b]. In fact, biotrophic pathogens like Yellow Rust usually benefit from high nitrogen availability [Simón et al., 2020]. However, in this study, no statistically significant differences were observed in the damage index to confirm this statement.

Lastly, it is worth noting that the grouping of $AUDPC_{DI}$ in Table 5.5 initially focused on nitrogen input and subsequently on fungicide treatment. This suggests that distinguishing between different nitrogen treatments may not be possible beyond a certain level of disease pressure, as the plant is unable to fully recover from the damage.

5.2. Disease Affects Reflectance in Two Ways

The findings of this study highlight the importance of accurately differentiating the elements present in multispectral imagery of crops. Specifically, the removal of soil and other background elements from crop scenes is crucial for the proper evaluation of crop phenotypes, as shown in Table 5.2.

Similar conclusions can be drawn regarding the impact of diseases. Diseases, through their symptoms, lead to a reduction in healthy areas [Bancal et al., 2008]. The results of the study indicate that damage symptoms have a significant effect on BRFs, with the 680nm wavelength exhibiting a particularly pronounced impact. This suggests that spectral data at this specific wavelength can serve as a valuable tool for distinguishing between healthy and diseased plants [Anderegg et al., 2019; Ashourloo et al., 2014; Whetton et al., 2018]. Furthermore, a strong correlation was observed between variations in BRFs and the extent of damage caused by the disease, indicating that as the disease progresses, the differences in BRFs become more pronounced.

The study also revealed that diseases affect the BRFs of the green area of the plant.

It was hypothesized that diseases can significantly impact the biophysical and biochemical properties of wheat plants, thereby influencing the measurement of BRFs and subsequent vegetation indices [Simón et al., 2020]. In the presence of disease, the measurement of $NDRE_{green}$ showed a stronger correlation with nitrogen status variables compared to $NDRE_{leaves}$. Specifically, it exhibited a high correlation with leaf nitrogen concentration and leaf nitrogen uptake, but not with other nitrogen status parameters. However, the nitrogen status of the leaves appeared also to deviate from other nitrogen parameters such as the NNI and the overall plant nitrogen concentration. This deviation is likely due to the complex influence of diseases on the overall nitrogen status of the plants [Simón et al., 2020]. In fact, yellow rust was found to impact the photosynthetic capacity of the green elements, but not the nitrogen content of the stem [Bancal et al., 2008]. Therefore, while spectral measurements, which primarily capture information from leaves, may effectively represent the nitrogen status of the leaves, they may not necessarily reflect the nitrogen status of the entire plant. This has important implications for fertilization decision-making tools. It was observed that a single value of $NDRE_{green}$ could represent different nitrogen input levels (Figure 5.6), which could potentially result in misleading interpretations. This is particularly relevant when considering the last fertilisation input made at GS 39 in Belgium.

To address this issue, it is important to carefully consider disease quantification in nitrogen estimation models based on spectral data. The use of $NDRE_{green}$, or the addition of features from the RGB image, such as the damage index, could aid in modeling nitrogen status variables. In addition, on each individual date, we observed a strong and significant correlation between both NDRE values. It is important to clarify that this correlation does not imply that the values are identical, but rather indicates a high degree of association between them. Consequently, using these NDRE measurements in relative terms may not pose any issues. However, in the context of a larger case study or when considering absolute values, it is possible that discrepancies or challenges may arise. In fact, it can be challenging to develop a model that accurately represents the nitrogen concentration for all growth stages, while the total nitrogen uptake is more easily assessed across all dates [Prey and Schmidhalter, 2019a]. Hyperspectral systems have an advantage in distinguishing between nitrogen deficiency and rust infection, as they use narrower wavebands [Devadas et al., 2015]. Additionally, from a nadir view, they are more effective at detecting diseases that develop in the canopy by sensing overall plant health [Yu et al., 2018]. However, their use in the field may be limited due to practical constraints, higher costs, and equipment complexity

Lastly, during the research, an unsupervised clustering model was tested using all image features. Initially, the model clustered the plots based on nitrogen levels and subsequently, from the middle of May, it differentiated them according to fungicide treatment, even when utilizing RGB features (data not shown). This further supports our earlier assertion that above a certain threshold of disease pressure, accurately determining the nitrogen treatment of a plot without historical information becomes chal-

lenging. It became evident that studying temporal features emerged as a reliable approach for disentangling stress factors [Anderegg et al., 2019; Berger et al., 2022]. Therefore, based on these findings, it is recommended to conduct further research on nitrogen stress modelling using spectral data in the presence of disease. It is worth noting that different diseases may exhibit distinct interactions with plant nitrogen status, as explained by [Simón et al., 2020], implying that the observations made in this study for Yellow rust may not necessarily apply to other diseases such as Septoria.

6. Supplementary Materials

Table 5.4: Vegetation indices selected in this study

Name	VI forumlation	Reference
NDRE	$\frac{\lambda 800-\lambda 720}{\lambda 800+\lambda 720}$	[Barnes et al., 2000]
mNDb	$\frac{\lambda 490-\lambda 720}{\lambda 490+\lambda 800}$	[Jay et al., 2017]
Cirede	$\frac{\lambda 800}{\lambda 720}-1$	[Gitelson et al., 2006]
NDVI	$\frac{\lambda 800-\lambda 680}{\lambda 800+\lambda 680}$	[Rouse et al., 1974]
PSRI	$\frac{\lambda 680-\lambda 490}{\lambda 800}$	[Anderegg et al., 2020; Merzlyak et al., 1999]

Table 5.5: Tukey HSD test to investigate the effect of the treatment on the $AUDPC_{DI}$ across the cropping season.

Treatment	19/04/2022	25/04/2022	02/05/2022	09/05/2022	17/05/2022	23/05/2022	30/05/2022	02/06/2022	13/06/2022	21/06/2022	04/07/2022
120_0F	a b	a b c	a b c d	a	a	a b	a b	a b	a b c	a	a
120_1F	a b	a b	a b c	a	a	a b	a b c d	a b c d	c d e	b c	b
120_2F	a	a	a	a	a	a b	a b c d	b c d e	d e f g	b c d	b c
120_3F	a b	a b	a b	a	a	a b	b c d	c d e	e f g	c d	c
180_0F	b c	a b c d	a b c d	a	a	a b	a b	a	a	a	a
180_1F	a b c	a b c d	a b c d	a	a	a b	a b c d	a b c d e	c d e f	b c	b
180_2F	b c	a b c d	a b c d	a	a	a b	b c d	c d e	d e f g	b c d	b c
180_3F	a b c	a b c d	a b c d	a	a	a b	c d	d e	f g	d	c
200_0F	c	b c d	a b c d	a	a	a	a	a	a	a	a
200_1F	b c	b c d	a b c d	a	a	a b	a b c d	a b c d e	c d e	b	b
200_3F	c	c d	b c d	a	a	a b	c d	d e	f g	d	c
260_0F	c	d	c d	a	a	a b	a b	a b	a b	a	a
260_1F	c	d	d	a	a	a b	a b c	a b c	b c d	b	b
260_2F	c	c d	b c d	a	a	a b	b c d	c d e	d e f g	b c d	b c
260_3F	c	d	b c d	a	a	b	d	e	g	d	c

Table 5.6: Comparison of variable between the whole leaf and the green leaf, using a paired t-test. The p-value indicates the level of statistical significance of the difference between the two groups.

	BRF_490	BRF_550	BRF_680	BRF_720	BRF_800	BRF_900	NDVI	NDRE	PSRI	mNDb	CIrede
02/06/2021	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
16/06/2021	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
17/03/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.457	<0.01
11/04/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
19/04/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
25/04/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
02/05/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
09/05/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.05	<0.01	<0.01	<0.01	<0.01	<0.01
17/05/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
23/05/2022	<0.05	0.083	<0.01	<0.05	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
30/05/2022	<0.01	0.063	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
02/06/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
13/06/2022	<0.01	<0.01	<0.01	<0.01	0.825	0.309	<0.01	<0.01	<0.01	<0.01	<0.01
21/06/2022	<0.01	<0.01	<0.01	<0.01	<0.01	0.573	<0.01	<0.01	<0.01	<0.01	<0.01
04/07/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Table 5.7: Comparison of variable between the whole leaf and the image, using a paired t-test. The p-value indicates the level of statistical significance of the difference between the two groups.

	BRF_490	BRF_550	BRF_680	BRF_720	BRF_800	BRF_900	NDVI	NDRE	PSRI	mNDb	CIrede
02/06/2021	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
16/06/2021	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
17/03/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
11/04/2022	0.371	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
19/04/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
25/04/2022	<0.01	<0.05	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
02/05/2022	0.380	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
09/05/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
17/05/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
23/05/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
30/05/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.142	<0.01
02/06/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.843	<0.01
13/06/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
21/06/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
04/07/2022	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Table 5.8: Pearson's correlation between $\Delta_{\text{leaves-green}}$ and the damage index.

	BRF_490	BRF_550	BRF_680	BRF_720	BRF_800	BRF_900	NDVI	NDRE	PSRI	mNDb	CIrede
02/06/2021	-0.55	-0.52	0.02	-0.84	-0.95	-0.95	-0.55	-0.62	0.48	-0.63	-0.46
16/06/2021	0.19	0.13	0.33	-0.05	-0.58	-0.62	-0.71	-0.51	0.54	-0.36	-0.11
17/03/2022	0.63	0.35	0.74	-0.07	-0.68	-0.66	-0.77	-0.78	0.80	0.40	-0.70
11/04/2022	0.59	0.29	0.64	0.35	-0.04	-0.01	-0.76	-0.72	0.69	-0.56	-0.57
19/04/2022	0.88	0.92	0.97	0.85	-0.68	-0.62	-0.98	-0.97	0.98	-0.97	-0.93
25/04/2022	0.44	0.58	0.89	-0.42	-0.81	-0.80	-0.91	-0.89	0.92	-0.81	-0.81
02/05/2022	0.88	0.96	0.96	0.95	0.67	0.74	-0.95	-0.95	0.95	-0.97	-0.76
09/05/2022	0.17	0.67	0.81	0.60	-0.40	-0.21	-0.84	-0.70	0.89	-0.75	-0.53
17/05/2022	0.45	0.73	0.92	0.69	-0.27	-0.23	-0.93	-0.92	0.93	-0.93	-0.84
23/05/2022	-0.46	-0.39	0.80	-0.36	-0.78	-0.74	-0.96	-0.64	0.98	-0.65	-0.58
30/05/2022	-0.76	0.21	0.98	0.04	-0.71	-0.71	-0.98	-0.91	0.99	-0.88	-0.88
02/06/2022	0.94	0.96	0.98	0.95	0.87	0.87	-0.99	-0.98	0.99	-0.97	-0.94
13/06/2022	0.89	0.91	0.97	0.90	0.73	0.79	-0.99	-0.96	0.99	-0.96	-0.86
21/06/2022	0.91	0.90	0.98	0.89	0.47	0.66	-0.96	-0.93	0.96	-0.90	-0.81
04/07/2022	0.97	0.90	0.93	0.93	0.92	0.93	-0.78	-0.81	0.89	-0.45	-0.17

Table 5.9: Pearson's correlation between $\Delta_{3F\text{-}xF}$ and the damage index.

	BRF_490_{green}	BRF_550_{green}	BRF_680_{green}	BRF_720_{green}	BRF_800_{green}	BRF_900_{green}	$NDVI_{green}$	$NDRE_{green}$	$PSRI_{green}$	$mNDb_{green}$	$Cirede_{green}$
02/06/2021	-0.04	-0.20	-0.15	-0.04	0.53	0.49	0.60	0.70	-0.26	0.65	0.67
16/06/2021	0.00	-0.24	-0.24	-0.22	0.43	0.33	0.47	0.67	-0.41	0.73	0.60
17/03/2022	-0.16	-0.15	-0.16	-0.15	-0.07	-0.03	0.21	0.31	-0.14	-0.04	0.35
11/04/2022	-0.16	-0.23	-0.18	-0.30	0.03	0.00	0.24	0.53	-0.50	0.33	0.60
19/04/2022	-0.40	-0.48	-0.47	-0.47	0.03	0.06	0.52	0.54	-0.36	0.52	0.49
25/04/2022	-0.22	-0.28	-0.31	-0.18	0.20	0.17	0.32	0.34	-0.29	0.34	0.34
02/05/2022	-0.28	-0.36	-0.45	-0.39	-0.05	-0.02	0.45	0.40	-0.65	0.42	0.38
09/05/2022	-0.35	-0.29	-0.27	-0.27	-0.21	-0.06	0.23	0.27	-0.03	0.20	0.28
17/05/2022	-0.37	-0.59	-0.58	-0.61	0.54	0.31	0.63	0.70	-0.66	0.73	0.72
23/05/2022	-0.56	-0.43	-0.59	-0.30	0.44	0.33	0.89	0.78	-0.77	0.66	0.79
30/05/2022	-0.85	-0.94	-0.93	-0.95	0.63	0.53	0.93	0.96	-0.96	0.97	0.95
02/06/2022	-0.57	-0.83	-0.83	-0.88	0.85	0.82	0.87	0.93	-0.95	0.96	0.90
13/06/2022	-0.59	-0.73	-0.88	-0.60	0.88	0.89	0.96	0.94	-0.98	0.95	0.90
21/06/2022	-0.35	-0.36	-0.45	-0.25	0.71	0.78	0.90	0.90	-0.83	0.86	0.91
04/07/2022	0.62	0.82	-0.78	0.90	0.95	0.95	0.98	0.92	-0.98	0.67	0.87

Table 5.10: R^2 of the linear regression associating several feature combinations with agronomic data. GF stands for green fraction and DI for damage index.

Date	Features	Nuptake total	Nuptake leaves	Nuptake stem	Nuptake Ears	%N total	%N leaves	%N stem	%N Ears	NNI
2021-GS39	$NDRE_{leaves}$	0.49	0.33	0.52		0.65	0.62	0.62		0.65
	$NDRE_{green}$	0.48	0.32	0.50		0.67	0.63	0.64		0.65
	$NDRE_{green}$ + GR	0.53	0.34	0.62		0.67	0.64	0.64		0.66
	$NDRE_{green}$ + GF + DI	0.53	0.34	0.62		0.68	0.64	0.64		0.66
2021-GS65	$NDRE_{leaves}$	0.43	0.49	0.46	0.05	0.66	0.63	0.62	0.38	0.58
	$NDRE_{green}$	0.42	0.47	0.45	0.05	0.66	0.63	0.61	0.37	0.57
	$NDRE_{green}$ + GR	0.46	0.50	0.48	0.15	0.66	0.68	0.61	0.37	0.58
	$NDRE_{green}$ + GF + DI	0.46	0.50	0.48	0.15	0.67	0.68	0.62	0.39	0.58
2022-GS30	$NDRE_{leaves}$	0.13	0.13			0.17	0.17			0.20
	$NDRE_{green}$	0.13	0.13			0.17	0.17			0.20
	$NDRE_{green}$ + GR	0.20	0.20			0.17	0.17			0.23
	$NDRE_{green}$ + GF + DI	0.21	0.21			0.42	0.42			0.32
2022-GS32	$NDRE_{leaves}$	0.61	0.58	0.52		0.53	0.59	0.33		0.63
	$NDRE_{green}$	0.59	0.57	0.49		0.54	0.61	0.33		0.62
	$NDRE_{green}$ + GR	0.59	0.57	0.50		0.54	0.62	0.33		0.62
	$NDRE_{green}$ + GR+ DI	0.59	0.57	0.50		0.59	0.62	0.47		0.63
2022-GS39	$NDRE_{leaves}$	0.28	0.38	0.08		0.27	0.50	0.01		0.30
	$NDRE_{green}$	0.32	0.41	0.11		0.27	0.55	0.01		0.33
	$NDRE_{green}$ + GR	0.34	0.42	0.15		0.28	0.56	0.02		0.35
	$NDRE_{green}$ + GF + DI	0.38	0.52	0.15		0.39	0.64	0.02		0.43
2022-GS65	$NDRE_{leaves}$	0.18	0.19	0.03	0.32	0.00	0.69	0.03	0.01	0.08
	$NDRE_{green}$	0.18	0.18	0.03	0.35	0.00	0.69	0.03	0.02	0.08
	$NDRE_{green}$ + GR	0.18	0.23	0.03	0.35	0.04	0.69	0.06	0.22	0.09
	$NDRE_{green}$ + GR+ DI	0.47	0.48	0.28	0.40	0.50	0.85	0.29	0.35	0.54
2022-GS73	$NDRE_{leaves}$	0.57	0.50	0.06	0.43	0.02	0.48	0.29	0.08	0.29
	$NDRE_{green}$	0.59	0.51	0.06	0.46	0.02	0.49	0.30	0.09	0.31
	$NDRE_{green}$ + GR	0.60	0.54	0.06	0.49	0.06	0.53	0.30	0.17	0.34
	$NDRE_{green}$ + GF + DI	0.67	0.70	0.14	0.51	0.11	0.57	0.32	0.18	0.41
All	$NDRE_{leaves}$	0.54	0.64	0.41	0.01	0.39	0.01	0.00	0.17	0.46
	$NDRE_{green}$	0.58	0.61	0.43	0.00	0.44	0.03	0.00	0.16	0.46
	$NDRE_{green}$ + GR	0.59	0.63	0.44	0.16	0.56	0.35	0.05	0.35	0.47
	$NDRE_{green}$ + GF + DI	0.64	0.63	0.59	0.40	0.72	0.50	0.10	0.35	0.48

6

Comparing CNNs and Traditional Machine Learning for Estimating Wheat Organs Biophysical Variables Using Proximal Sensing

1. Synopsis

This chapter relies on the following under-review paper:

Carlier, A., Dandrifosse, S., Dumont, B. & Mercatoris, B., 2023. Comparing CNNs and Traditional Machine Learning for Estimating Wheat Organs Biophysical Variables Using Proximal Sensing. Frontiers in Plant Science 2023

Biophysical vegetation variables are indicators of plant growth and health, providing essential information for understanding plant-environment interactions [Hawkesford and Riche, 2020; Lemaire and Ciampitti, 2020]. Leaf area index (LAI), aboveground biomass (AGB), and nitrogen uptake (Nupt) are some of the most critical variables that can aid in crop monitoring and yield prediction. These variables can also help understand the physiological processes that govern the associations between final yield, genotype and surrounding environment, and determine traits related to these processes such as the harvest index, the remobilization efficiency, and the nitrogen use efficiency. As concerns about climate change and its impact on human food security continue to grow, there is a growing need for reliable and efficient techniques for the quantitative assessment of vegetation variables [Hickey et al., 2019].

Convolutional neural networks (CNNs) have shown great potential in the field of phenotyping research. However, further investigation is necessary to determine their effectiveness in estimating biophysical variables such as AGB, LAI, nitrogen concentration, and nitrogen uptake from proximal images of wheat. To address this, this chapter conducted a thorough investigation of the use of CNNs for this purpose across several growing seasons, testing various CNN architectures and employing a customized training pipeline. The pipeline included a first training phase using transfer learning and a second phase using a customized pseudo-labeling method, which utilized the large amount of unlabeled data and biological temporal curve to correct the pseudo-labels. Models were also trained to predict the dry matter and nitrogen uptake partitioning among the different wheat organs. To evaluate the gain of such methods, the results were compared to a traditional machine learning approach. The chapter includes all the elements included in the corresponding research document with very few adaptations for this thesis.

2. Current State of the Art and Objectives

Recent developments in phenotyping systems have revolutionized the way crops are screened in high-throughput and non-destructive ways [Araus et al., 2022a; Reynolds et al., 2020; Sun et al., 2022]. These systems are able to measure most biophysical variables using different vision-based methods [Verrelst et al., 2019]. Methods for estimating above-ground biomass (AGB) and leaf area index (LAI) rely on crop architecture, vegetation indices, radiative transfer models, or a combination of these models

[Brocks and Bareth, 2018; Raj et al., 2021; Schiefer et al., 2021; Tilly et al., 2015; Wan et al., 2021; Yue et al., 2019]. Such methods are also widely used for assessing crop nitrogen status [Berger et al., 2020].

Assessing biophysical variables using remote sensing and proximal sensing methods requires expertise in both agronomy and data analysis, as these traits are highly complex and influenced by numerous factors. Achieving a comprehensive understanding of biophysical variables demands a holistic approach, considering a wide range of environmental, genetic and management variables [Hawkesford and Riche, 2020]. However, most of the phenotyping studies often focus on a local area or agricultural management, which limits their wider applications and the generalization of proposed models [Chao et al., 2019]. Therefore, it is important to design studies that capture the diversity of the crop populations and the variability of the growing conditions. External factors such as soil, weeds and biotic and abiotic stresses, and plant characteristics such as growth stage, canopy architecture significantly limit the possibility of accurate measurements. While remote sensing which have lower spatial resolution can capture the canopy in its entirety, it may be affected by these factors. In contrast, proximal sensing at the organ level can provide more precise measurements and limit the impact of them [Deery et al., 2014]. In fact, ground-based phenotyping systems equipped with multiple sensors can acquire high-resolution data, enabling better identification of plant organs, diseases, or yellow and green plant parts [Carlier et al., 2022; Dandrifosse, 2022; Serouart et al., 2022; Tanner et al., 2022; Xu and Li, 2022].

