Maynooth University
National University of Ireland Maynooth

Using proximal RGB and multi-spectral cameras to measure health parameters in cereal crops

Vytenis Edvardas Jakunskas

A thesis submitted for the degree of
Master of Science

Maynooth University
Faculty of Science and Engineering
Electronic Engineering Department

October 2025

Supervisor: Prof. Gerard Lacey

Abstract

Plant phenotyping enables quantitative monitoring of crop performance and stress responses. The LI-COR LI-600 is the gold standard for non-destructive leaf-level physiological measurement. However, it is slow, labour-intensive, and costly, limiting its suitability for large-scale or high-frequency studies. Remote sensing via satellites or UAVs enables large-scale monitoring but faces resolution and environmental limitations. Proximal imaging, conducted within 1-2 metres of the plant canopy, allows substantially higher spatial and temporal resolution. This enables yield and biomass prediction and pest, disease, and weed identification, all are difficult to perform reliably at UAV or satellite scales.

This thesis describes the design and evaluation of robotic High-Throughput Phenotyping Systems (HTPS) using proximal RGB and multispectral imaging. Two gantry-based robots were developed: a 3 $\times$ 6 m RGB HTPS for greenhouse imaging of potted oats, and an outdoor raised-bed HTPS using a custom multispectral camera to image oat and barley rows. Both systems were paired with LI-COR measurements; yield and biomass were measured outdoors. The image processing compared conventional plant phenotyping metrics with three deep learning models (ResNet50, Swin, EfficientNet-B5) to predict plant parameters.

This thesis answers two key research questions: (1) Can proximal RGB imaging predict LI-COR physiological measurements for individual oat plants in a greenhouse setting? The SWIN transformer achieved a prediction $R^2 = 0.27$ for leaf vapour pressure and $R^2 = 0.40$ for leaf vapour pressure deficit, and an $R^2 = 0.41$ for stomatal conductance. (2) Can proximal multispectral imaging estimate LI-COR physiological metrics, yield and biomass for oats and barley in outdoor field conditions? In field conditions, the SWIN transformer achieved a prediction $R^2 = 0.78$ for leaf vapour pressure, an $R^2 = 0.67$ for leaf vapour pressure deficit, and an $R^2 = 0.29$ for stomatal conductance. Resnet predicted yield in oats with $R^2 = 0.52$ and biomass in oats and barley combined with $R^2 = 0.56$. Two new datasets were produced and will be publicly released, including the first documenting visible predation in cereal crops.

Declaration

I declare that this thesis is entirely my own work and has not been submitted in whole or in part for the award of any other degree or qualification at this or any other university.

Where the work of others has been used, it is duly acknowledged and referenced.

Portions of this thesis were revised and refined using generative AI tools (OpenAI's ChatGPT) and grammar assistance software (Grammarly) for grammar and phrasing suggestions. All final content is my own, and no AI-generated text was included without substantive editing.

Edward Jakunskas

Acknowledgments

I would like to sincerely thank Dr. Gerard Lacey for his exceptional supervision throughout this project. His clear direction, insightful guidance, and continuous support were invaluable at every stage.

I am grateful to Conor Meade and Duyen Hoang for conducting the biology experiments and for their generous assistance with any biological questions I encountered. Thanks also to Adam McEneaney for his help in setting up the initial FarmBot hardware for the greenhouse RGB HTPS system.

Special thanks to James Petri and Greg Mantel for their wide-ranging technical contributions and consistent support throughout the project. James translated the FarmBot control system to ROS, developed the custom pivoting arm, configured the NAS, and assisted with numerous core components. Greg supported LI-COR data collection and FarmBot operations during the outdoor multispectral HTPS runs, and created and printed critical 3D designs, including the camera mount for the outdoor HTPS and cable carrier holders for both systems.

Finally, I would like to thank the AURA team — Majid Sorouri, Rana Umair Hameed, Cael Timmons, Heather Bruen, and Marina Martinez — for their help during the harvest, their support throughout the field days, and for being a great team to work with.

To my fiancée Beatriz de Oliveira Teixeira, thank you for your unwavering support, for helping harvest oats and barley on that one weekend, and for always being there when I needed you most.

Contents

List of Figures

List of Tables

Acronyms

AURA Automating Regenerative Agriculture.

CON Contrast.

COR Correlation.

DIS Dissimilarity.

ENT Entropy.

$\mathbf{F}_m$ Maximum Fluorescence.

$\mathbf{F}_s$ Steady-State Fluorescence.

$\mathbf{g}_{sw}$ Stomatal Conductance.

$\mathbf{g}_{tw}$ Total Conductance.