Recent advances in deep learning algorithms, such as convolutional neural networks (CNNs), have further enhanced the capabilities of these systems and opened up new avenues for the phenotyping community [Arya et al., 2022; Kattenborn et al., 2021; Singh et al., 2018]. CNNs have gained significant attention in the field of crop phenotyping due to their ability to outperform traditional machine learning algorithms and exhibit better generalization given sufficient data. For example, in estimating wheat biomass at early growth stages, CNNs were less influenced by plant density compared to other methods [Ma et al., 2019]. These novel methods improve the capacities to better estimate traits and can unlock the extraction of more advanced parameters, such as crop growth rate, when applied to time-series data [Buxbaum et al., 2022]. The availability of ready-to-use libraries, datasets, and emerging methodologies such as transfer learning have made these sophisticated algorithms transferable to crop characterization. However, it is crucial to ensure the accuracy and robustness of these models by validating them with ground-truth data and regression CNN is still in its infancy [Jiang and Changying Li, 2020].

Although preliminary attempts have been made to estimate crop yield, CNNs still face some challenges in phenotyping. The ever-growing availability of neural networks architectures and hyperparameters can present a challenge when it comes to selecting or designing the most suitable architecture. While some authors have success-

fully created their own neural architectures that perform comparably to well-known ones in terms of accuracy [Li et al., 2021b], it is still highly recommended to use established and widely recognized architectures. Furthermore, CNN models require a large amount of annotated or ground-truth data to determine the numerous trainable parameters. This is particularly challenging when dealing with biophysical variables, such as AGB, which require a significant amount of human labor to construct a dataset. To address this need of data, several methods have been proposed to train robust models with a limited amount of labeled data. One approach is to use pre-trained models with transfer learning, which has been successful in estimating forage biomass [Castro et al., 2020; de Oliveira et al., 2021]. However, when dealing with multispectral images, pre-trained models that are generally trained on RGB images may not perform well. Another approach is to use data augmentation to artificially increase the dataset size by applying transformations to the images. Advanced data augmentation methods, such as generative adversarial networks (GANs), have been used to improve wheat yield estimation [Zhang et al., 2022]. Finally, phenotyping centers often acquire large amounts of unlabeled data that still can be used to train CNNs. For this purpose, semi-supervised learning methods allow training the convolutional parts of CNNs from unlabeld datasets to retrieve relevant feature. For instance, the pseudo-labeling approach involves predicting unlabeled data that are subsequently inserted into the training dataset [Lee, 2013].

3. Materials and Methods

3.1. Machine Learning Approach

A conventional machine learning approach was tested to confront the CNN models. As machine learning algorithms require relevant image features to be extracted, additional processes were applied after performing the pre-processing steps as described in Chapter 3. Firstly, a stereovision process was used to extract plant height information using the 95th percentile of the height map (Section 4.3). Secondly, the plant ratio was computed as the proportion of plants in the scene, using a simple threshold method on the 800 nm image as detailed in (Section 4.2). Finally, three vegetation indices, namely the NDVI [Rouse et al., 1974], NDRE [Barnes et al., 2000], and CIrede [Gitelson et al., 2006], were calculated and averaged on the entire image and on the plant, i.e., after segmenting (Section 4.5).

A partial least square regression (PLSr) model was trained and validated using these features for DM, %N and Nupt of the entire plant. The trials used here were those presented in Table 3.1. Notably, the training data did not include the 2019 data, as only one RGB camera was available at that time. Additionally, the pseudo-labeling strategy mentioned in Section 3.2.3 was also tested for the PLSr.

3.2. CNN Training

3.2.1. Architecture

Three CNN architectures available in the python library Tensorflow 2.4. and Keras 2.4 were tested in this study. They were Resnet50 [He et al., 2015] and EfficientNetB0 and B4 [Tan and Le, 2020]. They represent the actual state-of-the-art CNN models with different properties (i.e., architecture and number of parameters) and purposes.

The CNN architectures were customized to perform two tasks: (i) a single-output model to estimate LAI, DM, %N and Nupt of the whole plant respectively; and (ii) a multi-output model to estimate organ-specific DM, %N, and Nupt, respectively. A linear activation function was considered for the last neuron of each single-output model. Regarding the multi-output models, four output neurons were considered, one for each organ. A linear activation function was used for the estimation of %N whereas softmax activation function were used for DM and Nupt to keep the values between 0 and 1. All models were initialized with weights from the ImageNet dataset [Deng et al., 2009].

The CNN architectures were originally designed for three-channel images, but the multispectral images used in this study had six channels. To accommodate this, a 2D convolutional layer with three filters and a kernel size of (1,1) was added at the beginning of each model when using multispectral images. It allowed to provide a three channels input required for the CNN models with pre-trained weights.

3.2.2. Dataset Configuration

In this chapter, all trials presented in Table 3.1 were used. Thus, the study used a dataset consisting of 1809 RGB images and 1391 multispectral images. These numbers correspond to the multiplication of the dates, samples, replicates, and images per plot. Each image was associated with a specific combination of agronomic variables. From this dataset, two treatments from F trials (Table 3.2 and Table 3.3), and one treatment from FP trials (Table 3.4) were selected for the validation dataset that included 424 RGB images and 341 multispectral images.

In addition to the images acquired on the same days as the manual sampling, each trial was monitored continuously throughout the season, as illustrated in Figure 3.4. All those acquisitions yielded a dataset comprising 16 812 RGB images and 14 491 multispectral images. To prepare the data for the CNN models, some pre-processing steps were taken.

The first pre-processing step involved determining the image size, which is a trade-off between retaining as much information as possible and limiting the computing time and resources required. Additionally, when using pre-trained models, it is recommended to set the input image size to match the size used during initial training. Therefore, all images were resized to 224 x 224 for the ResNet50 and EfficientNetB0

models, and to 380 x 380 for the EfficientNetB4 model. It is worth noting that the images were previously cropped into a square to avoid distortion.

In addition to image resizing, the pixel scaling was also adjusted for each model. For the RGB images, pixel scaling was adapted according to the Keras documentation and the requirements of each model. For the multispectral images, Bi-directional Reflectance Factor (BRF) values were first normalized between 0 and 1. Next, the data was standardized based on the mean and standard deviation of the training dataset. This ensures that the multispectral images are on the same scale as the RGB images, making it easier to combine the two types of images during training. To further enhance the dataset, data augmentation techniques, namely random flip up/down and right/left, were applied. These techniques increase the diversity of the dataset, which can improve the generalization performance of the models.

3.2.3. Training Pipeline

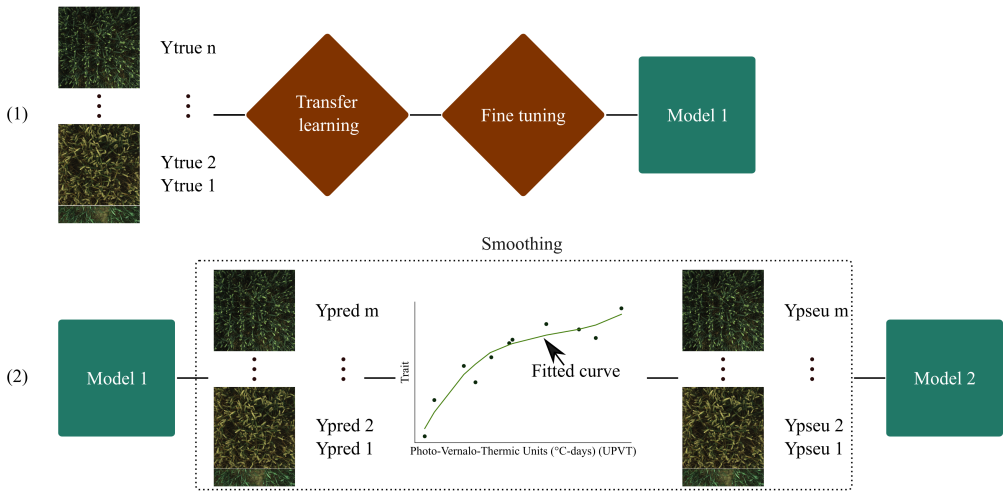

Figure 6.1: Proposed training pipeline. (1) is the training with transfer learning, and (2) is the training with pseudo-labels. Ytrue corresponds to the reference measurements of dry matter or nitrogen uptake of the plant and of each organ. Ypred are the predicted labels and Ypse the smoothed predicted values. n and m correspond to the number of reference measurements and total images respectively.

Exploiting unlabeled data to improve the performance of machine learning models is an underused approach. It relies on pseudo-labels, which can be introduced into the training dataset if they meet a certain confidence threshold. In classification tasks, pseudo-labels are assigned based on the predicted class probabilities. However, in regression tasks, the last activation function is linear, so probabilities are not calculated. To address this issue and correct the predicted values, the data from each micro-plot were plotted against time to generate a crop growth curve. This curve captures the

expected growth pattern of the crop over time and can be used to correct the predicted values. Specifically, the predicted values are adjusted to fit the growth curve, which can help to improve the accuracy of the regression models used in this study.

Based on this idea, a pipeline was designed (Figure 6.1). The pipeline involved training CNN models pre-trained on ImageNet using transfer learning for 40 epochs with a learning rate of 1×10^{-3}, followed by fine-tuning for 10 epochs with a learning rate of 1×10^{-5}. The transfer learning approach involved training only the last layer, which was the linear dense layer, while freezing the rest of the model. The last convolutional layer block was then unfrozen and retrained to obtain the Model 1.

Then, the Model 1 was used to generate predicted labels (Ypred) for the entire training dataset. Ypred labels were plotted against the Photo-Vernalo-Thermic Units (°C-days) [Duchene et al., 2021], and smoothed using a cubic B-Spline with a high smoothing condition for LAI and a cubic polynomial function for other variables. Basic correction conditions were also added, such as setting organ values to 0 when they were not present at a specific time. The resulting "corrected" pseudo-labels (Ypseu) were used in the second stage.

In the second stage, pre-trained CNNs from ImageNet were fine-tuned on Ypseu for 30 epochs with a learning rate of 1×10^{-5}. This resulted in the development of Model 2, which showed improved performance compared to Model 1.

The Mean Square Error (MSE) loss function and Adam optimizer were used in all models. However, in the case of %N of the organs, the MSE calculation was limited to true labels above 0. This means that if an organ was not yet visible (e.g., the ear during tillering growth stage), the loss function did not take it into account, which prevented it from interfering with the loss function. Additionally, a weight was applied to the loss calculation when working with relative values. Specifically, the flag leaf pool weights were multiplied by twenty to ensure consistency with the order of magnitude of the other organ pools. This helped to balance the contributions of different organ pools and prevent one pool from dominating the loss calculation. All models were trained on an NVidia Tesla V100 GPUs.

To evaluate the performance of all models, two metrics were used: the determination coefficient (R^2) and the root mean square error (RMSE).

4. Results

4.1. Biophysical Vegetation Variable Models

At the plant level, the EfficientNetB4 model trained on pseudo-labels achieved the best performance for DM (Table 6.1), with a R^2 of 0.92 and a RMSE of 1.50 on the validation dataset. The PLSr model had the lowest R^2 of 0.76 and the higher RMSE of 2.67. The ResNet50 trained on pseudo-labels had the highest R^2 of 0.82 (Table 6.2) for

Table 6.1: Model performances for DM of the plant.

Model	Data	RMSE train	R² train	RMSE val	R² val
EfficientNetB0	Ytrue	1.76	0.91	2.11	0.83
EfficientNetB0	pseudo	0.91	0.95	1.66	0.91
EfficientNetB4	Ytrue	1.38	0.93	1.89	0.86
EfficientNetB4	pseudo	1.09	0.96	1.50	**0.92**
Resnet50	Ytrue	0.78	0.98	1.87	0.89
Resnet50	pseudo	1.05	0.97	1.64	0.90
PLSr	Ytrue	2.88	0.74	2.67	0.76
PLSr	Ypseu	2.95	0.73	2.67	0.76

Table 6.2: Model performances for LAI.

Model	Data	RMSE train	R² train	RMSE val	R² val
EfficientNetB0	Ytrue	1.06	0.69	1.18	0.57
EfficientNetB0	Ypseu	0.72	0.86	0.80	0.80
EfficientNetB4	Ytrue	0.68	0.86	0.78	0.80
EfficientNetB4	Ypseu	0.67	0.87	0.78	0.81
Resnet50	Ytrue	0.27	0.98	0.79	0.79
Resnet50	Ypseu	0.66	0.88	0.78	**0.82**
PLSr	Ytrue	0.83	0.71	0.88	0.72
PLSr	Ypseu	0.84	0.71	0.93	0.68

LAI. Using the multispectral images, the best R² values were obtained with ResNet50 (0.80, Table 6.4) for %N and EfficientNetB0 (0.73, Table 6.3) for Nupt.

The other CNN models also showed very good and similar performances with the pseudo-labeling training pipeline. The differences between the results on the validation and training datasets were reduced when using pseudo-labels. In contrast, PLSr performance did not benefit from the pseudo-labeling method and was consistently worse than that of the CNN models.

The models were able to successfully assess the variables throughout the growing season, as evidenced by Figures 6.2 and 6.3. However, there were some outliers that significantly deviated from the ideal 1:1 relationship between predicted and true values. Additionally, we noticed a saturation effect where the models were unable to accurately capture the maximum values of each variable, leading to a lack of detail in certain growing seasons. The patterns observed in the data appeared scattered like a cloud rather than forming a clear line, indicating that there is still room for improvement in the models. Therefore, further refinement of the models is necessary to address

Table 6.3: Model performances for Nupt of the plant.

Model	Data	RMSE train	R² train	RMSE val	R² val
EfficientNetB0	Ytrue	42.42	0.65	37.97	0.66
EfficientNetB0	Ypseu	28.84	0.84	34.38	0.72
EfficientNetB4	Ytrue	37.46	0.70	43.39	0.47
EfficientNetB4	Ypseu	25.33	0.87	34.05	0.69
Resnet50	Ytrue	14.24	0.96	37.27	0.68
Resnet50	Ypseu	26.69	0.86	33.89	**0.73**
PLSr	Ytrue	38.84	0.67	42.92	0.66
PLSr	Ypseu	40.06	0.65	46.24	0.60

Table 6.4: Model performances for %N of the plant.

Model	Data	RMSE train	R² train	RMSE val	R² val
EfficientNetB0	Ytrue	0.37	0.74	0.36	0.72
EfficientNetB0	Ypseu	0.24	0.90	0.30	0.79
EfficientNetB4	Ytrue	0.33	0.75	0.32	0.55
EfficientNetB4	Ypseu	0.23	0.89	0.31	0.73
Resnet50	Ytrue	0.14	0.97	0.32	0.78
Resnet50	Ypseu	0.24	0.90	0.30	**0.80**
PLSr	Ytrue	0.56	0.56	0.41	0.58
PLSr	Ypseu	0.61	0.47	0.40	0.59

these issues and improve their accuracy.

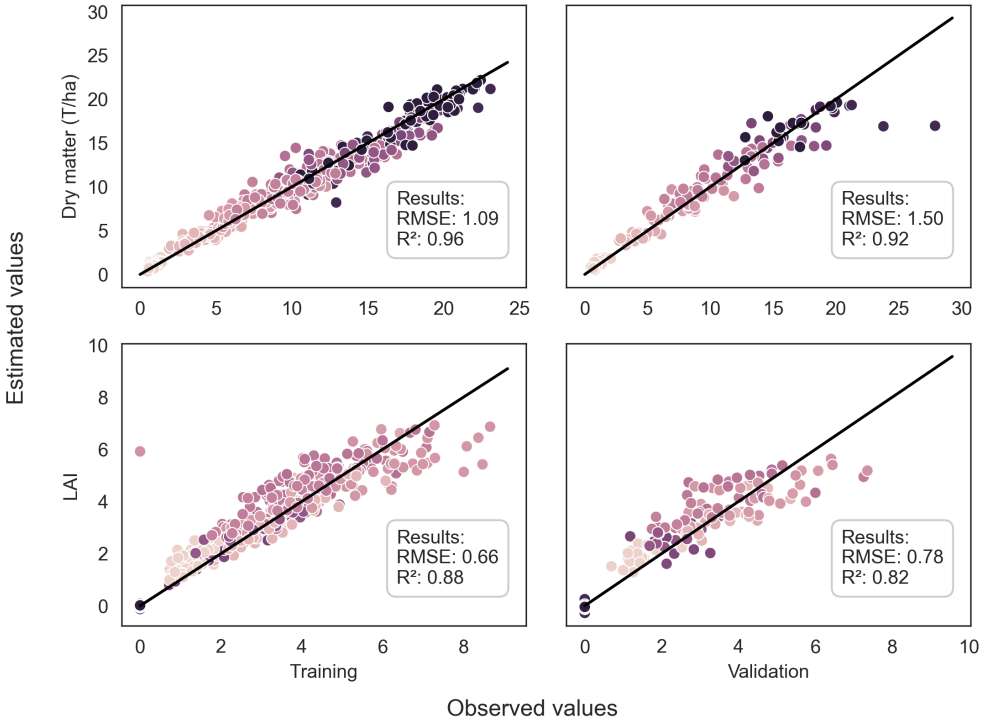

Figure 6.2: Comparison between observed and estimated values of DM of the whole plant and LAI for both training and validation datasets, using the EfficientNetB4 model for DM and the Resnet50 model for LAI. The dots are color-coded according to the stages in the season, with darker dots indicating later stages. The dark line represents the 1:1 line.

4.2. Temporal Wheat Dry Matter and Nitrogen Uptake Partitioning

The multi-output models produced diverse results for dry matter and nitrogen uptake partitioning, as indicated in Table 6.5. The ear and stems organs were more accurately estimated by the models, while the flag leaf (L1) was poorly estimated. In terms of %N models, the tiller and inferior leaf pools were best estimated.

EfficientNetB0 was found to be the best model for estimating Nupt at the organ level, with R^2 values of 0.7, 0.59, 0.69, and 0.86 for stem, Linf, L1 and ear, respectively. The pseudo-labeling method improved its performance slightly. However, this method did not improve the estimation of DM and %N of each organs. For both, the Resnet50 model was found to be the most effective, as shown in Table 6.5.

Figure 6.4 displays the estimated partitioning of wheat dry matter and nitrogen up-

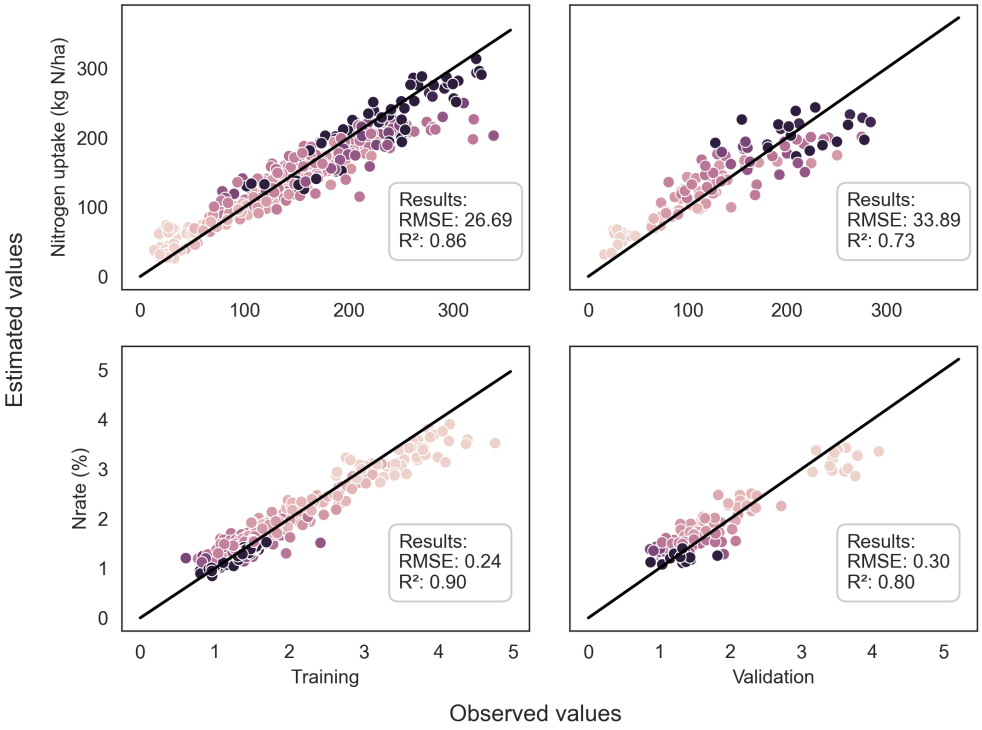

Figure 6.3: Comparison between observed and estimated values of %N and Nupt of the whole plant for both training and validation datasets, using the Resnet50 model. The dots are color-coded according to the stages in the season, with darker dots indicating later stages. The dark line represents the 1:1 line.

take over the growing season for a single micro-plot. These curves represent raw predicted data that have not undergone any filtering process, but they can be smoothed if necessary using a method similar to pseudo-labeling. Both the RGB and multispectral models successfully detected the emergence of new organs, including the flag leaf and ear. Notably, the dry matter model shows an earlier appearance of ears compared to the nitrogen uptake model in this particular example.

5. Discussion

The aim of this study was to demonstrate the effectiveness of recent CNNs in assessing important biophysical vegetation variables, such as dry matter, LAI, and nitrogen uptake and concentration. Traditionally, accurately measuring these traits from remote sensing data has required specialized knowledge in agronomy and computer vision. However, CNNs can learn to extract relevant features directly from the data, eliminating the need for manual feature extraction.

Table 6.5: R^2 of the different models to estimate DM, Nupt and %N of each organ using the multi-output models.