$\mathbf{H_2O_{leaf}}$ Leaf H_2O Mole Fraction.

HOM Homogeneity.

HTPS High-Throughput Plant Phenotyping System.

IDW Inverse Distance Weighting.

LI-600 LICOR LI-600 Porometer/Flourometer.

LiDAR Light Detection and Ranging.

MEA Mean.

MS Multispectral.

NAS Network-Attached Storage.

NIR Near-Infrared.

RE Red-Edge.

RealSense Intel RealSense Camera.

RGB Red-Green-Blue.

ROS Robot Operating System.

SAM Segment Anything Model.

SEC Second Moment.

$\mathbf{T_{leaf}}$ Leaf Temperature.

TGW Thousand Grain Weight.

VAR Variance.

VP Vapour Pressure.

VPD Vapour Pressure Deficit.

YOLO You Only Look Once.

Chapter 1

Introduction

1.1 Broad Context

Modern plant phenotyping enables researchers and farmers to assess crop performance and respond to environmental stress in a timely and data-driven manner. Regenerative agriculture in particular depends on detailed insight into plant responses to environmental inputs such as drought, nutrient availability, and soil structure.

While precision agriculture has made strides in crops like vegetables [3, 4, 5, 6] and vines [7, 8], cereal crops such as oats remain under-explored. This is despite their central role in global food systems, as cereals account for roughly 50% of global caloric intake and 41% of protein according to the United Nations [9].

Row crop farmers face a mix of pressures: labour shortages, weather-driven timing constraints, and the need to raise output while protecting soil and margins. While precision agriculture tools and methods can help alleviate these, various adoption frictions come up repeatedly in recent surveys and farmer interviews. These include high up-front cost with uncertain return on investment, technical complexity, weak interoperability with existing equipment, limited training, and-most relevant to this thesis-too much non-actionable data [10].

A recent study [10] combined a literature review with interviews of Irish and Atlantic-European farmers to map farmer needs to current technologies. One theme has stood out and is the main driver of this thesis: Farmers value per-plant, actionable outputs when those outputs shorten the path from detection to intervention (spot spraying, narrow-row mechanical weeding, targeted scouting). This is opposed to the gathering of large amounts of data that is difficult to act upon in practice.

In this context, there is a growing need for affordable, high-throughput phenotyping systems (HTPS) that provide interpretable and actionable information at the plant level. Such systems must balance the precision of direct physiological measurements with the scalability of imaging-based methods, bridging the gap between scientific measurement tools and on-farm decision making.

This work sits within Maynooth University's *Automating Regenerative Agriculture (AURA)* project, whose goal is to develop a fully electrified, automated robotic system capable of large-scale regenerative farming. The platform will integrate a HTPS to generate per-plant, actionable measurements at field scale and frequency, closing the loop between sensing and intervention. This thesis provides the imaging foundations for that platform: it evaluates sensor modality and pose, defines targets and metrics most predictive of LI-600 physiological traits, informs AURA's sensor selection, mounting geometry, and software stack [11].

1.2 High Throughput Phenotyping Systems

The LI-COR LI-600 Porometer/Fluorometer is the current gold standard for collecting accurate, direct measurements of plant physiological traits non-destructively in the field. Measurements are typically taken from the top-most fully expanded leaf, as it is the most physiologically active and first to exhibit visible symptoms of water stress. While accurate and reliable, the LI-600 is a handheld instrument requiring manual alignment and stabilisation, and its throughput is typically limited to approximately 60–120 measurements per hour. As a result, it is poorly suited to large-scale or high-frequency phenotyping workflows, except as a source of ground-truth data for training models.

To overcome these limitations, a range of HTPS have been developed to automate or accelerate plant trait measurement. HTPS are automated or semi-automated platforms designed to capture large-scale plant trait data using imaging, spectrometry, or other sensor modalities. UAVs are a great example, able to capture data over large areas, up to hundreds of hectares per flight. While making up completely for the biggest pain-points of handheld measurement, it introduces new issues. The downwash created by multi rotor drones can move leaves and alter the microclimate, such as reducing canopy temperature [12]. The low-level ag airspace is full of potential hazards, such as trees, poles, and wires. Area per image scales with the square of altitude, so reducing mission height leads to approximately quadratic increases in flight time (assuming roughly constant overlap and flight speed). This

is compounded by lower flight heights requiring slower flight speed to avoid motion blur. This increase in flight time means more battery swaps. All of these factors combine to typical drone imaging mission altitudes of 40m - 100m [13].