Model	Data	Dataset	Nuptake				DM				%N			
			Tiller	Linf	L1	Ear	Tiller	Linf	L1	Ear	Tiller	Linf	L1	Ear
EfficienNetB0	Ypseu	train	0.84	0.78	0.77	0.92	0.91	0.49	0.08	0.97	0.71	0.86	-0.08	-2.18
	Ytrue	train	0.72	0.67	0.7	0.91	0.90	0.49	0.17	0.96	0.60	0.75	0.60	-1.09
EfficienNetB4	Ypseu	train	0.54	0.02	0.65	0.75	0.90	0.53	-0.11	0.96	0.67	0.75	-0.10	-1.95
	Ytrue	train	0.6	0.48	-0.47	0.7	0.92	0.65	0.57	0.97	0.38	0.56	0.10	-2.46
Resnet50	Ypseu	train	0.8	0.75	0.76	0.93	0.88	0.51	0.15	0.94	0.73	0.87	-0.05	-2.10
	Ytrue	train	0.84	0.78	0.8	0.96	0.93	0.76	0.67	0.98	0.86	0.93	0.95	0.52
EfficienNetB0	Ypseu	val	0.7	0.59	0.69	0.86	0.83	0.28	-0.09	0.95	0.54	0.84	-0.43	-2.64
	Ytrue	val	0.63	0.47	0.54	0.86	0.82	0.34	-0.16	0.93	0.47	0.69	0.58	-1.72
EfficienNetB4	Ypseu	val	0.51	-0.2	0.52	0.66	0.83	0.38	-0.15	0.94	0.59	0.73	-0.46	-2.11
	Ytrue	val	0.41	0.3	-1.66	0.65	0.84	0.55	0.14	0.94	0.10	0.55	0.13	-3.91
Resnet50	Ypseu	val	0.64	0.52	0.64	0.85	0.87	0.57	0.22	0.94	0.48	0.84	-0.45	-3.17
	Ytrue	val	0.67	0.56	0.66	0.86	0.87	0.62	0.38	0.94	0.50	0.76	0.69	-1.07

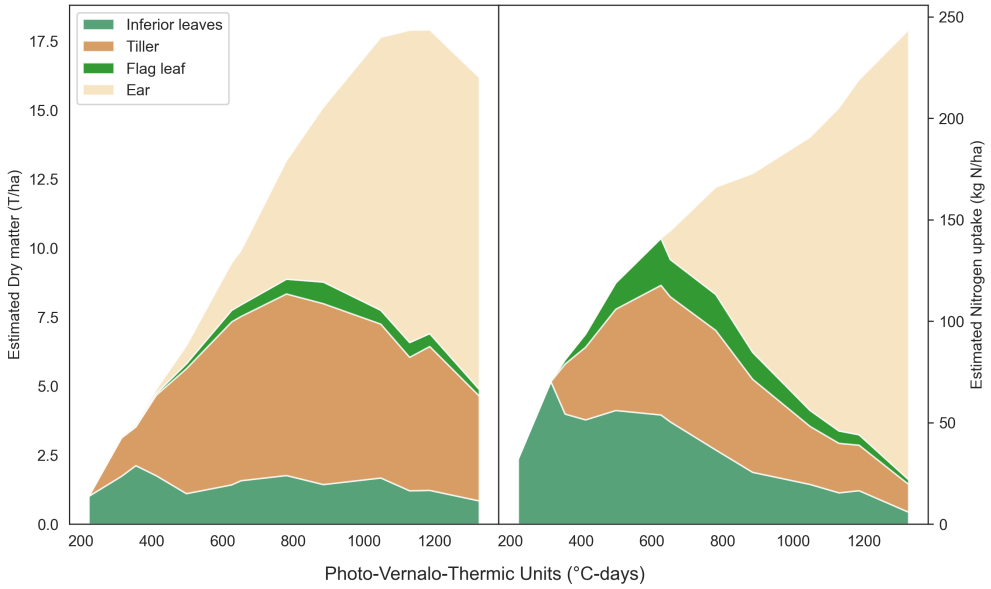

Figure 6.4: Estimated partitioning of dry matter and nitrogen uptake throughout the season for a micro-plot from the 22-F trial.

In this study, transfer learning, a common training method, outperformed traditional machine learning approaches, consistent with prior research [Castro et al., 2020; Ma et al., 2019]. This underscores the exceptional ability of convolutional neural network models to extract significant features, although interpreting their meaning remains a challenge. Some research is devoted to developing techniques to explain CNN predictions [Selvaraju et al., 2020]. Resnet50 has demonstrated good performance in similar studies [Castro et al., 2020], but EfficientNet has also produced promising results and was particularly effective for DM of the whole plant. However, it could be argued that

the advantage of EfficientNetB4 was due to its ability to capture finer details in larger images. Interestingly, recent research has shown that having larger images does not always result in better performance beyond a certain level [Li et al., 2021b]. This outcome may be architecture-dependent, as EfficientNet is specifically designed for easy scaling to optimize performance for specific datasets [Tan and Le, 2020].

Deep learning, especially in regression topics like biophysical variables, faces a significant challenge due to the scarcity of sufficient datasets for training purposes. To overcome this limitation, a training pipeline was devised to make use of the abundance of unlabeled data available in every high-throughput phenotyping installation. This proposed approach employs the temporal aspect of the data to correct pseudo-labels, leading to significant improvements in the performance of CNN models.

The smoothing method utilized in this study employed a cubic curve due to its ease in representing biophysical curves and its ease to be fitted. Finer curves, such as P-splines or logistic curves [van Eeuwijk et al., 2019], were also tested but the fitting process proved to be difficult. As noted by [Roth et al., 2020], one to two measurements per week are necessary to accurately fit such curves, but our data did not meet this frequency requirement. To limit the introduction of bias, we added some correcting conditions, which were particularly important for organ models. For instance, when an organ was not present at time t, we set the corresponding pseudo-label to 0, which may seem like a common-sense correction, but it was valuable.

The pseudo-labeling method effectively reduced overfitting of the model, as demonstrated by the smaller performance gap between the training and validation datasets. This suggests that the smoothed models were able to identify more general features than the transfer learning features method. During the research, more advanced data augmentation methods were also tested such as 90° rotation and color space transformations. However, caution should be exercised when applying such methods, as they may negatively impact model performance in certain cases, as observed in some models in [Castro et al., 2020]. On the other hand, [Ma et al., 2019] reported clear performance improvements with such methods. One possible explanation for this discrepancy is that the model may become overly reliant on certain features, such as the wheat lines in the case of image rotation. Therefore, data augmentation should be used judiciously.

An effective approach for evaluating model performance is to combine their predictions into a single variable. In this study, we used DM and %N of the plant estimated from their respective models to calculate the NNI. The R^2 values for the training and validation datasets were 0.71 and 0.33, respectively, indicating that this method could be useful for measuring NNI. Although these models require improvements in terms of trait saturation and lack of accuracy within specific growth stage, their potential is significant. They can be used to compute advanced traits such as growth rate and spot ideotypes using temporal curves, as demonstrated in a recent study [Roth et al.,

2022a].

The machine learning approach that combines features from the multi-sensor system consistently underperformed CNN models. Nonetheless, this approach still exhibited reasonable performance, with an R^2 value exceeding 0.6, akin to Yue et al. [2019]. This method could serve as a viable option for small datasets and may also be utilized as a pseudo-labeling tool.

Research on the allocation of major plant elements, such as sink/source regulation processes and their relationship with grain nitrogen content, is primarily based on dry matter and nitrogen uptake partitioning [Gaju et al., 2014; Martre et al., 2003]. The partitioning models proposed in this study have yielded promising results, as demonstrated in Table 6.5, with good performance in most cases. However, certain organs exhibited poor performance, such as %N of the ear, which could be attributed to the lack of visible traits that could account for it, such as a greener ear. The poor performance of DM and Nupt for L1 could mainly be attributed to the bad performances of the multi-ouputs model for this organ, which also showed poor performance in Table 6.6. Indeed, the flag leaf was not well-accounted for in the multi-task model, despite having a higher weight in the loss function. Some other tested weight values did not improved the results. It's possible that the models had too few images with only flag leaves, as the ear develops quite rapidly behind them.

The multi-ouput or multi-task model represent an exciting opportunity for plant phenotyping. In this study, a simple approach was used by sharing the loss function, but the benefits of multi-task learning can be substantial. For example, a single model that assesses both dry matter (DM) and leaf area index (LAI) can greatly reduce computational costs and processing time, while maintaining high accuracy for both tasks. Studies have shown that when tasks share complementary information, they can act as regularizers for each other, improving prediction performance for each task [Standley et al., 2020]. However, combining complex associations between tasks, such as classification and regression tasks, requires careful consideration of the model architecture, loss function, and training strategy to achieve optimal performance. Ongoing research is being conducted in this area [Vafaeikia et al., 2020; Vandenhende et al., 2020].

The models' ability to discover the appearance of new organs, such as ears and flag leaves, on their own is particularly interesting and opens up exciting new avenues for research. This suggests the feasibility of developing growth stage estimation models per RGB image in a similar way. Such models could be further used for various purposes, such as optimizing crop model [Yang et al., 2021a].

Overall, the performances of the models may be questioned in light of the limited dataset that only includes a few genotypes. However, the study underscores the potential of convolutional neural networks and highlights the necessity of data sharing to advance model performance and generalization capabilities. By processing diverse and large-scale data, CNNs can produce more robust and precise models. Therefore, the

phenotyping community should prioritize developing vast and well-annotated datasets for crucial phenotyping challenges, like the Global Wheat Head Detection (GWHD) dataset [David et al., 2021].

6. Conclusions

This chapter aimed to investigate the potential of convolutional neural networks (CNNs) for estimating biophysical vegetation variables in wheat from RGB and multispectral images. To overcome the need for agronomic data, a training pipeline was developed to generate pseudo-labels using transfer learning. These pseudo-labels were then corrected using the temporality of the data, greatly expanding the size of the training dataset. The performance of CNN models was enhanced by the use of these corrected pseudo-labels, outperforming the machine learning approach. In addition to estimating biophysical vegetation variables, models were also trained to estimate the dry matter and nitrogen uptake partitioning between various wheat organ pools, including the tiller, inferior leaf, flag leaf, and ear. These models demonstrated promising performance and highlighted the potential of multi-output model to estimate complex variables in plants. Overall, this chapter suggests that CNNs are a promising candidate for accurately retrieving biophysical parameters in wheat when given a sufficient dataset size. However, further data collection and investigation of models are still needed to fully unlock their capabilities.

7. Supplementary Materials

Table 6.6: R^2 of the different models to estimate DM and Nupt proportion of each organ.

Model	Data	Dataset	Nuptake				DM			
			Tiller	Linf	L1	Ear	Tiller	Linf	L1	Ear
EfficienNetB0	Ypseu	train	0.95	0.99	0.75	0.91	0.59	0.87	-0.14	0.95
EfficienNetB0	Ytrue	train	0.82	0.93	0.61	0.88	0.91	0.96	-0.02	0.93
EfficienNetB4	Ypseu	train	0.49	0.68	0.65	0.62	0.58	0.87	-0.37	0.93
EfficienNetB4	Ytrue	train	0.66	0.83	-1.38	0.55	0.93	0.98	0.5	0.95
Resnet50	Ypseu	train	0.81	0.95	0.73	0.92	0.52	0.85	0.04	0.86
Resnet50	Ytrue	train	0.97	0.99	0.84	0.99	0.98	0.99	0.71	0.98
EfficienNetB0	Ypseu	val	0.83	0.97	0.75	0.87	0.58	0.87	-0.17	0.94
EfficienNetB0	Ytrue	val	0.77	0.91	0.49	0.85	0.84	0.95	-0.31	0.92
EfficienNetB4	Ypseu	val	0.4	0.7	0.49	0.51	0.62	0.88	-0.24	0.93
EfficienNetB4	Ytrue	val	0.55	0.83	-1.8	0.44	0.87	0.96	0.2	0.94
Resnet50	Ypseu	val	0.63	0.92	0.73	0.88	0.59	0.86	0.12	0.87
Resnet50	Ytrue	val	0.78	0.96	0.66	0.9	0.91	0.98	0.46	0.96

7

Yield Estimation From Organ-Scale Traits

1. Synopsis

As agricultural systems approach the biophysical limits of crop yields, pursuing further advances in productivity becomes increasingly complex, requiring a deeper understanding of the deterministic pathways leading to improved yields [Reynolds et al., 2012]. A comprehensive understanding of yield-determining characteristics is essential, encompassing aspects such as the crop's capacity to efficiently capture and utilize available radiation, allocate dry matter across competing organs, successful generation of reproductive organs, and optimize water and nutrient absorption and utilization. Presently, drawing on insights drawn from research into yield components and mechanisms, breeders possess a more discerning awareness of which mechanisms or components to refine in order to ultimately improve yield. This informed perspective guides the selection of diverse directions that may elucidate the intricacies of yield [Reynolds et al., 2009; Slafer et al., 1996].

Within this chapter, we adopted a classical physiological approach focused on observable and countable traits (Figure 7.1). Our inquiry centers on elucidating yield and yield components through organ-scale traits [Slafer et al., 1996]. Here, yield (per m^{-2}) was expressed as:

$$Yield = Ear \times Grain\, ear^{-1} \times TKW \tag{7.1}$$

Where:

- Ear Density: Determined by the number of ears per square meter, ear density is influenced by plant density at the end of winter and the plant's ability to sustain them until the flowering phase.
- Grains per Ear: The establishment of the number of grains per ear hinges upon spikelet formation in each tiller, which subsequently generates flowers that necessitate effective fertilization during the flowering period.
- Thousand Kernel Weight (TKW): The weight of individual grains starts accumulation after the flowering period, culminating in ripening.

The concept of yield unfolds as a intricate narrative, with each yield component is susceptible to an array of stresses that ultimately impact the final yield [Khadka et al., 2020]. Consequently, strategic fertilization practices in Belgium are commonly employed at three pivotal growth stages: tillering (BBCH 28), initial stem elongation (BBCH 30), and the flag leaf (BBCH 39). The first application promotes tiller number, tthe second allows tillers to grow and develop their ear, and the third input enhances ear fertility, grain filling, and grain quality. However, an intricate trade-off exists among these components, where enhancing one could potentially lead to a decline in another [Quintero et al., 2014; Slafer et al., 2014]. Elements such as seeding rate and plant density can influence the number of ears per unit area, while genetic traits, environmental conditions, and other management practices impact grain weight

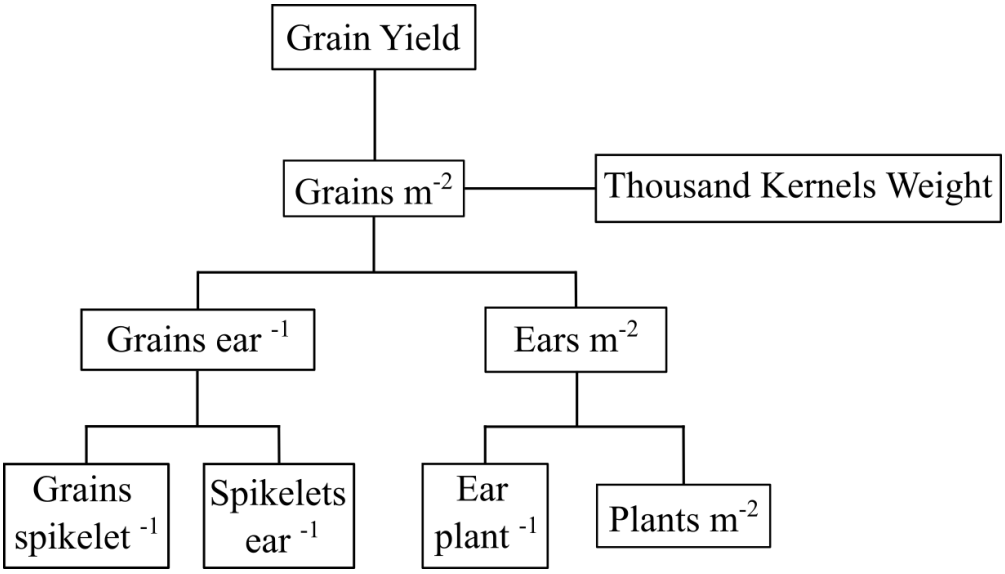

Figure 7.1: Key yield components (adapted from [Slafer et al., 1996]).

and the quantity of grains per ear. Thus, a comprehensive grasp of these yield components and their evolution throughout the crop's life cycle is imperative to formulate strategies for optimizing crop productivity and elevating wheat production.

Through this thesis and the encompassing literature, a substantial potential emerges for phenotyping systems to precisely evaluate crop traits, thereby assisting breeders and physiologists [Reynolds et al., 2020]. For instance, with a rich temporal dataset, Roth et al. [2022a] successfully construct ideotypes, representing ideal trait curves. Precise assessment of crop yield remains a challenge due to the diverse array of traits involved and the timing of their measurement [Jin et al., 2021; Prey and Schmidhalter, 2020]. In fact, the timing of data acquisition might hold equal importance to the traits themselves [Araus et al., 2022b]. While in early developmental stages, variations in vegetation indices may primarily stem from disparities in early vigor, in more advanced phenological stages characterized by denser canopies, discrepancies in vegetation indices might arise from other distinct crop characteristics. Although single NDVI measurements have demonstrated efficacy in assessing yield, this approach may fall short for larger datasets. To improve yield modeling, some researchers have suggested the use of temporal data [Sun et al., 2017], or by integrating much more other features such as weather report or crop growth models. Furthermore, one other hypothesis might be that the precision of yield assessment might improve as harvest approaches [Hassan et al., 2019].

In addition, despite the current emphasis on the evaluation of ear density (Chapter 4), little research has focused on the evaluation of the other yield components via pheno-

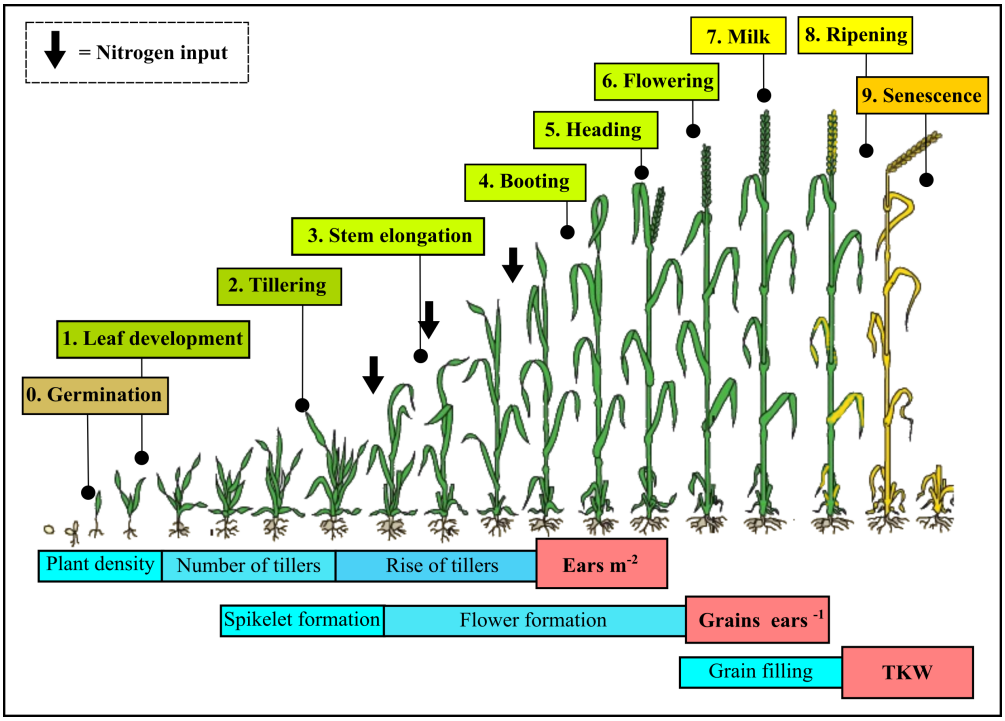

Figure 7.2: Wheat development stages and yield components. Each principal stage of the BBCH scale is indicated by a number from 0 to 9. The three yield components appear in the red boxes below the drawings. The blue boxes contain the elements influencing the elaboration of these yield components. The black down arrows represent nitrogen inputs according to a conventional fertilization strategy (from [Dandrifosse, 2022]).

typing methods. Thus, the aim of this chapter was to illustrate how the amalgamation of all extracted features from preceding chapters can offer insights into deciphering yield components. Furthermore, this also addressed a pivotal inquiry for phenotypers: that of ascertaining the most suitable criteria, timing, and methodologies for measuring these components.

2. Materials and Methods

2.1. Overview of Used Predicted Traits

The variables used in this chapter have been derived either from AI predictive models or directly computed using advanced computer vision techniques detailed in the previous chapter. A concise overview of these traits is presented in Table 7.2. Notably, all of these traits were consistently available across multiple trials, as detailed in Table 3.1, except for the 19-FP trial. Regrettably, an unexpected and severe storm to-

wards the end of the 2021 season had a detrimental impact on the micro-plots, leading to lodging for an extended duration. This weather event resulted in significant yield losses. As depicted in Figure 7.3, the micro-plots subjected to higher nitrogen inputs exhibited particularly poor yields due to increased lodging occurrences. Consequently, the data from the 2021 season has been omitted from this chapter. Furthermore, the damage removal technique employed in Chapter 5 was not applied in this context. This decision was influenced by the fact that only the 22-FP trial exhibited notable disease pressure, potentially disrupting the models presented within this chapter.

To ensure data consistency, the analysis was limited to a specific timeframe spanning from 330 to 1100 °C-Days (see Figure 3.4). This duration corresponds to the period between the second node growth stage (BBCH 32) and the early stages of maturity growth (BBCH 89). For the purpose of generating comprehensive time-series data, a cubic B-spline interpolation method with a high smoothing factor was employed. This approach facilitated the estimation of values for each individual °C-Day, resulting in a coherent and complete dataset.

Table 7.1: Output traits of this chapter.