This results in a ground sample distance (GSD) greater-than-centimetre in realistic missions, this kind of resolution doesn't allow measurement of important metrics like yield through individual grain counting. It must be noted that sub-mm has been shown in very specialized setups. For example; one study [14] used a 135 mm lens at 10 m height resulting in a resolution of around 0.3 mm/px. They manually positioned, aligned and hovered over each plot for taking images. This approach is difficult to scale due to long mission times, battery constrains, GNSS resolution errors forcing manual operation and the inherent risk of flying at such low altitudes. UAVs are also at the mercy of the elements, with drone manuals prohibiting flight missions in rain and strong wind [15]. This is of particular importance in counties like Ireland, where the number of wet days in 2024 ranged from 129 to 220 depending on location [16].

Another avenue of remote sensing is satellite imaging. They offer repeatable time-series with useful spectral bands, Sentinel-2 for example provides 13 spectral bands (four at 10 m GSD, six at 20 m GSD, three at 60 m GSD), covering a 290 km swathe of ground per pass, with a revisit schedule of approximately 5 days [17]. While specifics differ, other satellites operate on similar magnitudes of scale. This makes satellite data ideal for large-scale variability mapping and seasonal trend analysis. Their coarse GSD and weather sensitivity generally make them less suitable for plant-level diagnosis and fast intervention - ie. actionable data that farmers want.

Sitting somewhere in between UAV remote sensing and handheld sensor measurements, is proximal imaging. Proximal imaging in agriculture refers to ground-based sensing of crops from within the field at short stand-off distances using cameras or spectrometers mounted on handheld rigs, gantries, tractors or field robots. This approach enables much higher spatial and temporal resolution than UAV or satellite systems, while maintaining practical coverage and operational stability. If the camera is mounted on a field robot, mission duration becomes far less of a limitation than in UAV-based imaging, as battery weight is not a significant constraint on ground platforms. Robotic systems also offer the opportunity for seamless integration of imaging with other field operations such as pruning, harvesting, or targeted spraying. This enables repeatable, high-frequency data collection either as a dedicated phenotyping task or as a by-product of regular fieldwork. Moreover, the data gathered in real time can directly inform autonomous decisions, such as precision weeding or spot spraying, creating a closed loop between perception and action.

Although close-range imaging inherently requires longer mission times to cover a given area, this trade-off is less critical than for aerial systems because ground robots can operate continuously and autonomously. The short stand-off distance also unlocks new sensing possibilities: at proximal range, it is simple to achieve sub-millimetre GSD, a level of detail difficult to obtain from UAVs. Such resolution enables a new scope of analysis, including individual grain counting, lesion and pest detection, as well as assessment of subtle morphological cues such as leaf wilting, texture, or glossiness.

Finally, because the sensor height above the soil is fixed and known, it becomes possible to estimate absolute plant height directly from onboard depth sensors or stereo vision. This additional geometric information complements spectral and colour data.

1.3 LI-COR LI-600 Physiological Metrics

The LI-600 provides accurate, non-destructive measurements of key leaf-level physiological traits. The main parameters used in this work are summarised below to clarify their biological significance and their relationship to plant water status and stress responses.

Stomatal conductance to water (g_{sw}) is a measure of the degree of stomatal openness and density.This measurement can be useful for phenotyping and indicates a plant's physiological response to its environmental condition. [18]

Total conductance to water vapour (g_{tw}) is the combined measure of gas exchange through the stomata, boundary layer, and sometimes the cuticle, representing the plant's overall capacity for gas exchange. It was excluded from subsequent analyses because it was found to be a scaled version of g_{sw}, providing no independent information.[18]

Leaf vapor pressure ($\mathrm{VP_{leaf}}$) is the partial pressure of water vapour in the leaf microenvironment, while *vapor pressure deficit* ($\mathrm{VPD_{leaf}}$) indicates the difference between the leaf's vapour pressure and the saturation vapour pressure of the surrounding air. [19]

Leaf H_2O mole fraction ($\mathrm{H_2O_{leaf}}$) represents the proportion of water vapour near the leaf surface, and *leaf temperature* ($\mathrm{T_{leaf}}$) is the thermal measurement of the leaf, influencing physiological processes and gas exchange. The $\mathrm{T_{leaf}}$ column was excluded

from subsequent analyses because the LI-600 dataset contained corrupted values.[18]

Steady-state fluorescence (F_s) refers to the constant fluorescence emitted under continuous illumination, while *maximum fluorescence* (F_m) is maximum fluorescence yield in a light-adapted leaf. Both are typically high in healthy plants, reflecting efficient light absorption and energy transfer.[18]

In general, stomatal conductance decreases under water stress as stomata close to reduce transpiration. Leaf temperature increases under drought due to limited evaporative cooling. Increasing vapour pressure deficit indicates greater transpiration demand, while lower leaf vapour pressure and H_2O mole fraction values may signal stomatal closure and reduced moisture retention. These relationships make the LI-600 parameters valuable ground-truth indicators of plant water status for calibrating imaging-based models.

1.4 Imaging Modalities and Experimental Systems

Predicting stomatal conductance, water-related parameters, biomass, and yield using multispectral imaging is a well-established practice in precision agriculture. Numerous studies across different crops [20, 2, 21, 22] have demonstrated that vegetation indices and end-to-end machine learning models applied to multi-band data can estimate these metrics with reasonable accuracy.

However, there is a clear gap in the literature regarding how well multispectral imaging can predict LI-600 physiological parameters-particularly in oats and barley-and an overall lack of research on the use of proximal imaging for this purpose. Most existing work relies on UAV-based data, where the achievable spatial resolution and imaging frequency are constrained by flight logistics and regulations. This raises an important question: at close range, where sub-millimetre detail and stable environmental conditions are achievable, do the additional spectral bands of a multispectral camera provide a meaningful advantage over RGB imaging? To address this, we designed two complementary robotic HTPS: one equipped with RGB cameras for controlled greenhouse experiments, and another incorporating a custom-built multispectral (RGB+red edge (RE)+ near infrared (NIR)) camera for outdoor field imaging. This dual approach enables a direct comparison between RGB and multispectral imaging modalities under otherwise similar conditions, allowing us to evaluate whether increased spectral complexity improves prediction of these physiological parameters at proximal range.

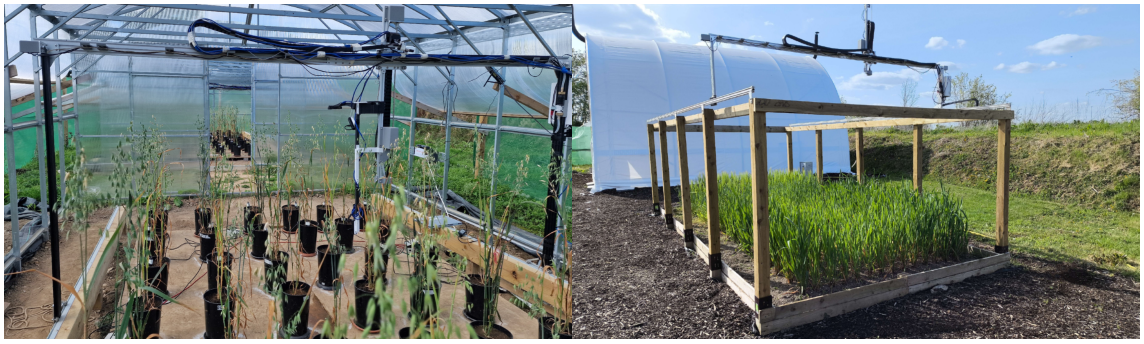

Figure 1.1: Left) The multi-RGB-camera greenhouse HTPS used for controlled drought experiments on potted oats. Right) The outdoor multispectral HTPS featuring a custom five-channel (RGB+RE+NIR) camera mounted above a raised bed containing oats and barley. These two complementary systems form the experimental foundation for the research questions outlined in the following section.

While hyperspectral cameras could, in principle, capture even more detailed spectral information, their cost, data volume, and acquisition speed make them impractical for autonomous, high-frequency field operation. Our focus therefore remains on the balance between information content, affordability, and scalability-factors that are critical for real-world agricultural deployment.

1.5 Research Questions and Contributions

This thesis addresses several gaps in current high-throughput plant phenotyping research. While previous work has demonstrated the potential of UAV and satellite imaging for estimating crop stress and productivity, there remains limited understanding of how close-range imaging can predict direct physiological measurements such as those obtained from the LI-600. This section outlines the key research questions investigated in this thesis and the main contributions that arose from addressing them.

1.5.1 Can proximal RGB imaging predict LI-COR metrics in controlled greenhouse settings?

To answer this question, a robotic High-Throughput Phenotyping System was developed for controlled greenhouse experiments using RGB imaging only. The platform consisted of a modified FarmBot Genesis XL gantry platform equipped with multiple Luxonis OAK-D Lite cameras capturing top-down and angled views of each potted oat plant. These plants were subjected to varying drought treatments and

periodically measured with the LI-600 to provide accurate physiological ground-truth data.