Trait	**Unit**	**Data Source**	**Measurement Method**
Grain Yield	$kg.m^{-2}$	Combine Harvester	Direct Harvesting
Thousand Kernel Weight (TKW)	g	Sample Measurement	Weighted
Ear Density		Chapter 4	CNN: YOLOv5
Grain ear^{-1}			$\frac{\text{Grain Yield}}{\text{TKW} \times \text{Ear Density}}$

Table 7.2: Input traits used in this chapter, derived from previous computer vision methodologies.

Trait	Unit	Data Source	Methodology	Additional Information
Dry Matter	$t.ha^{-1}$	Chapter 6	EfficientNetB4	
Nitrogen Concentration	%	Chapter 6	Resnet50 CNN	Applicable to tillers, ears, flag leaves, and other leaves
Nitrogen Uptake	$kgN.ha^{-1}$	Chapter 6	Resnet50 CNN	
Leaf Area Index		Chapter 6	Resnet50 CNN	
Nitrogen Nutrition Index		Chapter 6	Equation 3.1	
Green Ratio	%	Chapter 3 Section 4.2		
Ear and Leaf Ratio	%	Chapter 3 Section 4		
Vegetation Indices		Chapter 3 Section 4.5	See Formula in Table 7.5	Relevant for images, canopies, ears, and leaves
Height (mean and percentile 95)	cm	Section 4.3		

2.2. Analysis of Yield and Yield Components: Modeling and Correlation

The relationships between each yield component (Table 7.1) and a range of predicted traits (Table 7.2) throughout the growing season were assessed using Pearson correlation coefficients. To capture the evolving correlation patterns across the cropping season, the dataset was intelligently divided into four distinct periods, delineated by accumulated temperature (°C-Days): <502, 502-665, 665-891, and >891, aligning with the key growth stages 39, 65, 75, and 89 according to the BBCH scale. The computed correlation coefficients were then averaged for each period, and these averages were categorized for clarity: "o" (<0.2), "+" (<0.4), "++" (<0.6), and "+++" (>0.6). Notably, correlations surpassing 0.6 were considered remarkably strong, while a few cases of minimum correlations exceeding 0.8 were also observed.

Furthermore, a Two-Way Analysis of Variance (ANOVA) was conducted to scrutinize the treatment effects for each trial. Remarkably, the year's influence was intentionally omitted from the analysis due to its known variability; instead, the primary focus was directed towards highlighting the pronounced treatment effects. Note that to perform these analyses, the two fungicide treatments were removed because they contained one less fertilization treatment.

To comprehensively merge all input traits, a Random Forest regressor was employed. This advanced modeling technique was harnessed to decipher the intricate interactions within the dataset. Leveraging a robust 5-fold cross-validation approach for each timestamp, the model's performance underwent evaluation, with the R^2 metric serving as the benchmark for assessment.

3. Results and discussion

3.1. Yield Analysis

The analysis of variance (ANOVA) highlights an influence of fertilization on yield, thousand kernel weight (TKW), and ear density throughout each year of the fertilization trial (Table 7.3). Notably, while the number of grains per ear demonstrates a stronger correlation with genotype, the remaining yield components exhibit a pronounced sensitivity to environmental factors, as outlined by Slafer et al. [2014]. Among these factors, nitrogen input emerges as the dominant driver of these effects [Jaenisch et al., 2022]. Here, the correlation between yield and grains per ear exhibited a relatively stable trend, ranging from 0.20 to 0.80. Notably, across six trials, four instances revealed a correlation within the range of 0.47 to 0.60.

Turning our attention to the fertilization and fungicide treatments trials, a more intricate narrative unfolds. Despite the absence of interaction terms, the dynamics proved multifaceted. The narrower range of fertilization inputs (120-260 $kgN.ha^{-1}$), com-

pared to the fertilization trial, shows a less constant fertilization effect. Actually, this effect might have been obscured by the concurrent onset of disease. Specifically, in Chapter 5, we observed the substantial susceptibility of 22-FP to yellow rust, significantly impacting all three yield components (Table 7.3). The temporal variation in disease appearance between 2021 and 2022 exerts varying effects on ear density, with delayed disease incidence in 2021 sparing this component.

Indeed, the emergence of foliar diseases before anthesis, along with their persistence during grain-filling stages, triggers a decline in the plant's green leaf area. This, in turn, hampers the plant's efficiency in capturing and utilizing radiation, thereby reducing the production of essential assimilates crucial for grain development [Schierenbeck et al., 2019a]. Such a potential discord between diminished assimilate supply and elevated demand can culminate in the abortion of kernels and ultimately lead to a reduction in overall yield [González et al., 2011]. These physiological assertions provide insight into the observed fluctuating correlations depicted in Figure 7.4. When considering the thousand kernel weight (TKW), the presence of disease appears to enhance the correlation with yield, suggesting a potential interplay between disease-induced stress and grain development [Simón et al., 2020]. Conversely, for ear density, the strongest correlation is observed within the context of the fertilization trials, highlighting the dominant influence of fertilization on this particular yield component.

Notably, the influence of violent storms in 2021 emerges as a significant variable, substantially impacting higher nitrogen inputs due to long lodging period (Figure 7.3). Indeed, yield in 2021 were expected to exhibit an upward trajectory in line with patterns from 2020 and 2022.

3.2. Individual Traits Analyses

Grain yield reveals robust correlations with a range of individual traits, as demonstrated in Table 7.4. Specifically, parameters such as DM_{plant} and $Nupt_{plant}$ exhibit a consistent enhancement in correlation, progressing from 0.5 during tillering to 0.75 during grain filling. This ascending pattern is visually depicted in Figure 7.6. Similarly, NNI and LAI exhibit parallel trends, maintaining alignment with grain yield until the end of flowering. Notably, as the season progresses, $\%N_{L1/Linf/tiller}$ demonstrates an inverse correlation of approximately -0.5, while DM_{ear} and $Nupt_{ear}$ attain a notable correlation of 0.7. Furthermore, an analogous trajectory to DM_{plant} emerges within most VIs and the leaf-to-plant ratio, as evident from Figure 7.7.

While only a handful of traits surpass a 0.4 correlation threshold with TKW, it's pertinent to highlight certain traits such as $\%N_{L1/Linf/tiller}$, $Nupt_{L1/Linf/tiller}$, $Nupt_{ear}$, and selected VIs(Table 7.4). In particular, the correlation with TKW accentuates post-flowering. However, TKW's determination primarily occurs during grain filling and pertains to grain dimensions—such as width, thickness, and surface area [Simmonds et al., 2014; Xie et al., 2015]—which are less directly captured by the camera setup.

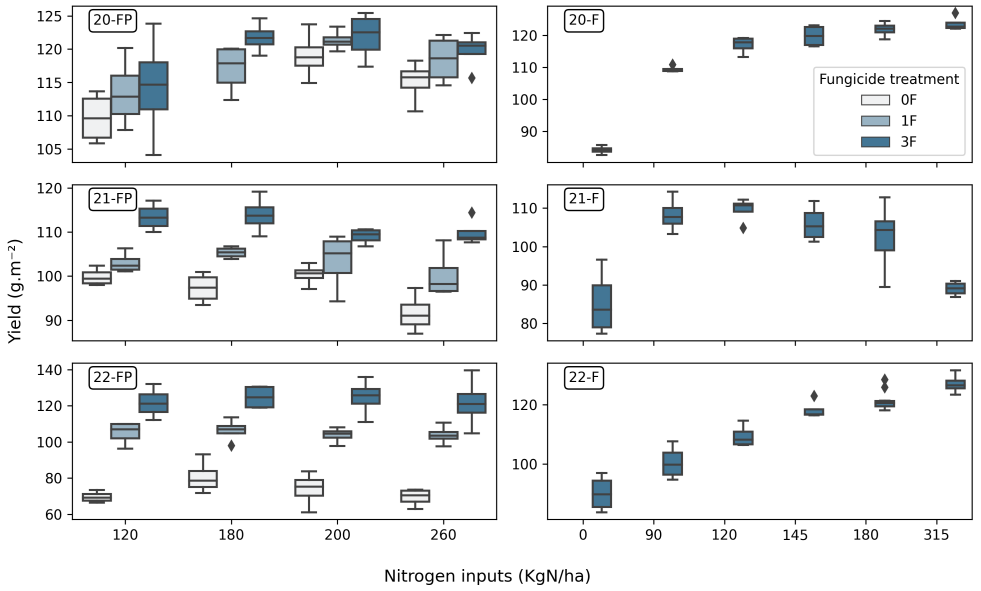

Figure 7.3: Boxplot illustrating yield variations across different trials in response to varying levels of nitrogen input. Notably, the impact of violent storms in 2021 is evident as lodging persisted for several weeks, influencing the results.

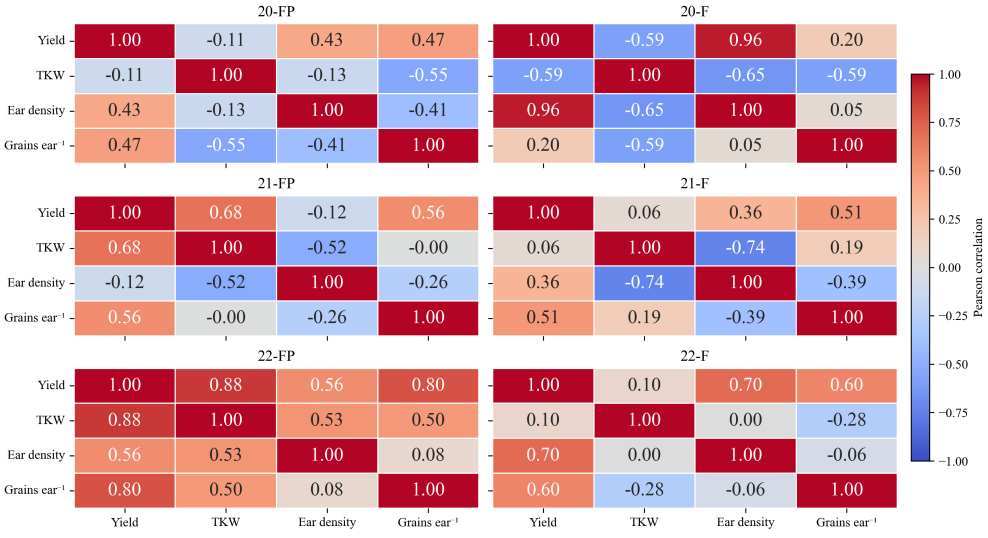

Figure 7.4: Pearson correlation between each yield component.

Table 7.3: ANOVA analysis for each trial and each yield component according to the fertilization (F) and fungicide (P) treatments and their interactions (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

Trial	Treatment	Yield	TKW	Ear density	Grains ear -1
20-FP					
	F	**		***	
	P	**			
	F x P				
21-FP					
	F	**	***	***	
	P	***	***		***
	F x P				
22-FP					
	F			*	
	P	***	***	***	***
	F x P				
20-F					
	F	***	***	***	
21-F					
	F	***	***	***	
22-F					
	F	***	**	***	

Addressing this challenge has proven complex, with a few studies attaining satisfactory outcomes [Garriga et al., 2017; Prey et al., 2020]. Thus, a deeper exploration of grain filling and ensuing senescence phases becomes imperative. Time-related metrics, like senescence duration, could potentially offer enhanced insight into these periods, enabling the identification of traits more closely related to TKW.

Pre-flowering, the number of grains per ear correlates most prominently with traits indicative of the plant's nitrogen status, encompassing NNI, %Nrate, and the plant ratio (Table 7.4). Vegetation indices also exhibit a good correlation with this variable, as they are also sensible to the presence/absence of plant. This parameter is notably sensitive to the plant's well-being during stem elongation, while also being susceptible to stresses during anthesis [González et al., 2011]. Consequently, capturing wheat phenotypes during this critical period becomes essential. Nevertheless, post-flowering, as the number of grain is already established, correlations assume greater relevance to grain yield itself, in accordance with recent findings [Jaenisch et al., 2022].

The establishment of ear density primarily occurs before the flag leaf growth stage (BBCH 39). Prominent correlations exceeding 0.6 emerge between vegetation indices associated with nitrogen, such as NDRE, GNDVI, CIREDE, and ear density (Table 7.4). Likewise, traits linked to nitrogen status exhibit correlations surpassing 0.4 pre-flag leaf growth stage, encompassing NNI, $Nupt_{plant/Linf}$, LAI, and DM_{Linf}. These findings corroborate the well-established understanding of ear density's strong reliance on nitrogen availability, a fundamental aspect of agricultural fertilization practices [Koppensteiner et al., 2022]. Beyond this stage, robust correlations emerge due to the visual distinctiveness of ears, facilitating accurate enumeration through deep learning techniques (as detailed in Chapter 4).

Table 7.4: Correlation between image traits and agronomic data by growing period (BBCH 39, 65, 75, 89). The strength of the correlation was classified using a scale where "o" indicates a correlation under 0.2, "+" indicates a correlation under 0.4, "++" indicates a correlation under 0.6, and "+++" indicates a correlation above 0.6. L1 stand for flag leaf and Linf for inferior leaves.

	Grain yield				TKW				Grain ear^{-1}				Ear density			
Image trait	< 39	< 65	< 75	< 89	< 39	< 65	< 75	< 89	< 39	< 65	< 75	< 89	< 39	< 65	< 75	< 89
NNI	++	+++	+++	++	o	o	o	+	++	++	++	+	++	++	++	++
LAI	o	++	+++	++	+	o	o	o	o	+	+	+	++	+++	+++	+
DM_{plant}	o	+++	+++	+++	o	o	o	o	o	+	++	++	+	+++	++	++
DM_{Linf}	+	+	++	++	o	o	o	o	o	o	o	+	++	+++	+++	++
DM_{L1}	o	++	++	++	o	o	o	o	o	o	+	++	o	+++	+++	++
DM_{ear}	o	+	++	+++	o	o	o	+	o	+	+	++	o	o	+	+
DM_{tiller}	o	+++	+++	++	o	o	o	o	o	++	++	+	+	++	++	++
Nup_{plant}	++	+++	+++	+++	o	o	o	+	+	++	++	+	++	++	+++	+++
$Nupt_{Linf}$	+	++	++	o	o	o	o	+	o	+	+	o	++	+++	+++	o
$Nupt_{L1}$	+	+++	+++	o	o	o	o	+	o	++	++	o	o	++	++	+
$Nupt_{ear}$	o	o	++	+++	o	o	+	++	o	+	+	++	o	o	o	+
$Nupt_{tiller}$	++	+++	++	o	o	o	o	+	+	++	+	+	+	++	+++	+
$Nrate_{plant}$	++	++	+	o	o	o	o	o	++	++	o	o	+	+	+	o
$Nrate_{Linf}$	+	++	+	+	o	o	o	+	+	++	+	+	+	o	+	o
$Nrate_{L1}$	+	+	o	+	o	o	o	+	+	+	+	+	o	o	o	o
$Nrate_{ear}$	+	o	o	+	o	o	o	o	o	o	o	o	o	o	o	o
$Nrate_{tiller}$	++	++	o	++	o	o	o	+	++	++	+	++	+	+	+	o
Leaf ratio	++	+++	+++	+++	o	o	o	o	++	++	++	+++	+	+	++	+
Ear ratio	+	+	+	++	o	o	+	o	++	++	+	++	o	o	o	+
Plant ratio	++	+++	+++	+++	o	o	o	o	++	++	++	+++	+	+	++	++
Height (average)	o	++	+++	+++	o	o	o	o	o	o	+	+	+	+++	+++	++
Height (percentile 95)	o	o	++	++	+	o	o	o	+	+	o	o	o	++	+++	++
$NDVI_{plant}$	++	+++	+++	+++	o	o	+	++	+	++	+++	+++	++	++	+	+
$GNDVI_{plant}$	++	+++	+++	+++	o	o	+	+	o	++	+++	+++	+++	++	++	++
$NDRE_{plant}$	++	+++	+++	+++	o	o	+	+	o	++	+++	++	+++	++	++	++
$MCARI_{plant}$	++	++	o	++	o	o	o	+	++	++	+	++	+	+	+	o
$CIREDE_{plant}$	++	+++	+++	+++	o	o	o	o	o	++	++	++	+++	+++	++	++
$NPCI_{plant}$	++	+++	+++	+++	o	o	+	+	+	++	+++	++	+	+	+	++

3.3. Random Forest Modeling

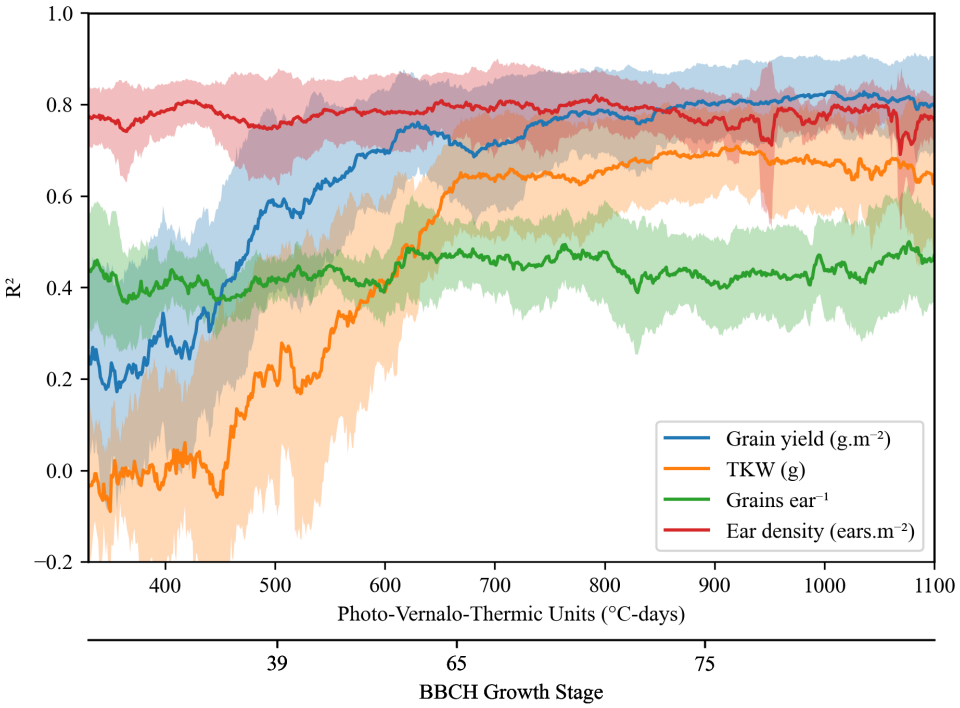

Figure 7.5: Random Forest R^2 to assess yield components in a 5 folds cross-validation. Shaded ribbons represent the standard deviation from the cross-validation.

Random Forest modeling was employed to predict yield components and yield, and the results of this analysis are presented in Figure 7.5. Initially, the R^2 of predicting grain yield and TKW was modest, but both exhibited steady improvement until plateauing around 650°C-days after ear emergence. Subsequently, the yield estimation demonstrated good performance, achieving an R^2 value of up to 0.8, while TKW yielded an R^2 value of 0.7. This pattern is consistent with plant physiology, as TKW is set post-flowering (Figure 7.2). The similarity between the trends of TKW and yield could suggest a potential high correlation between these variables. However, a closer examination of the correlation reveals that this relationship is not uniformly consistent across all trials (Figure 7.4). This indicates that the Random Forest model likely identified additional relevant features beyond this correlation.

Additionally, ear density consistently showed high and stable R^2 values ranging from 0.7 to 0.8. Conversely, the number of grains per ear also had stable results, but with a lower performance of around 0.4. This is also in line with the physiology of the plant, as both variables are established during the vegetative phase and then do not

change afterwards (Figure 7.2). The moderate performance of the Random Forest model to predict the number of grain per ear, when compared to the high correlation exhibited by individual traits, suggests that each of these traits did not provide any additional information on the number of grains per ear. Thus, only few of them might be sufficient.

These findings resonate with various studies that have successfully harnessed remote sensing techniques for yield estimation [Abbas et al., 2020; Li et al., 2022b; Wang et al., 2019]. Importantly, akin to this study, many of these investigations demonstrated low heterogeneity within their datasets, facilitating straightforward correlations. However, a comprehensive review delineating the panorama of yield prediction studies remains absent, hindering a comprehensive understanding of possibilities, challenges, and limitations. Nonetheless, the amalgamation of machine learning techniques, leveraging diverse features, emerges as a robust tool for yield estimation [Li et al., 2022a].

In light of these insights, it becomes apparent that some traits pose greater complexity for estimation through phenotyping systems. This study opted for a statistical modeling approach based on phenotyping features, resulting in dataset-specific models that yielded intriguing results. It's important to acknowledge that this approach's goals and use might differ from mechanistic approaches. The selected approach may prove effective for generating precise data on a local area, whereas methods such as crop growth model could excel at broader scales. While few recent studies compare phenotyping and crop growth modeling specifically for yield estimation, numerous successful endeavors have effectively fused remote sensing and crop growth models [Della Nave et al., 2022; Yang et al., 2021b; Yue et al., 2021].

4. Conclusions

This study delved into an intricate analysis of the correlations between various yield components and a diverse array of image-derived traits, detailed in the preceding chapter through computer vision methodologies. The correlations were analyzed across different growth periods, revealing shifting relationships over time. Each yield component showed strong correlations with multiple traits, ranging from hard-to-obtain traits, such as NNI, to more readily available traits, such as various vegetation indices.

To comprehensively integrate these complex interactions, a Random Forest regressor was employed for modeling. The R^2 values obtained from this analysis illustrated the predictive power of image-based traits in estimating yield components. Notably, the model exhibited increasing accuracy as the season progressed, reflecting the temporal dynamics of yield determination. These statement are also in accordance with the physiological knowledge. Ear density and yield were particularly well-predicted by the model, suggesting that image-derived traits effectively capture the underlying dynamics of these components. However, the prediction of the number of grains per

ear showed a modest performance, indicating that additional factors beyond the traits considered in the analysis may contribute to its determination.

Overall, this chapter provided valuable insights into the complex relationships between image-based traits and key yield components in wheat. It showcased the potential of advanced computer vision techniques and statistical modeling to enhance our understanding of crop physiology and improve yield estimation. The findings accentuate the need for a judicious choice of traits and a holistic consideration of their temporal dynamics across the agricultural season. As such, these innovative phenotyping systems stand as a potent tool in aiding crop improvement.

5. Supplementary Materials

Table 7.5: Description of the 15 vegetation indices and the normalized RGB components used as features for the machine learning algorithms aiming to separate leaves and ears. r, g and b are the channels of the RGB camera. blue, green, red, rededge and nir are the channels of the multi-spectral camera array, respectively at 490, 550, 680, 720 and 800 nm.

Index	Index formula	Reference
NDVI	$\frac{nir-red}{nir+red}$	[Rouse et al., 1974]
GNDVI	$\frac{nir-green}{nir+geen}$	[Gitelson et al., 1996]
NDRE	$\frac{nir-rededge}{nir+rededge}$	[Barnes et al., 2000]
MCARI	$((rededge-red)-0.2\times(rededge-green))\times \frac{rededge}{red}$	[Daughtry et al., 2000]
CIrede	$\frac{nir}{rededge}-1$	[Roujean and Breon, 1995]
NPCI	$\frac{red-blue}{red+blue}$	[Devadas et al., 2009]

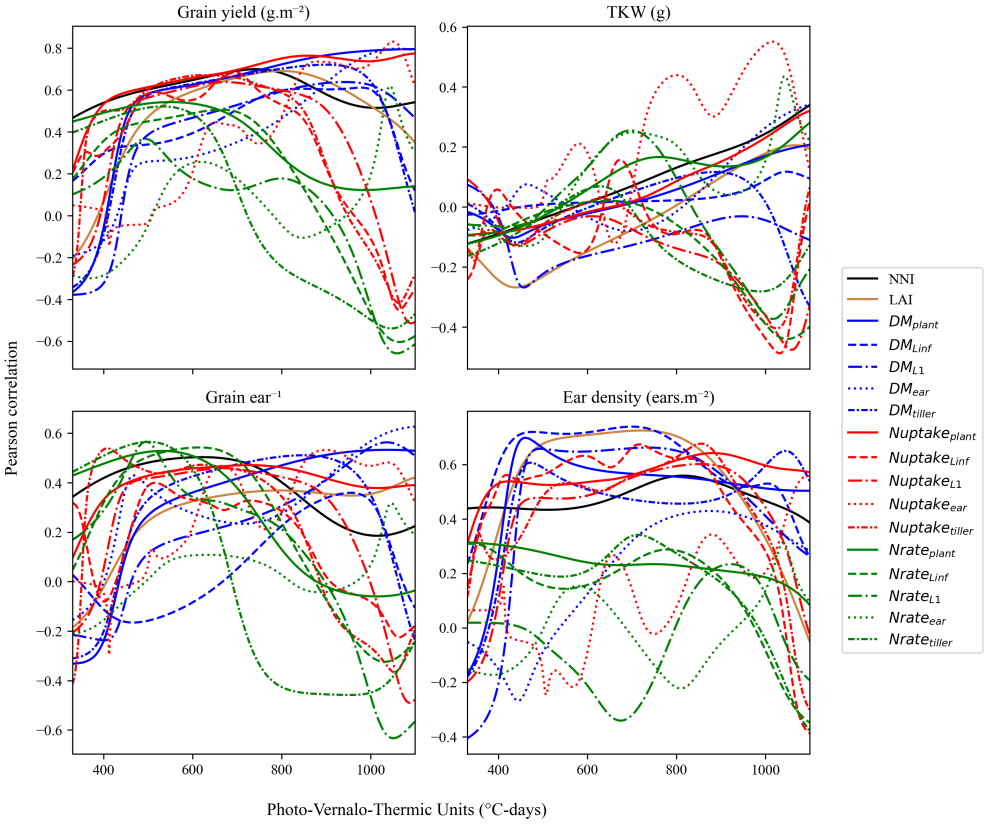

Figure 7.6: Pearson correlation between yield components and modeled traits.

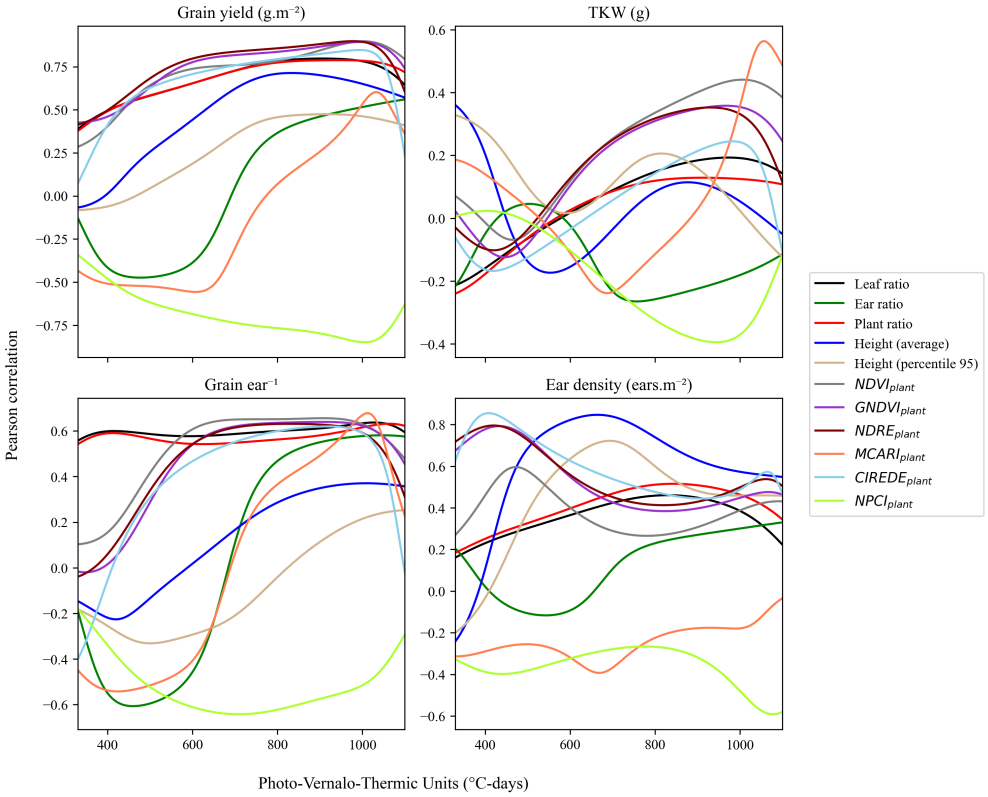

Figure 7.7: Pearson correlation between yield components and images features.

8

General Discussion

1. An Emerging Science With Great Potential

The field of plant phenotyping is rapidly emerging as a dynamic and innovative scientific frontier, as evidenced by its exponential growth [Morisse et al., 2022; Sun et al., 2022]. A remarkable testament to this evolution can be found in a study by Saint Cast et al. [2022], revealing an astounding 2600% surge in published papers within the plant phenotyping domain between 2015 and 2019. This unprecedented trajectory is largely propelled by technological advancements, amplified computational capacities, and notably, the advancement of AI [Arya et al., 2022]. However, despite these remarkable strides, the field remains in its infancy as a scientific discipline, still grappling with diverse and formidable challenges tied to extracting agronomic insights from sensor data collected directly in field.

This thesis demonstrates the efficacy of an affordable phenotyping system in capturing a diverse array of data on wheat, encompassing crucial metrics such as plant height, biomass, and disease impact. Each of these traits, either individually or gather, contributes valuable insights toward understanding grain yield. The advent of high-throughput phenotyping systems is enabling a transformation in crop improvement by facilitating rapid and objective assessments, and by shedding light on new traits [Araus et al., 2022b; Reynolds et al., 2020]. While final yield remains the pivotal metric evaluated, the capacity to measure an array of traits throughout the crop cycle has expanded significantly. In essence, remote and proximal sensing offer a repertoire of methods and tools/devices, ranging from vegetation indices, BRFs, stereovision, and AI to RGB, multi and hyperspectral cameras, as well as thermal sensors, all aimed at retrieving these intricate traits. As illustrated in this thesis, the amalgamation of RGB and multispectral images, when combined with AI models, yields valuable insights into traits such as biomass, nitrogen content, LAI, plant height, and the impact of diseases.

Moreover, this thesis presents a temporal analysis of estimated traits and their correlation with yield components, which are typically not directly measurable in the field except for ear density. These findings provide insights into the optimal timing for observing specific traits throughout the growth cycle. However, the quest for more extensive data persists, necessary to draw sturdier conclusions from this approach and to conduct meta-analyses, ultimately leading to a deeper comprehension of the intricate relationships among these traits, thereby enhancing our understanding of crop yield dynamics [Roth et al., 2022b; van Eeuwijk et al., 2019].

This reservoir of data serves as a guiding compass for breeders, aiding in the selection of optimal plant varieties for specific growth conditions, thereby culminating in amplified yields and enhanced crop quality. Furthermore, these insights facilitate the identification of ideotypes, unraveling temporal patterns and extracting time-linked traits, in addition to genes/QTL [Mir et al., 2019]. With continued research in this area, we can further harness the power of high-throughput phenotyping to improve

crop yields and ensure food security for future generations.

As technology continues to progress, the decreasing costs of components render phenotyping systems increasingly accessible to a broader audience. The system outlined in this thesis exhibits relative affordability in terms of hardware, with the camera representing the most substantial expense. Specifically, while multispectral cameras often cost around €10,000, an RGB-D camera can be acquired for €350. Additionally, the integration of deep learning has evolved into an indispensable facet of all phenotyping systems, empowering researchers to overcome once insurmountable challenges. A salient instance of this progress is the automated counting of wheat heads, facilitated by object detection algorithms and a publicly available extensive dataset [David et al., 2021]. Notwithstanding these strides, a disparity persists between phenotyping systems in research and their widespread utilization in large-scale field trials. Bridging this gap is imperative to fully realize the potential of phenotyping systems in addressing agricultural challenges.

2. Present Challenges

2.1. Technical Obstacles

Over the course of several years of development, we encountered numerous challenges that shed light on the slow uptake of this approach among users. Our experience highlighted one prominent hurdle: technical challenges arising from the use of custom-made hardware and software. These challenges inevitably led to setbacks. Across a four-year span, we faced multiple obstacles that complicated the implementation of this approach. Yet, we consider these challenges integral to the research process and will present the limitations we encountered as examples of the complexities associated with this approach. Despite the difficulties we encountered, we remained resolute in our pursuit, gaining valuable insights and lessons along the way.

As noted by Dandrifosse [2022], the use of a multimodal set-up increased the complexity of the system, making it more susceptible to issues. This was true for both the mobile platform and the sensor set-up. As the platform and electronic system had to be reassembled for each acquisition, there was a risk that some cables might be forgotten in the laboratory, screws might not be tightened correctly, or devices could be damaged. Given the large number of cables required to connect all the sensors, managing the system and guarantee its robustness were undoubtedly challenging.

One other issue to manage was that the cameras used in our system required specific calibrations and settings. We found that the stereovision cameras needed to be recalibrated periodically, possibly due to imperfections in the metal support. The relative exposure time of the multispectral camera had to be adjusted several times per year to prevent over- or under-saturated pixels, depending on the crop growth stage. Unfortunately, these cameras saved the images onto SD cards, making it challenging to verify

them in the field. These SD cards also added multiple steps to the image unloading process and increased the burden on external storage. However, we did not encounter any issues related to data storage. The key was to stay organized and follow a basic folder structure on each external disk, given that each acquisition date generated several gigabytes of data.

The homemade acquisition system we implemented came with its fair share of bugs. Frequently, situations arose where perfect lab performance didn't translate to the field. This was more common initially, but it provided valuable lessons on project management. One of the most challenging aspects was synchronizing all the sensors, each of which had its own Python Software Development Kits (SDK) working differently with varying levels of detail.

Obtaining high-quality data is essential for effective phenotyping and model building. Addressing this challenge consumed a significant amount of time, particularly in Sebastien Dandrifosse's thesis. The image analysis pipeline also underwent significant changes throughout our research. A large Python script incorporating most methods was written to process the images, with each pipeline step often relying on the output of a preceding step. Insufficient verification phases, however, occasionally resulted in bugs and code crashes, particularly concerning given the code's computationally intensive nature.

Field conditions present a more complex environment than laboratories, where the plant itself primarily drives changes. Weather is a major factor that can disrupt data acquisition and affect data quality. Wind can impede the use of sensors that require a significant amount of time to acquire data, such as hyperspectral imaging systems [Bebronne et al., 2020]. Rain can also hinder data acquisition as many technologies are not designed to withstand water, and can make the field less accessible. Finally, the sun, in combination with clouds, can be a significant challenge. Numerous studies have demonstrated the effects of sunlight on data quality, particularly in image data, where high illumination conditions enhance contrast and could hide the bottom of the canopy [Dandrifosse et al., 2022b; Serouart et al., 2022]. Solutions exist to mitigate these issues, such as using a flash or concealing the area, but these solutions can also introduce biases into the data and increase the complexity of the system. External factors such as soil color, soil heterogeneity, or the presence of weeds can affect plant measurements. To mitigate external factors, we implemented various methods such as soil segmentation, but achieving consistent results still requires additional knowledge, especially when working outside controlled environments.

Many studies have successfully identified valuable plant traits but there are still many other traits that require further investigation. For example, although there have been notable improvements in identifying diseases in crops [Tanner et al., 2022], predicting diseases before symptoms appear, differentiating between diseases with similar symptoms, and detecting diseases hidden within a canopy, such as STB in wheat, still

present significant challenges. Thermal imaging technology is another promising tool for detecting water stress, but no operational method currently allows for large-scale application. Yet, it is certain that these challenges would greatly benefit from a wealth of data.

The challenges we faced during this work, particularly in regards to image acquisition and analysis, underscore the fact that implementing a phenotyping system is a complex task and not easy to handle by novices. Currently, there are very few user-friendly commercial solutions available, despite the proliferation of commercial cameras and UAVs. While these technologies are becoming more affordable and accessible, there remains a significant knowledge barrier to their effective use. A solution that can instantly capture an image and produce a list of traits is not yet available, and users must possess a high level of technical expertise to operate these commercial cameras and UAVs from image to knowledge. Furthermore, the diverse needs of different end-users mean that there is no one-size-fits-all image analysis pipeline, and each problem may require a bespoke solution. Technical challenges represent one major but not the only barrier to the adoption of phenotyping systems.

2.2. Matching Needs and Utilization of Phenotyping Systems

Plant phenotyping systems cater to a diverse range of end users with distinct needs and purposes. Nevertheless, all these users play a pivotal role in enhancing crop quality. Despite the potential advantages that phenotyping systems offer to plant breeding, their adoption remains in its early stages. Current research in this domain often stems from available technologies rather than being driven by breeders' specific demands, even though breeders are the ultimate intended users of these systems.

The potential of phenotyping techniques is intriguing, and the rapid production of massive amounts of imaging and remote-sensing data is impressive. However, the complexity involved in integrating this data with genotypic and environmental information poses the risk of lingering unresolved. For instance, trying to predict grain yield by correlating a single vegetation index with final yield might not be suitable across all regions due to the multifaceted nature of yield as a trait. A simplistic approach of this kind wouldn't satisfy breeders, given that yield is a multifarious trait that can't be reliably predicted from a single measurement.

Breeders' requirements primarily revolve around yield and technical crop parameters, such as disease resistance, rather than precise factors like plant height or nitrogen content in the context of wheat. Moreover, breeders might have needs that current technology cannot meet, like determining kernel weights under field conditions. Additionally, considering the crop type and the stage of the breeding program, the breeders' needs can vary in terms of measurement throughput and trait priorities.

On contrary, agronomists or physiologists might necessitate more specific measurements to provide advice to farmers, evaluate input efficacy, or address research queries.

Thus, when designing phenotyping systems, it's imperative to factor in the end users' needs to ensure successful adoption and effective utilization in crop enhancement.

Modern high-throughput phenotyping systems can now measure numerous previously unmeasured or underexplored traits, such as color data, canopy coverage, and temperature. While these traits can serve as supplementary variables or be correlated with established variables, their role as predictive elements raises questions about their agronomic significance. The debate over whether end users or technologies should adapt to each other forms part of a wider societal discourse on embracing new agricultural technologies. The optimal approach likely lies in a combination of strategies, each contributing an added advantage.

2.3. Data: The Core of Phenotyping

To design reliable and meaningful models, the quality of the data itself stands as the primary factor to reckon with. The concept of reproducibility within the field of phenotyping might still be relatively unexplored [Kedron and Frazier, 2022], a situation that extends to deep learning models and the training phase as well [Maxwell et al., 2022]. Field conditions inherently vary, and the repeatability or heritability of genetically intricate traits can differ due to interactions with the environment. Thus, it's imperative to either measure or control these environmental factors as meticulously as possible, to maximize heritability as traits advance through the selection process.

The expression of most traits also interacts with the growth stage. This is evident when assessing phenology or conducting growth analysis, for instance. Yet, for other traits where changes might not be easily visible, this interaction can be less apparent. Sensors can also introduce inaccuracies. When collecting data using certain sensors, particularly spectroscopy sensors, external inputs like reflectance calibration panels are necessary to ensure precision. Sensor parameters such as exposure time and aperture need consideration to avoid image saturation and blurriness. Devices like UAVs require external tools like georeferenced targets to capture consistently located images, along with an understanding of flying parameters. Thankfully, hardware manufacturers are incorporating these features to automate data acquisition to the fullest extent. However, it's worth noting that high-quality data remains vital for specific phenotyping tasks, such as wheat ear counting or canopy cover estimation. It's been demonstrated that high throughput can offset lower data quality [Lane and Murray, 2021]. Thus, the trade-off between high throughput and high-resolution data depends on the task or requirement.

Developing truly "user-friendly" phenotyping systems remains a challenge, with data handling arguably being a significant bottleneck. At the core of phenomics data management lies the pivotal process of translating raw sensor data into biologically pertinent traits. While there are various data processing options available for numerous applications, data processing still demands familiarity with data science and coding.

Moreover, modeling a singular target trait could entail several approaches, leaving users grappling with choosing the best one. While certain user-friendly tools like PlantCV software [Fahlgren et al., 2015] have surfaced to streamline image analysis pipelines, many applications still necessitate specific requirements that must be included in the code. Consequently, crafting user-friendly phenotyping systems capable of accommodating diverse personalized tasks remains a challenge.

Agronomic reference measurements like biomass, LAI, yield, and plant height are frequently assessed, modeled, or predicted through various methodologies. Statistical approaches, the prevalent choice [Berger et al., 2020; Verrelst et al., 2019], establish correlations between these variables and features extracted from images. While these methods offer adaptability in connecting target values with phenotyping data, their broader applicability is limited since they often hinge on relationships created for the specific task. Their performance in a different context could significantly diverge.

Although acquiring all these target measurements is crucial, it's labor-intensive, demanding, and can be costly to measure, as seen with nitrogen uptake that require both biomass samples and laboratory analyses. Additionally, models are invariably tested on a subset of the dataset to evaluate their performance. Selecting the appropriate subset can be intricate; it must be independent yet within the dataset's variability to avoid overfitting. Nevertheless, over the course of this thesis, it has become apparent that constructing and segmenting such datasets can be tricky, still they retain their fundamental importance for evaluating a model's efficacy. Therefore, the incorporation of image features like Vegetation Index (VI) as alternative variables in breeding pipelines could potentially surmount the constraints imposed by manual measurements.

While phenotyping systems generate substantial amounts of phenotypic data, the concept of big data introduces challenges and considerations to be tackled. Big data necessitates considerable storage and computational resources, and the infrastructure to support these needs can be expensive and challenging to manage. Managing and storing extensive data can also raise concerns about data privacy and security.

2.4. Enhancing Collaboration through Standardization

The significance of open and shared data in addressing phenotyping challenges cannot be emphasized enough. However, the proliferation of distinct data storage systems can complicate data sharing. To facilitate collaboration and data exchange within the scientific community, a standardized approach to data management and storage is imperative. This strategy could help surmount the challenges posed by big data and facilitate the culture of open and shared data.

Standardization and harmonization of measurement protocols are critical for the exchange of this data. Phenomic experiments are susceptible to environmental variations, making replication a challenge [Kedron and Frazier, 2022]. Thus, it is essential to ensure that any scientist, including those in the future, can reuse phenotypic data

and replicate data workflows for large-scale analyses. This necessity has given rise to the development of norms known as FAIR (findable, accessible, interoperable, and reusable) principles, which concentrate on tracing data, protocols, methods, and workflows [Wilkinson et al., 2016]. The meticulous collection of metadata holds significance, encompassing basic data linked to field attributes, management practices, crop phenology, data acquisition dates, and setup parameters. In this regard, initiatives like the Minimum Information About a Plant Phenotyping Experiment (MIAPPE) were introduced to offer a checklist for accurately describing phenotyping experiments. However, there is still a considerable effort ahead to establish a unified language within the phenotyping community, facilitating seamless information exchange and extending its impact into domains such as crop modeling [Saint Cast et al., 2022]. This harmonization would undoubtedly yield mutual benefits, enhancing the compatibility of both disciplines and allowing them to mutually refine and amplify each other's efficacy.