Statistical and machine-learning models were trained on the paired image and LI-600 dataset. The results demonstrate that proximal RGB imaging captures meaningful visual cues correlated with physiological traits, despite relying solely on visible light. The best-performing model (Swin Transformer) achieved $R^2 = 0.35$ for leaf water content, $R^2 = 0.40$ for leaf vapour pressure, and $R^2 = 0.41$ for stomatal conductance. These findings confirm that close-range RGB imaging, combined with modern computer vision and machine learning, can approximate LI-COR-derived physiological parameters.

1.5.2 Can proximal multispectral imaging estimate LI-COR metrics, yield and biomass in open field conditions?

A second robotic HTPS was implemented outdoors using a custom-built five-channel multispectral camera mounted above a raised bed containing both oats and barley. This setup enabled close-range imaging under natural light while collecting paired LI-COR measurements through targeted sampling, along with biomass and yield data.

Machine-learning models were trained to predict both physiological and agronomic traits from the multispectral imagery. The Swin Transformer achieved $R^2 = 0.69$ for leaf water content and vapour pressure deficit and $R^2 = 0.32$ for stomatal conductance using the five-channel input. The ResNet model predicted yield with $R^2 = 0.71$ and biomass with $R^2 = 0.49$. These results confirm that low-cost multispectral imaging can extend proximal phenotyping from controlled greenhouse environments to realistic multi-species field conditions, providing accurate predictions of both physiological and yield-related traits.

1.5.3 Image Processing Pipeline for Plant and Organ Segmentation

As part of this research, a modular image-processing pipeline was developed for plant instance and organ-level segmentation, integrating recent models such as Depth Anything, Segment Anything, SAM2, and YOLOv11. This pipeline isolates plant organs such as the top-most leaf for targeted trait prediction. The methods and datasets developed here contribute a reusable workflow for organ-level phenotyping

from RGB imagery.

1.5.4 Dataset Release and Novel Field Conditions

The datasets generated in this work include paired RGB and multispectral imagery, LI-600 measurements, biomass, and yield data for oats and barley grown under both greenhouse and field conditions. A substantial subset of the field dataset also contains visible signs of bird predation-an underexplored problem in cereal imaging. The release of these annotated datasets introduces a new benchmark for low-cost phenotyping and supports future research in plant–environment interaction.

1.5.5 Synthetic Training Data via Interpolation of LI-600 Measurements

In the outdoor experiment, LI-COR measurements were collected for approximately 15% of the plants due to the time-intensive nature of manual data acquisition. To overcome this limitation, the thesis explores whether spatial interpolation of the sparse LI-600 readings can be used to generate synthetic training data, increasing the effective sample size without additional measurements. This approach demonstrates a practical strategy for improving regression model performance when direct physiological data collection is constrained by labour or time.

The following chapters expand on these contributions in detail, describing the design, implementation, and evaluation of both the greenhouse RGB and outdoor multispectral HTPS.

1.6 Map of thesis

Chapter 1 introduced the motivation for this research within the broader context of plant phenotyping and regenerative agriculture. It outlined the limitations of existing sensing methods and established the rationale for developing proximal, robotic imaging systems capable of linking visual data with physiological plant traits.

Chapter 2 reviews the relevant literature on proximal imaging platforms, RGB and multispectral phenotyping, and the application of deep learning in plant science. This positions the work within existing research and identifies the technological and methodological gaps that the thesis addresses.

Chapter 3 describes the design of both HTPS, outlining their mechanical construction, imaging hardware, agronomic setup, and data-collection protocols.

Chapter 4 details the HTPS implementation and the development of a unified image-processing pipeline, integrating segmentation and deep-learning components to extract plant and organ-level features for trait prediction.

Chapter 5 validates each system through controlled experiments, quantifying imaging and segmentation accuracy, and assessing preprocessing methods such as colour correction.

Chapter 6 presents the results of the predictive analyses, comparing traditional statistical approaches with three deep-learning architectures (ResNet50, EfficientNet-B5, and Swin Transformer) for both greenhouse and field datasets.

Finally, Chapter 7 discusses the findings in the context of the two research questions, summarising the main contributions, limitations, and implications for future high-throughput phenotyping research and on-farm decision support.

Chapter 2

Literature Review

This chapter reviews prior work relevant to the development and application of HTPS. It is structured by sensing modality and analytical approach, beginning with robotic and imaging platforms, followed by RGB and MS-based phenotyping methods, deep learning and image segmentation strategies, and yield + biomass estimation studies.

2.1 Proximal Imaging Platforms

Recent reviews highlight the rapid development of proximal imaging systems for high-throughput phenotyping in response to climate-related challenges and the need for scalable, non-destructive monitoring of plant traits [23, 24].

Early gantry-based systems were primarily developed for controlled environments such as greenhouses, where repeatable positioning and consistent lighting simplify image acquisition. Examples include the LemnaTec Scanalyzer and other conveyor-based plant phenotyping systems, which remain a benchmark for precision but are costly and infrastructure-heavy. More affordable open-source alternatives, such as the FarmBot platform, have been adapted for automated imaging and trait extraction. One study [6] has demonstrated a modified FarmBot for estimating crop coverage and growth dynamics in mixed-species plots, illustrating how off-the-shelf components can enable robust plant monitoring at low cost. The greenhouse system developed in this thesis builds on a similar approach, integrating low-cost RGB cameras and custom control software to achieve fine-grained imaging of individual pots.

Beyond fixed gantries, several mobile robotic platforms have been developed to enable proximal imaging in open-field environments [25, 26, 27, 28]. Farm-ng's Amiga, shown

Figure 2.1: The Farm-ng Amiga Platform

in Figure 2.1 is a lightweight, modular platform designed for research and small-scale farming. Its compact form factor and open hardware architecture make it well suited for sensor integration. Saga Robotics' Thorvald and Thorvald II [29, 30], shown in Figure 2.2 are larger, highly configurable platforms that can support a variety of sensors, including RGB, MS, and thermal cameras, across diverse environments such as open fields, poly-tunnels, and vineyards. Other commercial systems such as Robotti, Solix, and Orio [10, 31, 32] further demonstrate the growing diversity of agricultural robots capable of supporting proximal imaging systems. While these platforms are essential for practical field-scale applications, our research focuses on precise tracking of individual plants, which would be challenging to achieve with the coarse positional accuracy of mobile robot GNSS systems. Moreover, working with a lightweight, static gantry system provides faster iteration cycles and reduced operational overhead, enabling rapid development and testing of imaging pipelines before transitioning to larger, more complex robotic systems.

Our raised-bed phenotyping platform combines the positional accuracy of a gantry with the environmental realism of field conditions, forming an intermediate step toward deployment on a mobile robot.

Figure 2.2: The Thorvald II Agricultural Robotic System

2.2 RGB Phenotyping

RGB imaging remains one of the most accessible and cost-effective sensing modalities for plant phenotyping. Despite lacking the narrow-band spectral detail of multi-spectral or hyperspectral systems, RGB cameras offer fine spatial resolution, high acquisition speed, and rich colour and texture information that can serve as indirect indicators of physiological status.

While colour index analysis most often involves multispectral bands such as RE or NIR, numerous works demonstrate that RGB indices alone can be highly effective for stress or disease detection. For instance, one study [33] quantified rust disease progression in pea plants using various colour and shape metrics. Among these, the Normalized Green-Red Difference Index (NGRDI) proved most effective for leaflet analysis, achieving over 98% accuracy compared to visual pustule counting. Even when image resolution was reduced to 60-80%, NGRDI maintained high performance. Other indices, such as the Primary Colours Hue Index (HI), also performed strongly, while the Green Leaf Area Index (GLAI) and a* chrominance from the LAB colour space provided moderate but meaningful results.

The Triangular Greenness Index (TGI)[34] estimates leaf chlorophyll content in corn using only RGB bands. Unlike many other vegetation indices that rely on narrow spectral ranges, TGI is specifically designed for broadband digital cameras. Other visible-light indices used in that study include the NGRDI, Green Leaf Index (GLI), and Visible Atmospherically Resistant Index (VARI), with GLI serving as a broadband approximation of the GLAI.

A closely related study [35] developed a manually operated proximal phenotyping cart equipped with two Canon 1300D DSLR cameras, one was a standard RGB camera and one was modified for NDVI imaging to assess drought tolerance in 142 winter wheat genotypes under both field and controlled conditions. Using features such as NDVI, relative growth rate, and RGB-derived biomass, they trained a random forest regression model that achieved $R^2 = 0.54$ when predicting visually scored drought tolerance.

While this demonstrates that low-cost RGB and NDVI imaging can capture drought-related variation, their work focused on qualitative drought scoring rather than direct physiological measurement. In contrast, this thesis investigates the ability of RGB and multispectral proximal imaging to predict *quantitative* LI-600 physiological parameters such as stomatal conductance and fluorescence, enabling direct validation against ground truth sensor data rather than subjective visual ratings.

The growing evidence that RGB-based indices and deep learning regressors can capture physiologically relevant information supports the exploration of RGB imaging as a standalone sensing modality within proximal HTPS. This also provides a logical bridge to multispectral phenotyping, discussed in the following section.