These protocols are equally essential for the reference measurements linked to phenotyping data. While many of these measurements are inherently standardized, such as area-based variables like biomass or yield, recent open databases associated with deep learning, like the Global Wheat Head Dataset 2 (GWHD2), demand consistent annotations such as bounding boxes and segmentation for constructing accurate models [David et al., 2021]. In fact, to ensure annotation consistency, the dataset authors have opted to outsource the annotation process rather than allowing individual contributors to annotate their own images.

Lastly, the community could greatly benefit from a form of quality assurance. This would assure relevance to end-users as well as competent authorities responsible for certifying agricultural products [Morisse et al., 2022]. Such authorities could recommend these tools in the certification process, elevating these tools to highly recommended status.

3. Conclusions and Perspectives

Phenotyping systems have become increasingly popular among researchers owing to remarkable technological advancements. While this field has witnessed substantial growth within the research community, its broader adoption is still in its infancy. Presently, technology is predominantly embraced by early adopters on the Rogers curve [Rogers, 1995] (Figure 8.1). These early adopters are visionary, recognizing the project's trajectory and embracing an evolving product. In 2023, we consider that some few companies have already adopted such technologies in routine. However, for wider dissemination, the technology must traverse the Chasm, a critical juncture between early adopters and the more pragmatic early majority. Crossing this Chasm is pivotal for engaging the rest of the end-users.

This thesis offers insights into the potential and limitations of phenotyping systems,

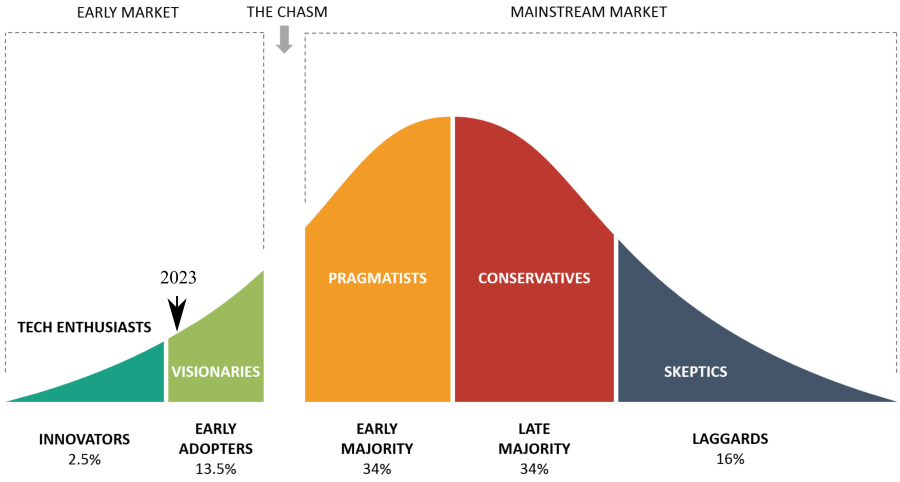

Figure 8.1: Rogers Curve illustrating the estimated positioning of phenotyping technologies in 2023.

encapsulated in the SWOT analysis (Table 8.1). Contemporary measurement techniques expand the horizon for rapid, precise, and comprehensive data collection for diverse objectives, even unveiling novel and significant traits. The horizon of phenotyping systems is promising, particularly with the burgeoning influence of AI. The research community is currently working on gathering large datasets that will benefit all phenotyping users and help build increasingly accurate models. Nevertheless, data collection, especially under field conditions, confronts numerous challenges, with weather proving a pivotal variable. Furthermore, the diverse array of sensors and analysis pipelines demands expertise that might frighten some end-users. To alleviate this, standardization and quality assurance mechanisms, whether services or user-driven initiatives, could inspire confidence and allow users to fully exploit the benefits of phenotyping systems. These measures could even gain traction in the field of agricultural input approvals by policymakers, potentially positioning certain phenotyping systems as recommended tools for this function. Consequently, phenotyping system adoption could significantly reduce the laborious task of annotations and help overcome agricultural challenges more quickly.

While the sector is in its developmental phase, its organizational framework remains nascent. The query arises whether breeders should construct their own phenotyping systems and competencies or rely on a burgeoning market to provide systems and services. If breeders opt for in-house development, it may disproportionately favor larger corporations endowed with ample resources, potentially marginalizing smaller entities clinging to traditional approaches. Yet, these industry giants can also serve as

beacons, expanding phenotyping systems' reach beyond the Chasm and democratizing access across the sector. Ultimately, the choice between internal system development and outsourcing will hinge on individual breeder circumstances, constituting a long-term goal.

Table 8.1: SWOT analysis on the adoption of phenotyping solutions.

Strengths	**Weaknesses**
Technological Innovations: Advanced sensors, high-resolution imaging, and AI-driven analytics. **Abundance of traits**: High-throughput, objective and non-destructive measure. **New traits opportunities**: New traits are available such as vegetation indices.	**Data Complexity**: Managing and interpreting large datasets require specialized skills. **Multifacets**: Affected by multiple factors such as the weather. **Sensor and Limitations**: Sensors can't capture all traits. **Dependency on reference data**: Over-reliance on agronomic reference data. **Cost and Accessibility**: Expensive technology limits accessibility for smaller organizations.
Opportunities	**Threats**
AI Integration and improvement: Automation for data processing and analytics. **Data Collaboration**: Standardized protocols for enhanced cross-laboratory collaboration. **Sensing Technology**: Continued improvement of sensors. **Agricultural Challenges**: Addressing global agricultural issues. **Cost Reduction**: Lowering expenses for wider adoption. **Growing Community**: Expanding phenotyping expertise.	**Resistance to Change**: Overcoming reluctance to adopt new technologies. **Matching Needs**: Driven by technologies instead of needs. **Data Privacy and Security**: Ensuring data privacy and protection. **Environmental Impact**: Considering the ecological implications of technology. **Regulatory Challenges**: Potential regulatory challenges related to safety, environment, and intellectual property. **Skills Gap**: Disparities in technical skills may widen.

9

General Conclusions and Future Perspectives

1. Summarizing Key Conclusions

The field of high-throughput plant phenotyping emerges as a potent tool, holding the promise to support crop enhancement and address the intricate challenges in agriculture to sustain global food security in a sustainable manner. However, the use of these new technologies to assess crop traits within field conditions is laden with complexities, given that the plant is one factor in a broader context. The focal aim of this thesis was to bridge the chasm between raw data, predominantly images, and agronomic knowledge through the utilization of a functional mobile platform. Employing diverse image analysis methodologies, this work has showcased the capacities in extracting a spectrum of wheat traits across the growing season. Moreover, it underscores the feasibility of leveraging these traits collectively to explore their potential in explaining grain yield and its components.

A sensor pod equipped with two RGB cameras and a multispectral camera, placed on the mobile platforms, served several purposes. Primarily, the stereovision technique harnessed the twin RGB cameras to capture height data. Additionally, a deep learning algorithm displayed a F1-score of 0.91 for ear segmentation through YOLOv5 and DeepMac models, outperforming a superpixel-based machine learning approach. Furthermore, the SegVeg method exhibited remarkable success in segmenting the wheat canopy into soil, green, and damaged pixels. Capitalizing on ResNet50 and EfficientNet models, the RGB images demonstrated remarkable promise in quantifying wheat biomass, distributing dry matter, and assessing LAI. These deep learning models outperformed a conventional PLSr methodology.

Turning to the multispectral camera, a pivotal preliminary step entailed normalizing the values through bi-directional reflectance factor calculations to accommodate varying illumination conditions. Particularly noteworthy, the 800 nm image proved helpful in discerning soil from crops. By integrating diverse canopy masks coming from the previous segmentation, an array of vegetation indices were computed at the organ scale. Our investigation revealed that diseases not only perturb the vegetation indices of the damaged parts but also impact those of unaffected green elements, potentially undermining the reliability of reflectance-based decision-making tools when diseases are present. Additionally, the multispectral images exhibited substantial promise in assessing nitrogen-related variables, spanning uptake and concentration across plants and individual organs. In this regard, the ResNet CNN model outperformed a traditional PLSr method as well.

Each image underwent meticulous registration, a pivotal step facilitating the overlay of images and masks, a prerequisite for subsequent organ-scale studies. In the culminating phases, all predicted traits were validated as robust indicators in explicating yield development and its constituents. Notable instances include the Random Forest model, yielding an enhanced grain yield estimation post-flowering with an R^2 of 0.8, alongside the Thousand Kernel Weight model that yielded around 0.7 R^2. Consistent

performances marked ear density, grains per ear count, and grain nitrogen concentration across the growth stages, showing R^2 values of approximately 0.8, 0.4, and 0.4 respectively. Furthermore, the potency of several vegetation indices in correlating with yield components underscores their efficacy as a resource-efficient tool for this pursuit.

While this study elucidated the potential of proximal field trial phenotyping in extracting detailed phenotypic insights, the real-world uptake of such systems remains limited amongst end-users. A major hindrance is the huge array of sensors, platforms, and analytical approaches available, often leaving users perplexed in their quest for the most suitable fit. Moreover, some technologies are operationally challenging and prohibitively expensive, further dampening adoption rates. A lack of benchmark dataset, coupled with a lack of protocol and data standardization, compounds the challenge of data exchange, process streamlining, and model generalization. Such standardization would not only facilitate cross-study comparisons but also bolster our comprehension of plant traits and their interplay with yield.

Propelled by strides in technology and artificial intelligence, the research community is getting ready to offer more user-friendly and economically viable tools accessible to a broader spectrum of users. As these tools become increasingly accessible, the latent benefits of field trial phenotyping can be harnessed by a diverse cohort of stakeholders, including plant breeders, farmers, and researchers. Ultimately, increased adoption of phenotyping systems could potentially accelerate the development of new crop varieties and amplify agricultural productivity in a sustainable way globally.

2. Perspectives and Future Work

The landscape of addressing agricultural challenges is tackled by the complementary roles of a multitude of actors. Scientific inquiry drives the innovation of novel solutions, while breeders and agronomists must align with societal needs to surmount agricultural hurdles. A dynamic interplay exists, with researchers benefiting from laboratory flexibility to ideate and refine solutions, yet culminating in a need for robust standardization to serve end-users. This transition necessitates a symbiotic alliance among researchers, breeders, and agronomists to ensure lab-derived solutions harmonize with practicality, efficacy, and the needs of the farming community.

Through the course of building an affordable multimodal camera prototype, that promises faster and more efficient data acquisition, our exploration of homemade equipment has also underscored challenges requiring technical expertise. In the pursuit of improving the system, two paths are available to us. Primarily, an elevated iteration of the current platform, characterized by comprehensive automation, promises accelerated acquisition and free up time for other tasks. Alternatively, the development of a portable system that allows for simpler sensor integration could also be explored. These two pathways could be intended to technicians at the Gembloux Agro-Bio Tech

Experimental Farm, thus aligning developments with the practical requisites and constraints of end-users.

In the course of this work, the power of Convolutional Neural Network (CNN) models has manifested remarkably, charting a course for further exploration. The decade ahead brims with extraordinary prospects. Hence, I advocate accentuating CNN's trajectory to refine numerous methodologies elucidated in this work, encompassing:

- Pioneering the integration of CNN to amplify disparity maps within the stereovision process, while delving into monocular depth perception.
- Leveraging recent strides in image registration to accelerate and improve the registration procedure.
- Advancing and expanding upon the Unet model introduced in Chapter 5 for canopy segmentation.
- Exploring multi-task models as an avenue to curtail computational demands and potentially enhance model performance for individual traits.

The future holds a tapestry of promising avenues for exploration. Delving into high dynamic range techniques that yield diffuse RGB images regardless of sunlight conditions bears immense potential. Another area of investigation concerns the analysis of the influence of data acquisition conditions, such as lighting on the repeatability and reproducibility of measurements. Moreover, a multi-scale strategy employing Unmanned Aerial Vehicles (UAVs) and/or satellites could amplify the efficiency of data collection and analysis across expansive landscapes. Moreover, even though our study focused on wheat, the adaptive nature of the system can extend to various crops. Various image analysis methodologies can be directly transferrable, with some requiring adaptation. Thus, further research can be directed towards optimizing the system for other crops and identifying new traits that can be measured within the system.

Notwithstanding the considerable strides made in phenotyping, a range of challenges still shadows the community. While certain challenges are actively engaged, others persist with enigmatic resolutions. Of these, the potential of thermal imagery remains underexplored, holding promise for discerning biotic and abiotic stress but braked by the complexities of environmental influences. Disease detection, especially for in-canopy and early-stage identification, remains a formidable enigma, compounded by the need to differentiate diverse stress factors. Another crucial frontier is root phenotyping, at the crossroads of importance and complexity. However, the pursuit of accurate and comprehensive root phenotyping data encounters formidable challenges.

In summation, the field of phenotyping calls for sustained dedication, with recent breakthroughs in machine learning, computer vision, and image processing opening up tantalizing avenues. The amalgamation of these innovations with the fervor of breeders, the tenacity of researchers, and the entrepreneurial spirit of farmers holds the promise of transforming the agroecosystem. As we navigate this synergy and leverage novel tools and techniques, a more sustainable and resilient future for agriculture

and food production awaits, underpinned by collaborative endeavor and technological prowess.

List of Achievements

Published Papers as First Author

Carlier, A., Dandrifosse, S., Dumont, B. and Mercatoris, B., 2023. Comparing CNNs and traditional machine learning for estimating wheat organs biophysical variables using proximal sensing. Frontiers in Plant Science 2023. (Under review)

A. Carlier, S. Dandrifosse, B. Dumont, and B. Mercatoris, 'To what extent do leaf disease pressures affect remotely sensed nitrogen status? The case study of yellow rust', Plant Phenomics, vol. 0, no. ja, Aug. 2023, doi: 10.34133/plantphenomics.0083.

A. Carlier, S. Dandrifosse, B. Dumont, and B. Mercatoris, 'Wheat Ear Segmentation Based on a Multisensor System and Superpixel Classification', Plant Phenomics, vol. 2022, Jan. 2022, doi: 10.34133/2022/9841985.

Romain Bebronne, Alexis Carlier, Remi Meurs, Vincent Leemans, Philippe Vermeulen, Benjamin Dumont, Benoit Mercatoris, 'In-field proximal sensing of septoria tritici blotch, stripe rust and brown rust in winter wheat by means of reflectance and textural features from multispectral imagery', Biosystems Engineering, vol. 197, 2020, doi: 10.1016/j.biosystemseng.2020.06.011. (Sharing first author)

Other Published Papers

S. Dandrifosse, E. Ennadifi, A. Carlier, B. Gosselin, B. Dumont, and B. Mercatoris, 'Deep learning for wheat ear segmentation and ear density measurement: From heading to maturity', Computers and Electronics in Agriculture, vol. 199, p. 107161, Aug. 2022, doi: 10.1016/j.compag.2022.107161.

S. Dandrifosse, A. Carlier, B. Dumont, and B. Mercatoris, 'In-Field Wheat Reflectance: How to Reach the Organ Scale?', Sensors, vol. 22, no. 9, Art. no. 9, Jan. 2022, doi: 10.3390/s22093342.

S. Dandrifosse, A. Carlier, B. Dumont, and B. Mercatoris, 'Registration and Fusion of Close-Range Multimodal Wheat Images in Field Conditions', Remote Sensing, vol. 13, no. 7, p. 1380, Apr. 2021, doi: 10.3390/rs13071380.

E. David et al., 'Global Wheat Head Detection 2021: an improved dataset for benchmarking wheat head detection methods', Plant Phenomics, vol. 2021, 2021, doi:

10.34133/2021/9846158.

Scientific Oral Presentations and Posters

A. Carlier, S. Dandrifosse, B. Dumont, and B. Mercatoris, 'Organ scale nitrogen map: a novel approach for leaf nitrogen concentration estimation' Oral presentation, ICPA 2022, June 2022

A. Carlier, S. Dandrifosse, B. Dumont, and B. Mercatoris, 'Proximal sensing for wheat field trials phenotyping', Oral presentation, Erasmus + KA2 Smart-Farming closing conference, Oct 2020

A. Carlier, S. Dandrifosse, B. Dumont, and B. Mercatoris,'Wheat nitrogen and senescence dynamics in field assessment through two phenotyping approaches late in season' Oral presentation, XVI European Society for Agronomy Congress, Sept 2020

A. Carlier, S. Dandrifosse, and B. Mercatoris 'PhenWheat: A field phenotyping platform for wheat trials' Poster, 25th National Symposium for Applied Biological Sciences, Jan 2020

Other Communications

A. Carlier, S. Dandrifosse, A. Bouvry, E Bustillo Vazquez, B. Dumont, and B. Mercatoris,'Le phénotypage numérique en champs : un outil de caractérisation des cultures' article du Livre Blanc céréales, Feb 2020

A. Carlier and S. Dandrifosse, 'Presentation de la phenomobile lors d'un événement organisé par walagri' June 2019

Bibliography

Abbas, F., Afzaal, H., Farooque, A. A., and Tang, S. (2020). Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms. *Agronomy* 10, 1046. doi:10.3390/agronomy10071046

Affouard, A., Goëau, H., Bonnet, P., Lombardo, J.-C., and Joly, A. (2017). Pl@ntNet app in the era of deep learning. In *ICLR: International Conference on Learning Representations*

Al-Tamimi, N., Langan, P., Bernád, V., Walsh, J., Mangina, E., and Negrão, S. (2022). Capturing crop adaptation to abiotic stress using image-based technologies. *Open Biology* 12, 210353. doi:10.1098/rsob.210353

Alves, K. S., Guimarães, M., Ascari, J. P., Queiroz, M. F., Alfenas, R. F., Mizubuti, E. S. G., et al. (2022). RGB-based phenotyping of foliar disease severity under controlled conditions. *Tropical Plant Pathology* 47, 105–117. doi:10.1007/s40858-021-00448-y

Anderegg, J., Hund, A., Karisto, P., and Mikaberidze, A. (2019). In-Field Detection and Quantification of Septoria Tritici Blotch in Diverse Wheat Germplasm Using Spectral–Temporal Features. *Frontiers in Plant Science* 10

Anderegg, J., Yu, K., Aasen, H., Walter, A., Liebisch, F., and Hund, A. (2020). Spectral Vegetation Indices to Track Senescence Dynamics in Diverse Wheat Germplasm. *Frontiers in Plant Science* 10. doi:10.3389/fpls.2019.01749

Andrade, S. M., Teodoro, L. P. R., Baio, F. H. R., Campos, C. N. S., Roque, C. G., da Silva Júnior, C. A., et al. (2021). High-throughput phenotyping of soybean genotypes under base saturation stress conditions. *Journal of Agronomy and Crop Science* 207, 814–822. doi:10.1111/jac.12513

Araus, J. L., Buchaillot, M. L., and Kefauver, S. C. (2022a). High Throughput Field Phenotyping. In *Wheat Improvement*, eds. M. P. Reynolds and H.-J. Braun (Cham: Springer International Publishing). 495–512. doi:10.1007/978-3-030-90673-3_27

Araus, J. L. and Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science* 19, 52–61. doi:10.1016/j.tplants.2013.09.008

Araus, J. L., Kefauver, S. C., Vergara-Díaz, O., Gracia-Romero, A., Rezzouk, F. Z., Segarra, J., et al. (2022b). Crop phenotyping in a context of global change: What to measure and how to do it. *Journal of Integrative Plant Biology* 64, 592–618. doi:10.1111/jipb.13191

Arya, S., Sandhu, K. S., Singh, J., and kumar, S. (2022). Deep learning: As the new frontier in high-throughput plant phenotyping. *Euphytica* 218, 47. doi:10.1007/s10681-022-02992-3

Arzt, M., Deschamps, J., Schmied, C., Pietzsch, T., Schmidt, D., Tomancak, P., et al. (2022). LABKIT: Labeling and Segmentation Toolkit for Big Image Data. *Frontiers in Computer Science* 4

Ashourloo, D., Mobasheri, M. R., and Huete, A. (2014). Evaluating the Effect of Different Wheat Rust Disease Symptoms on Vegetation Indices Using Hyperspectral Measurements. *Remote Sensing* 6, 5107–5123. doi:10.3390/rs6065107

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing* 5, 949–981. doi:10.3390/rs5020949

Bancal, M.-O., Roche, R., and Bancal, P. (2008). Late Foliar Diseases in Wheat Crops Decrease Nitrogen Yield Through N Uptake Rather than Through Variations in N Remobilization. *Annals of Botany* 102, 579–590. doi:10.1093/aob/mcn124

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., et al. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. In *Proceedings of the Fifth International Conference on Precision Agriculture* (Bloomington, Madison, USA)

Bebronne, R., Carlier, A., Meurs, R., Leemans, V., Vermeulen, P., Dumont, B., et al. (2020). In-field proximal sensing of septoria tritici blotch, stripe rust and brown rust in winter wheat by means of reflectance and textural features from multispectral imagery. *Biosystems Engineering* 197. doi:10.1016/j.biosystemseng.2020.06.011

Berger, K., Machwitz, M., Kycko, M., Kefauver, S. C., Van Wittenberghe, S., Gerhards, M., et al. (2022). Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: A review. *Remote Sensing of Environment* 280, 113198. doi:10.1016/j.rse.2022.113198

Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Wocher, M., Strathmann, M., et al. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment* 242, 111758. doi:10.1016/j.rse.2020.111758

[Dataset] Birodkar, V., Lu, Z., Li, S., Rathod, V., and Huang, J. (2021). The surprising impact of mask-head architecture on novel class segmentation. doi:10.48550/arXiv.2104.00613

Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*

Brocks, S. and Bareth, G. (2018). Estimating Barley Biomass with Crop Surface Models from Oblique RGB Imagery. *Remote Sensing* 10, 268. doi:10.3390/rs10020268

Buxbaum, N., Lieth, J. H., and Earles, M. (2022). Non-destructive Plant Biomass Monitoring With High Spatio-Temporal Resolution via Proximal RGB-D Imagery and End-to-End Deep Learning. *Frontiers in Plant Science* 13