2.3 Multi-Spectral Phenotyping

Multispectral imaging extends beyond the RGB range, providing spectral information that can reveal subtle physiological changes in plants not visible to the naked eye. An example of how this looks in practice is shown in figure 2.3.

In [36], thermal and hyperspectral imaging approaches are shown to effectively monitor plant responses to drought stress. Similarly, [37] demonstrated that hyperspectral imaging estimated photosynthesis performance with an accuracy of 0.8 for shaded leaves and 0.6 for sunlit leaves. They also validated their approach using an LI-600 sensor for ground truth, highlighting the importance of using proximal Porometer/Fluorometer sensors to validate imaging-based predictions.

Recently, a dataset of wheat leaf images with accompanying LI-600 measurements of stomatal conductance and chlorophyll fluorescence was published for the Hyper-Leaf2024 competition [38]. However, datasets combining LI-600 data and RGB imagery for cereal crops are extremely rare. To our knowledge, there is no publicly available paired LI-600 + RGB dataset for oats, and very few for any cereal. Most cereal-focused imaging collections-such as Global Wheat Head Detection or UAV-

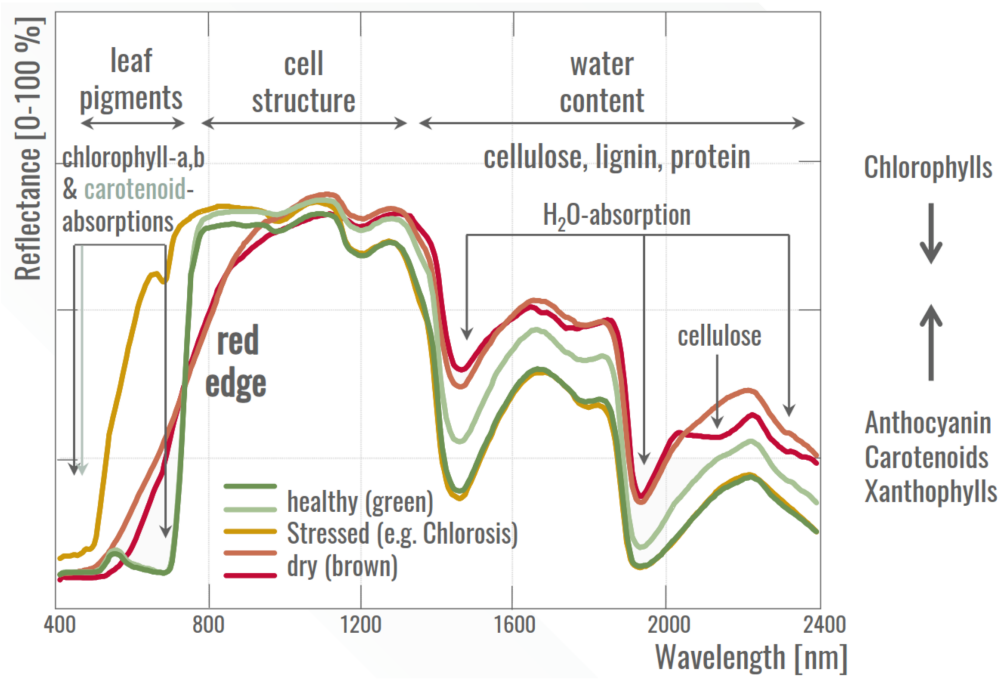

Figure 2.3: Spectral signatures of healthy, stressed and dry plants. Figure adapted from [1]

based wheat canopy surveys-lack localized LI-600 measurements and instead focus on canopy-level traits like head density, yield prediction, biomass, leaf area index, or classification [39, 40]. This scarcity significantly limits opportunities to train and validate low-cost sensing approaches for stress physiology. Paired LI-600/RGB datasets are crucial for developing models that can predict photosynthetic and transpiration traits directly from RGB cameras and in so doing create scalable, cost-efficient field phenotyping.

One study [20] has estimated stomatal conductance (of CO_2) in citrus trees in China based on multispectral UAV imagery and k-nearest neighbour regression. They used a K4 airborne multispectral imager with 560, 650, 725, and 808 nm spectral filters and a grayscale plate for radiometric calibration. Ground truth measurements were collected using a SYS-GH30D photosynthesis analyzer. The eight image texture features evaluated for correlation with photosynthetic parameters of citrus canopy leaves were mean (MEA), variance (VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (ENT), second moment (SEC), and correlation (COR). They also segmented the canopy using HSV filtering for green. Their dataset included 120 samples for model training, achieving $R^2 = 0.92$. The vegetation indices they used are summarized in Table 2.1.