Cammarano, D., Fitzgerald, G. J., Casa, R., and Basso, B. (2014). Assessing the Robustness of Vegetation Indices to Estimate Wheat N in Mediterranean Environments. *Remote Sensing* 6, 2827–2844. doi:10.3390/rs6042827

Cao, X., Liu, Y., Yu, R., Han, D., and Su, B. (2021). A Comparison of UAV RGB and Multispectral Imaging in Phenotyping for Stay Green of Wheat Population. *Remote Sensing* 13. doi:10.3390/rs13245173

Carlier, A., Dandrifosse, S., Dumont, B., and Mercatoris, B. (2022). Wheat Ear Segmentation Based on a Multisensor System and Superpixel Classification. *Plant Phenomics* 2022. doi:10.34133/2022/9841985

Carlier, A., Dandrifosse, S., Dumont, B., and Mercatoris, B. (2023). To what extent do leaf disease pressures affect remotely sensed nitrogen status? The case study of yellow rust. *Plant Phenomics* 0. doi:10.34133/plantphenomics.0083

Castro, W., Marcato Junior, J., Polidoro, C., Osco, L. P., Gonçalves, W., Rodrigues, L., et al. (2020). Deep Learning Applied to Phenotyping of Biomass in Forages with UAV-Based RGB Imagery. *Sensors* 20, 4802. doi:10.3390/s20174802

Cavaco, A., Utkin, A., da Silva, J., and Guerra, R. (2022). Making Sense of Light: The Use of Optical Spectroscopy Techniques in Plant Sciences and Agriculture. *Applied Sciences (Switzerland)* 12. doi:10.3390/app12030997

Chakraborty, D., Sehgal, V. K., Sahoo, R. N., Pradhan, S., and Gupta, V. K. (2015). Study of the Anisotropic Reflectance Behaviour of Wheat Canopy to Evaluate the Performance of Radiative Transfer Model PROSAIL5B. *Journal of the Indian Society of Remote Sensing* 43, 297–310. doi:10.1007/s12524-014-0411-7

Chao, Z., Liu, N., Zhang, P., Ying, T., and Song, K. (2019). Estimation methods developing with remote sensing information for energy crop biomass: A comparative review. *Biomass and Bioenergy* 122, 414–425. doi:10.1016/j.biombioe.2019.02.002

Cheshkova, A. F. (2022). A review of hyperspectral image analysis techniques for plant disease detection and identif ication. *Vavilovskii Zhurnal Genetiki I Selektsii* 26, 202–213. doi:10.18699/VJGB-22-25

Chowhan, P. and Chakraborty, A. (2022). Remote Sensing Technology—A New Dimension in Detection, Quantification and Tracking of Abiotic and Biotic Stresses. *Advances in Science, Technology and Innovation* , 445–457doi:10.1007/978-3-030-95365-2_27

Cointault, F., Guerin, D., Guillemin, J.-P., and Chopinet, B. (2008). In-field Triticum aestivum ear counting using colour-texture image analysis. *New Zealand Journal of Crop and Horticultural Science* 36, 117. doi:10.1080/01140670809510227

Comar, A., Baret, F., Viénot, F., Yan, L., and de Solan, B. (2012). Wheat leaf bidirectional reflectance measurements: Description and quantification of the volume, specular and hot-spot scattering features. *Remote Sensing of Environment* 121, 26–35. doi:10.1016/j.rse.2011.01.028

[Dataset] CRA-W (2023). Yield Statistics Belgium. https://bcgms.be/fr/

Dandrifosse, S. (2022). *Dynamics of wheat organs by close-range multimodal machine vision*. Ph.D. thesis, ULiège. GxABT - Liège Université. Gembloux Agro-Bio Tech, Gembloux, Belgium

Dandrifosse, S., Bouvry, A., Leemans, V., Dumont, B., and Mercatoris, B. (2020). Imaging Wheat Canopy Through Stereo Vision: Overcoming the Challenges of the Laboratory to Field Transition for Morphological Features Extraction. *Frontiers in Plant Science* 11, 96. doi:10.3389/fpls.2020.00096

Dandrifosse, S., Carlier, A., Dumont, B., and Mercatoris, B. (2021). Registration and Fusion of Close-Range Multimodal Wheat Images in Field Conditions. *Remote Sensing* 13, 1380. doi:10.3390/rs13071380

Dandrifosse, S., Carlier, A., Dumont, B., and Mercatoris, B. (2022a). In-Field Wheat Reflectance: How to Reach the Organ Scale? *Sensors* 22, 3342. doi:10.3390/s22093342

Dandrifosse, S., Ennadifi, E., Carlier, A., Gosselin, B., Dumont, B., and Mercatoris, B. (2022b). Deep learning for wheat ear segmentation and ear density measurement: From heading to maturity. *Computers and Electronics in Agriculture* 199, 107161. doi:10.1016/j.compag.2022.107161

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., and McMurtrey, J. E. (2000). Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sensing of Environment* 74, 229–239. doi: 10.1016/S0034-4257(00)00113-9

David, E., Serouart, M., Smith, D., Madec, S., Velumani, K., Liu, S., et al. (2021). Global Wheat Head Detection 2021: An improved dataset for benchmarking wheat head detection methods. *Plant Phenomics* 2021. doi:10.34133/2021/9846158

de Oliveira, G. S., Marcato Junior, J., Polidoro, C., Osco, L. P., Henrique Siqueira, Rodrigues, L., et al. (2021). Convolutional Neural Networks to Estimate Dry Matter Yield in a Guineagrass Breeding Program Using UAV Remote Sensing. *Sensors* 21, 3971. doi:10.3390/s21123971

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., and Furbank, R. (2014). Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping. *Agronomy* 4, 349–379. doi:10.3390/agronomy4030349

Deery, D. M. and Jones, H. G. (2021). Field Phenomics: Will It Enable Crop Improvement? *Plant Phenomics* 2021. doi:10.34133/2021/9871989

Della Nave, F. N., Ojeda, J. J., Irisarri, J. G. N., Pembleton, K., Oyarzabal, M., and Oesterheld, M. (2022). Calibrating APSIM for forage sorghum using remote sensing and field data under sub-optimal growth conditions. *Agricultural Systems* 201, 103459. doi:10.1016/j.agsy.2022.103459

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. doi:10.1109/CVPR.2009.5206848

Devadas, R., Lamb, D. W., Backhouse, D., and Simpfendorfer, S. (2015). Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat. *Precision Agriculture* 16, 477–491. doi: 10.1007/s11119-015-9390-0

Devadas, R., Lamb, D. W., Simpfendorfer, S., and Backhouse, D. (2009). Evaluating ten spectral vegetation indices for identifying rust infection in individual wheat leaves. *Precision Agriculture* 10, 459–470. doi:10.1007/s11119-008-9100-2

Duchene, O., Dumont, B., Cattani, D. J., Fagnant, L., Schlautman, B., DeHaan, L. R., et al. (2021). Process-based analysis of Thinopyrum intermedium phenological development highlights the importance of dual induction for reproductive growth and agronomic performance. *Agricultural and Forest Meteorology* 301–302, 108341. doi:10.1016/j.agrformet.2021.108341

Erenstein, O., Jaleta, M., Mottaleb, K. A., Sonder, K., Donovan, J., and Braun, H.-J. (2022). Global Trends in Wheat Production, Consumption and Trade. In *Wheat Improvement*, eds. M. P. Reynolds and H.-J. Braun (Cham: Springer International Publishing). 47–66. doi:10.1007/978-3-030-90673-3_4

Fahlgren, N., Feldman, M., Gehan, M. A., Wilson, M. S., Shyu, C., Bryant, D. W., et al. (2015). A Versatile Phenotyping System and Analytics Platform Reveals Diverse Temporal Responses to Water Availability in Setaria. *Molecular Plant* 8, 1520–1535. doi:10.1016/j.molp.2015.06.005

Fan, X., Zhou, R., Tjahjadi, T., Das Choudhury, S., and Ye, Q. (2022). A Segmentation-Guided Deep Learning Framework for Leaf Counting. *Frontiers in Plant Science* 13

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

Fernandez-Gallego, J. A., Buchaillot, M. L., Aparicio Gutiérrez, N., Nieto-Taladriz, M. T., Araus, J. L., and Kefauver, S. C. (2019). Automatic Wheat Ear Counting Using Thermal Imagery. *Remote Sensing* 11, 751. doi:10.3390/rs11070751

Fernandez-Gallego, J. A., Kefauver, S. C., Gutiérrez, N. A., Nieto-Taladriz, M. T., and Araus, J. L. (2018). Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. *Plant Methods* 14, 22. doi:10.1186/s13007-018-0289-4

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342. doi:10.1038/nature10452

Food and Agriculture Organization of the United Nations (ed.) (2017). *The Future of Food and Agriculture: Trends and Challenges* (Rome: Food and Agriculture Organization of the United Nations)

Foulkes, M. J., Molero, G., Griffiths, S., Slafer, G. A., and Reynolds, M. P. (2022). Yield Potential. In *Wheat Improvement*, eds. M. P. Reynolds and H.-J. Braun (Cham: Springer International Publishing). 379–396. doi:10.1007/978-3-030-90673-3_21

Furbank, R. T. and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science* 16, 635–644. doi:10.1016/j.tplants.2011.09.005

Gaju, O., Allard, V., Martre, P., Le Gouis, J., Moreau, D., Bogard, M., et al. (2014). Nitrogen partitioning and remobilization in relation to leaf senescence, grain yield

and grain nitrogen concentration in wheat cultivars. *Field Crops Research* 155, 213–223. doi:10.1016/j.fcr.2013.09.003

Garcin, C., Joly, A., Bonnet, P., Affouard, A., Lombardo, J.-C., Chouet, M., et al. (2021). Pl@ntNet-300K image dataset doi:10.5281/zenodo.5645731

Garriga, M., Romero-Bravo, S., Estrada, F., Escobar, A., Matus, I. A., del Pozo, A., et al. (2017). Assessing Wheat Traits by Spectral Reflectance: Do We Really Need to Focus on Predicted Trait-Values or Directly Identify the Elite Genotypes Group? *Frontiers in Plant Science* 8. doi:10.3389/fpls.2017.00280

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* 58, 289–298. doi:10.1016/S0034-4257(96)00072-7

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment* 80, 76–87. doi:10.1016/S0034-4257(01)00289-9

Gitelson, A. A., Keydan, G. P., and Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters* 33. doi:10.1029/2006GL026457

González, F. G., Miralles, D. J., and Slafer, G. A. (2011). Wheat floret survival as related to pre-anthesis spike growth. *Journal of Experimental Botany* 62, 4889–4901. doi:10.1093/jxb/err182

Gonzalez-Dugo, V., Hernandez, P., Solis, I., and Zarco-Tejada, P. J. (2015). Using High-Resolution Hyperspectral and Thermal Airborne Imagery to Assess Physiological Condition in the Context of Wheat Phenotyping. *Remote Sensing* 7, 13586–13605. doi:10.3390/rs71013586

Görlich, F., Marks, E., Mahlein, A.-K., König, K., Lottes, P., and Stachniss, C. (2021). Uav-based classification of cercospora leaf spot using rgb images. *Drones* 5. doi:10.3390/drones5020034

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., et al. (2021). UAS-Based Plant Phenotyping for Research and Breeding Applications. *Plant Phenomics* 2021, 1–21. doi:10.34133/2021/9840192

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment* 81, 416–426. doi:10.1016/S0034-4257(02)00018-4

Hamuda, E., Mc Ginley, B., Glavin, M., and Jones, E. (2017). Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Computers and Electronics in Agriculture* 133, 97–107. doi: 10.1016/j.compag.2016.11.021

Hassan, M. A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., et al. (2019). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Science* 282, 95–103. doi:10.1016/j.plantsci.2018.10.022

Hawkesford, M. and Riche, A. (2020). Impacts of G x E x M on Nitrogen Use Efficiency in Wheat and Future Prospects. *Frontiers in Plant Science* 11. doi: 10.3389/fpls.2020.01157

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep Residual Learning for Image Recognition doi:10.48550/arXiv.1512.03385

Henke, M., Junker, A., Neumann, K., Altmann, T., and Gladilin, E. (2019). Comparison and extension of three methods for automated registration of multimodal plant images. *Plant Methods* 15, 44. doi:10.1186/s13007-019-0426-8

Henrich, V., Götze, C., Jung, A., Sandow, C., Thürkow, D., and Cornelia, G. (2009). *Development of an Online Indices Database: Motivation, Concept and Implementation*

Hickey, L. T., N. Hafeez, A., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nature Biotechnology* 37, 744–754. doi:10.1038/s41587-019-0152-9

[Dataset] Iakubovskii, P. (2019). Segmentation Models. https://github.com/qubvel/segmentation_models

Jaenisch, B. R., Munaro, L. B., Jagadish, S. V. K., and Lollato, R. P. (2022). Modulation of Wheat Yield Components in Response to Management Intensification to Reduce Yield Gaps. *Frontiers in Plant Science* 13, 772232. doi:10.3389/fpls.2022.772232

Jay, S., Gorretta, N., Morel, J., Maupas, F., Bendoula, R., Rabatel, G., et al. (2017). Estimating leaf chlorophyll content in sugar beet canopies using millimeter- to centimeter-scale reflectance imagery. *Remote Sensing of Environment* 198, 173–186. doi:10.1016/j.rse.2017.06.008

Jay, S., Rabatel, G., and Gorretta, N. (2014). In-field crop row stereo-reconstruction for plant phenotyping. In *Second International Conference on Robotics and Associated High-Technologies and Equipment for Agriculture and Forestry (RHEA-2014)* (Madrid, Spain), 10 p.

Jiang, J., Atkinson, P. M., Zhang, J., Lu, R., Zhou, Y., Cao, Q., et al. (2022). Combining fixed-wing UAV multispectral imagery and machine learning to diagnose winter wheat nitrogen status at the farm scale. *European Journal of Agronomy* 138, 126537. doi:10.1016/j.eja.2022.126537

Jiang, Y. and Changying Li (2020). Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. *Plant Phenomics* 2020, 1–22. doi:10.34133/2020/4152816

Jiang, Y., Li, C., Robertson, J. S., Sun, S., Xu, R., and Paterson, A. H. (2018). GPhenoVision: A Ground Mobile System with Multi-modal Imaging for Field-Based High Throughput Phenotyping of Cotton. *Scientific Reports* 8, 1213. doi:10.1038/s41598-018-19142-2

Jin, X., Zarco-Tejada, P. J., Schmidhalter, U., Reynolds, M. P., Hawkesford, M. J., Varshney, R. K., et al. (2021). High-Throughput Estimation of Crop Traits: A Review of Ground and Aerial Phenotyping Platforms. *IEEE Geoscience and Remote Sensing Magazine* 9, 200–231. doi:10.1109/MGRS.2020.2998816

Justes, E. (1994). Determination of a Critical Nitrogen Dilution Curve for Winter Wheat Crops. *Annals of Botany* 74, 397–407. doi:10.1006/anbo.1994.1133

Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture* 147, 70–90. doi:10.1016/j.compag.2018.02.016

Kattenborn, T., Leitloff, J., Schiefer, F., and Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 173, 24–49. doi:10.1016/j.isprsjprs.2020.12.010

Kedron, P. and Frazier, A. E. (2022). How to Improve the Reproducibility, Replicability, and Extensibility of Remote Sensing Research. *Remote Sensing* 14, 5471. doi:10.3390/rs14215471

Khadka, K., Earl, H., Raizada, M., and Navabi, A. (2020). A Physio-Morphological Trait-Based Approach for Breeding Drought Tolerant Wheat. *Frontiers in Plant Science* 11. doi:10.3389/fpls.2020.00715

Khanna, R., Schmid, L., Walter, A., Nieto, J., Siegwart, R., and Liebisch, F. (2019). A spatio temporal spectral framework for plant stress phenotyping. *Plant Methods* 15, 13. doi:10.1186/s13007-019-0398-8

Koc, A., Odilbekov, F., Alamrani, M., Henriksson, T., and Chawade, A. (2022). Predicting yellow rust in wheat breeding trials by proximal phenotyping and machine learning. *Plant Methods* 18, 30. doi:10.1186/s13007-022-00868-0

Koppensteiner, L. J., Kaul, H.-P., Piepho, H.-P., Barta, N., Euteneuer, P., Bernas, J., et al. (2022). Yield and yield components of facultative wheat are affected by sowing time, nitrogen fertilization and environment. *European Journal of Agronomy* 140, 126591. doi:10.1016/j.eja.2022.126591

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60, 84–90. doi:10.1145/3065386

Kuska, M., Heim, R., Geedicke, I., Gold, K., Brugger, A., and Paulus, S. (2022). Digital plant pathology: A foundation and guide to modern agriculture. *Journal of Plant Diseases and Protection* 129, 457–468. doi:10.1007/s41348-022-00600-z

Lancashire, P. D., Bleiholder, H., Boom, T. V. D., Langelüddeke, P., Stauss, R., Weber, E., et al. (1991). A uniform decimal code for growth stages of crops and weeds. *Annals of Applied Biology* 119, 561–601. doi:10.1111/j.1744-7348.1991.tb04895.x

Lane, H. M. and Murray, S. C. (2021). High throughput can produce better decisions than high accuracy when phenotyping plant populations. *Crop Science* 61, 3301–3313. doi:10.1002/csc2.20514

Lee, D.-H. (2013). Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*

Leemans, V., Dumont, B., and Destain, M.-F. (2013). Assessment of plant leaf area measurement by using stereo-vision. In *2013 International Conference on 3D Imaging*. 1–5. doi:10.1109/IC3D.2013.6732085

Lemaire, G. and Ciampitti, I. (2020). Crop Mass and N Status as Prerequisite Covariables for Unraveling Nitrogen Use Efficiency across Genotype-by-Environment-by-Management Scenarios: A Review. *Plants* 9, 1309. doi:10.3390/plants9101309

Li, L., Wang, B., Feng, P., Li Liu, D., He, Q., Zhang, Y., et al. (2022a). Developing machine learning models with multi-source environmental data to predict wheat yield in China. *Computers and Electronics in Agriculture* 194, 106790. doi:10.1016/j.compag.2022.106790

Li, Q., Jin, S., Zang, J., Wang, X., Sun, Z., Li, Z., et al. (2022b). Deciphering the contributions of spectral and structural data to wheat yield estimation from proximal sensing. *The Crop Journal* 10, 1334–1345. doi:10.1016/j.cj.2022.06.005

Li, W., Jiang, J., Weiss, M., Madec, S., Tison, F., Philippe, B., et al. (2021a). Impact of the reproductive organs on crop BRDF as observed from a UAV. *Remote Sensing of Environment* 259, 112433. doi:10.1016/j.rse.2021.112433

Li, W., Niu, Z., Chen, H., Li, D., Wu, M., and Zhao, W. (2016). Remote estimation of canopy height and aboveground biomass of maize using high-resolution stereo images from a low-cost unmanned aerial vehicle system. *Ecological Indicators* 67, 637–648. doi:10.1016/j.ecolind.2016.03.036

Li, Y., Liu, H., Ma, J., and Zhang, L. (2021b). Estimation of leaf area index for winter wheat at early stages based on convolutional neural networks. *Computers and Electronics in Agriculture* 190, 106480. doi:10.1016/j.compag.2021.106480

Li, Y., Zhan, X., Liu, S., Lu, H., Jiang, R., Guo, W., et al. (2023). Self-Supervised Plant Phenotyping by Combining Domain Adaptation with 3D Plant Model Simulations: Application to Wheat Leaf Counting at Seedling Stage. *Plant Phenomics* 5, 0041. doi:10.34133/plantphenomics.0041

Lin, Y. (2015). LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics? *Computers and Electronics in Agriculture* 119, 61–73. doi:10.1016/j.compag.2015.10.011

Liu, S., Baret, F., Abichou, M., Manceau, L., Andrieu, B., Weiss, M., et al. (2021). Importance of the description of light interception in crop growth models. *Plant Physiology* 186, 977–997. doi:10.1093/plphys/kiab113

Lobet, G., Draye, X., and Périlleux, C. (2013). An online database for plant image analysis software tools. *Plant Methods* 9, 38. doi:10.1186/1746-4811-9-38

Ma, J., Li, Y., Chen, Y., Du, K., Zheng, F., Zhang, L., et al. (2019). Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *European Journal of Agronomy* 103, 117–129. doi:10.1016/j.eja.2018.12.004

Ma, J., Li, Y., Liu, H., Du, K., Zheng, F., Wu, Y., et al. (2020). Improving segmentation accuracy for ears of winter wheat at flowering stage by semantic segmentation. *Computers and Electronics in Agriculture* 176, 105662. doi:10.1016/j.compag.2020.105662

Machwitz, M., Pieruschka, R., Berger, K., Schlerf, M., Aasen, H., Fahrner, S., et al. (2021). Bridging the Gap Between Remote Sensing and Plant Phenotyping—Challenges and Opportunities for the Next Generation of Sustainable Agriculture. *Frontiers in Plant Science* 12, 749374. doi:10.3389/fpls.2021.749374

Madec, S., Irfan, K., Velumani, K., Baret, F., David, E., Daubige, G., et al. (2023). VegAnn, Vegetation Annotation of multi-crop RGB images acquired under diverse conditions for segmentation. *Scientific Data* 10, 302. doi:10.1038/s41597-023-02098-y

Mahlein, A.-K. (2016). Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease* 100, 241–251. doi:10.1094/PDIS-03-15-0340-FE

Martre, P., Porter, J. R., Jamieson, P. D., and Triboï, E. (2003). Modeling Grain Nitrogen Accumulation and Protein Composition to Understand the Sink/Source Regulations of Nitrogen Remobilization for Wheat. *Plant Physiology* 133, 1959–1967. doi:10.1104/pp.103.030585

Maxwell, A. E., Bester, M. S., and Ramezan, C. A. (2022). Enhancing Reproducibility and Replicability in Remote Sensing Deep Learning Research and Practice. *Remote Sensing* 14, 5760. doi:10.3390/rs14225760

Maydup, M. L., Antonietta, M., Guiamet, J. J., Graciano, C., López, J. R., and Tambussi, E. A. (2010). The contribution of ear photosynthesis to grain filling in bread wheat (Triticum aestivum L.). *Field Crops Research* 119, 48–58. doi:10.1016/j.fcr.2010.06.014

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., and Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum* 106, 135–141. doi:10.1034/j.1399-3054.1999.106119.x

Meyer, G. E. and Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture* 63, 282–293. doi:10.1016/j.compag.2008.03.009

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science* 282, 60–72. doi:10.1016/j.plantsci.2019.01.007

Morisse, M., Wells, D. M., Millet, E. J., Lillemo, M., Fahrner, S., Cellini, F., et al. (2022). A European perspective on opportunities and demands for field-based crop phenotyping. *Field Crops Research* 276, 108371. doi:10.1016/j.fcr.2021.108371

Mortensen, A. K., Dyrmann, M., Karstoft, H., Jørgensen, R. N., and Gislum, R. (2016). Semantic Segmentation of Mixed Crops using Deep Convolutional Neural Network

Müller-Linow, M., Pinto-Espinosa, F., Scharr, H., and Rascher, U. (2015). The leaf angle distribution of natural plant populations: Assessing the canopy with a novel software tool. *Plant Methods* 11, 11. doi:10.1186/s13007-015-0052-z

Mustafa, G., Zheng, H., Khan, I., Tian, L., Jia, H., Li, G., et al. (2022). Hyperspectral Reflectance Proxies to Diagnose In-Field Fusarium Head Blight in Wheat with Machine Learning. *Remote Sensing* 14. doi:10.3390/rs14122784

Nabwire, S., Suh, H.-K., Kim, M. S., Baek, I., and Cho, B.-K. (2021). Review: Application of Artificial Intelligence in Phenomics. *Sensors* 21, 4363. doi: 10.3390/s21134363

Nielsen, D. C., Miceli-Garcia, J. J., and Lyon, D. J. (2012). Canopy Cover and Leaf Area Index Relationships for Wheat, Triticale, and Corn. *Agronomy Journal* 104, 1569–1573. doi:10.2134/agronj2012.0107n

Niu, Y., Zhang, L., Zhang, H., Han, W., and Peng, X. (2019). Estimating above-ground biomass of maize using features derived from UAV-based RGB imagery. *Remote Sensing* 11. doi:10.3390/rs11111261

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9, 62–66. doi:10.1109/TSMC.1979.4310076

Pandey, P., Irulappan, V., Bagavathiannan, M. V., and Senthil-Kumar, M. (2017). Impact of Combined Abiotic and Biotic Stresses on Plant Growth and Avenues for Crop Improvement by Exploiting Physio-morphological Traits. *Frontiers in Plant Science* 8

Pauli, D., Chapman, S. C., Bart, R., Topp, C. N., Lawrence-Dill, C. J., Poland, J., et al. (2016). The Quest for Understanding Phenotypic Variation via Integrated Approaches in the Field Environment. *Plant Physiology* 172, 622–634

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830

Peng, J., Manevski, K., Kørup, K., Larsen, R., and Andersen, M. N. (2021). Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. *Field Crops Research* 268, 108158. doi:10.1016/j.fcr.2021.108158

Peng, Y., Dallas, M. M., Ascencio-Ibáñez, J. T., Hoyer, J. S., Legg, J., Hanley-Bowdoin, L., et al. (2022). Early detection of plant virus infection using multispectral imaging and spatial–spectral machine learning. *Scientific Reports* 12, 3113. doi:10.1038/s41598-022-06372-8

Pineda, M., Barón, M., and Pérez-Bueno, M. (2020). Thermal Imaging for Plant Stress Detection and Phenotyping. *Remote Sensing* 13, 68. doi:10.3390/rs13010068

Prey, L., Hu, Y., and Schmidhalter, U. (2020). High-Throughput Field Phenotyping Traits of Grain Yield Formation and Nitrogen Use Efficiency: Optimizing the Selection of Vegetation Indices and Growth Stages. *Frontiers in Plant Science* 10

Prey, L. and Schmidhalter, U. (2019a). Sensitivity of Vegetation Indices for Estimating Vegetative N Status in Winter Wheat. *Sensors* 19, 3712. doi:10.3390/s19173712

Prey, L. and Schmidhalter, U. (2019b). Temporal and Spectral Optimization of Vegetation Indices for Estimating Grain Nitrogen Uptake and Late-Seasonal Nitrogen Traits in Wheat. *Sensors* 19, 4640. doi:10.3390/s19214640

Prey, L. and Schmidhalter, U. (2020). Deep Phenotyping of Yield-Related Traits in Wheat. *Agronomy* 10, 603. doi:10.3390/agronomy10040603

Prey, L., von Bloh, M., and Schmidhalter, U. (2018). Evaluating RGB imaging and multispectral active and hyperspectral passive sensing for assessing early plant vigor in winter wheat. *Sensors (Switzerland)* 18. doi:10.3390/s18092931

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment* 48, 119–126. doi:10.1016/0034-4257(94)90134-1

Quintero, A., Molero, G., Reynolds, M., Gouis, J. L., and Calderini, D. F. (2014). Trade-off between grain weight and grain number and key traits for increasing potential grain weight in CIMCOG population

Raj, R., Walker, J. P., Pingale, R., Nandan, R., Naik, B., and Jagarlapudi, A. (2021). Leaf area index estimation using top-of-canopy airborne RGB images. *International Journal of Applied Earth Observation and Geoinformation* 96, 102282. doi:10.1016/j.jag.2020.102282

Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., et al. (2018). Trends in Global Agricultural Land Use: Implications for Environmental Health and Food Security. *Annual Review of Plant Biology* 69, 789–815. doi:10.1146/annurev-arplant-042817-040256

Ray, D., Gerber, J., Macdonald, G., and West, P. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications* 6. doi:10.1038/ncomms6989

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE* 8, e66428. doi:10.1371/journal.pone.0066428

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., and Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nature Communications* 3, 1293. doi:10.1038/ncomms2296

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N., et al. (2020). Breeder friendly phenotyping. *Plant Science* 295, 110396. doi:10.1016/j.plantsci.2019.110396

Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., et al. (2012). Achieving yield gains in wheat. *Plant, Cell & Environment* 35, 1799–1823. doi:10.1111/j.1365-3040.2012.02588.x

Reynolds, M., Foulkes, M. J., Slafer, G. A., Berry, P., Parry, M. A. J., Snape, J. W., et al. (2009). Raising yield potential in wheat. *Journal of Experimental Botany* 60, 1899–1918. doi:10.1093/jxb/erp016

Rogers, E. M. (1995). Lessons for Guidelines from the Diffusion of Innovations. *The Joint Commission Journal on Quality Improvement* 21, 324–328. doi:10.1016/S1070-3241(16)30155-9

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment* 55, 95–107. doi:10.1016/0034-4257(95)00186-7

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, eds. N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Cham: Springer International Publishing), Lecture Notes in Computer Science, 234–241. doi:10.1007/978-3-319-24574-4_28

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65, 386–408. doi:10.1037/h0042519

Roth, L., Barendregt, C., Bétrix, C.-A., Hund, A., and Walter, A. (2022a). High-throughput field phenotyping of soybean: Spotting an ideotype. *Remote Sensing of Environment* 269, 112797. doi:10.1016/j.rse.2021.112797

Roth, L., Camenzind, M., Aasen, H., Kronenberg, L., Barendregt, C., Camp, K.-H., et al. (2020). Repeated Multiview Imaging for Estimating Seedling Tiller Counts of Wheat Genotypes Using Drones. *Plant Phenomics* 2020. doi:10.34133/2020/3729715

Roth, L., Piepho, H.-P., and Hund, A. (2022b). Phenomics data processing: Extracting dose–response curve parameters from high-resolution temperature courses and repeated field-based wheat height measurements. *in silico Plants* 4, diac007. doi:10.1093/insilicoplants/diac007

Roth, L. and Streit, B. (2018). Predicting cover crop biomass by lightweight UAS-based RGB and NIR photography: An applied photogrammetric approach. *Precision Agriculture* 19, 93–114. doi:10.1007/s11119-017-9501-1

Roujean, J.-L. and Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* 51, 375–384. doi:10.1016/0034-4257(94)00114-3

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 211–252. doi:10.1007/s11263-015-0816-y

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., and Hawkesford, M. J. (2019). DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Frontiers in Plant Science* 10

Saint Cast, C., Lobet, G., Cabrera-Bosquet, L., Couvreur, V., Pradal, C., Tardieu, F., et al. (2022). Connecting plant phenotyping and modelling communities: Lessons from science mapping and operational perspectives. *in silico Plants* 4, diac005. doi:10.1093/insilicoplants/diac005

Sarić, R., Nguyen, V., Burge, T., Berkowitz, O., Trtílek, M., Whelan, J., et al. (2022). Applications of hyperspectral imaging in plant phenotyping. *Trends in Plant Science* 27, 301–315. doi:10.1016/j.tplants.2021.12.003

Schiefer, F., Schmidtlein, S., and Kattenborn, T. (2021). The retrieval of plant functional traits from canopy spectra through RTM-inversions and statistical models are both critically affected by plant phenology. *Ecological Indicators* 121. doi:10.1016/j.ecolind.2020.107062

Schierenbeck, M., Fleitas, M. C., Cortese, F., Golik, S. I., and Simón, M. R. (2019a). Nitrogen accumulation in grains, remobilization and post-anthesis uptake under tan spot and leaf rust infections on wheat. *Field Crops Research* 235, 27–37. doi:10.1016/j.fcr.2019.02.016

Schierenbeck, M., Fleitas, M. C., and Simón, M. R. (2019b). Nitrogen fertilization and fungicide mixtures in wheat: How do they affect the severity, yield and dynamics of nitrogen under leaf rust infections? *European Journal of Plant Pathology* 155, 1061–1075. doi:10.1007/s10658-019-01832-w

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji - an Open Source platform for biological image analysis. *Nature methods* 9, 10.1038/nmeth.2019. doi:10.1038/nmeth.2019

Schirrmann, M., Landwehr, N., Giebel, A., Garz, A., and Dammer, K.-H. (2021). Early Detection of Stripe Rust in Winter Wheat Using Deep Residual Neural Networks. *Frontiers in Plant Science* 12

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision* 128, 336–359. doi:10.1007/s11263-019-01228-7

Serouart, M., Madec, S., David, E., Velumani, K., Lopez Lozano, R., Weiss, M., et al. (2022). SegVeg: Segmenting RGB Images into Green and Senescent Vegetation by Combining Deep and Shallow Methods. *Plant Phenomics* 2022. doi:10.34133/2022/9803570

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., and Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security* 5, 291–317. doi:10.1007/s12571-013-0263-y

Simmonds, J., Scott, P., Leverington-Waite, M., Turner, A. S., Brinton, J., Korzun, V., et al. (2014). Identification and independent validation of a stable yield and thousand grain weight QTL on chromosome 6A of hexaploid wheat (Triticum aestivum L.). *BMC Plant Biology* 14, 191. doi:10.1186/s12870-014-0191-9

Simón, M. R., Fleitas, M. C., Castro, A. C., and Schierenbeck, M. (2020). How Foliar Fungal Diseases Affect Nitrogen Dynamics, Milling, and End-Use Quality of Wheat. *Frontiers in Plant Science* 11

Singh, A., Ganapathysubramanian, B., Singh, A., and Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science* 21, 110–124. doi:10.1016/j.tplants.2015.10.015

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in Plant Science* 23, 883–898. doi:10.1016/j.tplants.2018.07.004

Slafer, G. A., Calderini, D. F., and Miralles, D. J. (1996). Yield components and compensation in wheat: Opportunities for further increasing yield potential. *Increasing yield potential in wheat: Breaking the Barriers* , 101–133

Slafer, G. A., Savin, R., and Sadras, V. O. (2014). Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research* 157, 71–83. doi:10.1016/j.fcr.2013.12.004

Song, X., Yang, G., Xu, X., Zhang, D., Yang, C., and Feng, H. (2022). Winter Wheat Nitrogen Estimation Based on Ground-Level and UAV-Mounted Sensors. *Sensors* 22. doi:10.3390/s22020549

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging* 32, 1153–1190. doi:10.1109/TMI.2013.2265603

Sripada, R. P., Heiniger, R. W., White, J. G., and Meijer, A. D. (2006). Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agronomy Journal* 98, 968–977. doi:10.2134/agronj2005.0200

[Dataset] Standley, T., Zamir, A. R., Chen, D., Guibas, L., Malik, J., and Savarese, S. (2020). Which Tasks Should Be Learned Together in Multi-task Learning? doi: 10.48550/arXiv.1905.07553

Sun, D., Robbins, K., Morales, N., Shu, Q., and Cen, H. (2022). Advances in optical phenotyping of cereal crops. *Trends in Plant Science* 27, 191–208. doi:10.1016/j.tplants.2021.07.015

Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J.-L., and Sorrells, M. E. (2017). Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *The Plant Genome* 10, plantgenome2016.11.0111. doi:10.3835/plantgenome2016.11.0111

Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E., and Mittler, R. (2014). Abiotic and biotic stress combinations. *New Phytologist* 203, 32–43. doi:10.1111/nph.12797

Tan, M. and Le, Q. V. (2020). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks doi:10.48550/arXiv.1905.11946

Tanner, F., Tonn, S., de Wit, J., Van den Ackerveken, G., Berger, B., and Plett, D. (2022). Sensor-based phenotyping of above-ground plant-pathogen interactions. *Plant Methods* 18, 35. doi:10.1186/s13007-022-00853-7

Tayade, R., Yoon, J., Lay, L., Khan, A., Yoon, Y., and Kim, Y. (2022). Utilization of Spectral Indices for High-Throughput Phenotyping. *Plants* 11. doi:10.3390/plants11131712

Thenkabail, P. S. and Lyon, J. G. (eds.) (2011). *Hyperspectral Remote Sensing of Vegetation* (Boca Raton: CRC Press). doi:10.1201/b11222

Tilly, N., Aasen, H., and Bareth, G. (2015). Fusion of Plant Height and Vegetation Indices for the Estimation of Barley Biomass. *Remote Sensing* 7, 11449–11480. doi:10.3390/rs70911449

Tilneac, M., Dolga, V., Grigorescu, S., and Bitea, M. A. (2012). 3D Stereo Vision Measurements for Weed-Crop Discrimination. *Elektronika ir Elektrotechnika* 123, 9–12. doi:10.5755/j01.eee.123.7.2366

[Dataset] Vafaeikia, P., Namdar, K., and Khalvati, F. (2020). A Brief Review of Deep Multi-task Learning and Auxiliary Task Learning. doi:10.48550/arXiv.2007.01126

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., et al. (2014). Scikit-image: Image processing in Python. *PeerJ* 2, e453

van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., et al. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science* 282, 23–39. doi:10.1016/j.plantsci.2018.06.018

Vandenhende, S., Georgoulis, S., Proesmans, M., Dai, D., and Gool, L. (2020). Revisiting Multi-Task Learning in the Deep Learning Era. *ArXiv*

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., et al. (2019). Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics* 40, 589–629. doi:10.1007/s10712-018-9478-y

Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., and Hawkesford, M. J. (2017). Field Scanalyzer: An automated robotic field phenotyping platform for detailed crop monitoring. *Functional Plant Biology* 44, 143. doi:10.1071/FP16163

Voss-Fels, K. P., Stahl, A., and Hickey, L. T. (2019). Q&A: Modern crop breeding for future food security. *BMC Biology* 17, 18. doi:10.1186/s12915-019-0638-4

Wan, L., Li, H., Li, C., Wang, A., Yuheng, Y., and Wang, P. (2022). Hyperspectral Sensing of Plant Diseases: Principle and Methods. *Agronomy* 12, 1451. doi:10.3390/agronomy12061451

Wan, L., Zhang, J., Dong, X., Du, X., Zhu, J., Sun, D., et al. (2021). Unmanned aerial vehicle-based field phenotyping of crop biomass using growth traits retrieved from PROSAIL model. *Computers and Electronics in Agriculture* 187, 106304. doi:10.1016/j.compag.2021.106304

Wang, F., Yang, M., Ma, L., Zhang, T., Qin, W., Li, W., et al. (2022a). Estimation of Above-Ground Biomass of Winter Wheat Based on Consumer-Grade Multi-Spectral UAV. *Remote Sensing* 14, 1251. doi:10.3390/rs14051251

Wang, K., Huggins, D. R., and Tao, H. (2019). Rapid mapping of winter wheat yield, protein, and nitrogen uptake using remote and proximal sensing. *International Journal of Applied Earth Observation and Geoinformation* 82, 101921. doi:10.1016/j.jag.2019.101921

Wang, W., Wu, Y., Zhang, Q., Zheng, H., Yao, X., Zhu, Y., et al. (2021). AAVI: A Novel Approach to Estimating Leaf Nitrogen Concentration in Rice From Unmanned Aerial Vehicle Multispectral Imagery at Early and Middle Growth Stages. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 6716–6728. doi:10.1109/JSTARS.2021.3086580

Wang, W., Zheng, H., Wu, P., Yao, X., Zhu, Y., Cao, W., et al. (2022b). An assessment of background removal approaches for improved estimation of rice leaf nitrogen concentration with unmanned aerial vehicle multispectral imagery at various observation times. *Field Crops Research* 283, 108543. doi:10.1016/j.fcr.2022.108543

Weiss, M., Jacob, F., and Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment* 236, 111402. doi:10.1016/j.rse.2019.111402

Whetton, R. L., Hassall, K. L., Waine, T. W., and Mouazen, A. M. (2018). Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 1: Laboratory study. *Biosystems Engineering* 166, 101–115. doi:10.1016/j.biosystemseng.2017.11.008

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018. doi:10.1038/sdata.2016.18

Xie, P., Du, R., Ma, Z., and Cen, H. (2023). Generating 3D Multispectral Point Clouds of Plants with Fusion of Snapshot Spectral and RGB-D Images. *Plant Phenomics* 5, 0040. doi:10.34133/plantphenomics.0040

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, 6715–6730. doi:10.1093/jxb/erv378

Xu, R. and Li, C. (2022). A Review of High-Throughput Field Phenotyping Systems: Focusing on Ground Robots. *Plant Phenomics* 2022. doi:10.34133/2022/9760269

Yang, K.-W., Chapman, S., Carpenter, N., Hammer, G., McLean, G., Zheng, B., et al. (2021a). Integrating crop growth models with remote sensing for predicting biomass yield of sorghum. *in silico Plants* 3, diab001. doi:10.1093/insilicoplants/diab001

Yang, S., Hu, L., Wu, H., Ren, H., Qiao, H., Li, P., et al. (2021b). Integration of Crop Growth Model and Random Forest for Winter Wheat Yield Estimation From UAV Hyperspectral Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 6253–6269. doi:10.1109/JSTARS.2021.3089203

Yu, K., Anderegg, J., Mikaberidze, A., Karisto, P., Mascher, F., McDonald, B. A., et al. (2018). Hyperspectral Canopy Sensing of Wheat Septoria Tritici Blotch Disease. *Frontiers in Plant Science* 9

Yu, K., Kirchgessner, N., Grieder, C., Walter, A., and Hund, A. (2017). An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping. *Plant Methods* 13, 15. doi:10.1186/s13007-017-0168-4

Yue, J., Feng, H., Li, Z., Zhou, C., and Xu, K. (2021). Mapping winter-wheat biomass and grain yield based on a crop model and UAV remote sensing. *International Journal of Remote Sensing* 42, 1577–1601. doi:10.1080/01431161.2020.1823033

Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., and Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS Journal of Photogrammetry and Remote Sensing* 150, 226–244. doi:10.1016/j.isprsjprs.2019.02.022

Zarco-Tejada, P. J., Poblete, T., Camino, C., Gonzalez-Dugo, V., Calderon, R., Hornero, A., et al. (2021). Divergent abiotic spectral pathways unravel pathogen stress signals across species. *Nature Communications* 12, 6088. doi:10.1038/s41467-021-26335-3

Zhang, C. and Kovacs, J. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture* 13, 693–712. doi:10.1007/s11119-012-9274-5

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., et al. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture* 165, 104943. doi:10.1016/j.compag.2019.104943

Zhang, J., Tian, H., Wang, P., Tansey, K., Zhang, S., and Li, H. (2022). Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China.

Computers and Electronics in Agriculture 192, 106616. doi:10.1016/j.compag.2021.106616

Zhou, B., Elazab, A., Bort, J., Vergara, O., Serret, M. D., and Araus, J. L. (2015). Low-cost assessment of wheat resistance to yellow rust through conventional RGB images. *Computers and Electronics in Agriculture* 116, 20–29. doi:10.1016/j.compag.2015.05.017

Zhou, C., Liang, D., Yang, X., Xu, B., and Yang, G. (2018). Recognition of Wheat Spike from Field Based Phenotype Platform Using Multi-Sensor Fusion and Improved Maximum Entropy Segmentation Algorithms. *Remote Sensing* 10, 246. doi:10.3390/rs10020246

Zhu, Y., Sun, G., Ding, G., Zhou, J., Wen, M., Jin, S., et al. (2021). Large-scale field phenotyping using backpack LiDAR and CropQuant-3D to measure structural variation in wheat. *Plant Physiology* 187, 716–738. doi:10.1093/plphys/kiab324