A study in 2021 [21] estimated stomatal conductance of winter wheat using UAV multispectral imaging in China with a DJI M600 UAV and a Tetracam Micro-MCA

Table 2.1: Summary of vegetation indices (VIs) used in the study.

Vegetation Index (with citation)	Formula
Normalized Difference Vegetation Index (NDVI) [41]	$\frac{\mathrm{NIR}-\mathrm{R}}{\mathrm{NIR}+\mathrm{R}}$
Chlorophyll Vegetation Index (CVI) [42]	$\frac{\mathrm{NIR}\cdot\mathrm{R}}{\mathrm{G}^2}$
Red-edge Chlorophyll Index ($\mathbf{CI_{RE}}$) [43]	$\frac{\mathrm{NIR}}{\mathrm{RE}}-1$
Green Normalized Difference Vegetation Index (GNDVI) [44]	$\frac{\mathrm{NIR}-\mathrm{G}}{\mathrm{NIR}+\mathrm{G}}$
Normalized Green-Red Difference Index (NGRDI) [45]	$\frac{\mathrm{G}-\mathrm{R}}{\mathrm{G}+\mathrm{R}}$
Normalized Difference Red Edge Index (NDRE) [46]	$\frac{\mathrm{NIR}-\mathrm{RE}}{\mathrm{NIR}+\mathrm{RE}}$
Ratio Vegetation Index (RVI) [47]	$\frac{\mathrm{NIR}}{\mathrm{R}}$

camera (bands at 490, 550, 680, 720, 800 and 900 nm), achieving ground resolutions between 0.008 m and 0.2 m. They used an LI-6400XT photosynthetic system for ground truth measurements, achieving $R^2 = 0.899$.

Another study [2] predicted stomatal conductance using UAV-based multispectral and thermal infrared imagery in almond orchards in Portugal. They used a DJI P4 multispectral camera (blue 450 nm, green 560 nm, red 650 nm, RE 730 nm, NIR 840 nm) and a DJI Mavic 3T (RGB + Thermal Infrared (TIR), 4-14 μm), with ground resolutions of 3.1-3.5 cm (MS), 7.6 cm (TIR), and 1.9 cm (RGB). The LI-600 was used for ground truth, and they reached $R^2 = 0.87$. These results show that combining multispectral vegetation indices with additional features can effectively estimate porometer-measured stomatal conductance under field conditions. A list of the VIs used in their study is shown in Table 2.1. This approach is well established in crops such as wheat [21], maize/soy, and tree crops such as citrus [20] and almonds [2]. *Note:* GBVI has been replaced with NGBDI [48] as that is the more recognized term for that specific index.

While passive multispectral cameras cannot directly measure active fluorescence parameters like a dedicated fluorometer such as the LI-600, there is precedent for

using multispectral reflectance to predict chlorophyll fluorescence parameters (Fs or Fm') through indirect correlations. A study in 2019 [22], while focusing on hyperspectral indices not reproducible by our setup, found that CI_{RE} (R800/R720, approximately NIR/RE) gave $R^2 = 0.39$ for Fv/Fm and 0.32 for Fv'/Fm' against a validation set in winter wheat. This shows that although the relationship is modest, it is nonetheless meaningful and worth further investigation. While Fs and Fm' are not identical to Fv/Fm or Fv'/Fm', all of these parameters derive from related physiological processes.

2.4 Deep Learning for Phenotyping and Segmentation

While much of the existing phenotyping literature relies on vegetation indices or handcrafted features, this work explores both conventional and deep learning approaches for predicting physiological parameters directly from imagery.

Image regression models have been used to predict numerical plant traits such as area, roundness, and shades of green [55]. Building on this, we hypothesise that numerical physiological metrics from the LI-600 can be predicted in a similar manner using RGB imagery. To test this, we assess three deep learning architectures commonly used for image regression: ResNet-50, EfficientNet, and the Swin Transformer.

ResNet-50 [56] is a widely used convolutional neural network architecture built on residual learning, which allows for the training of very deep models by enabling skip connections across layers. This structure helps mitigate vanishing gradient issues and supports robust feature extraction. ResNet variants remain a strong baseline in image classification and regression pipelines, including in plant phenotyping and detection tasks [55, 57].

EfficientNet [58] is a more recent CNN family that achieves state-of-the-art accuracy through compound scaling of network depth, width, and resolution. Its architecture is known for balancing high performance with relatively low parameter counts, making it a strong candidate for image-based trait prediction tasks in plant phenotyping [55].

The Swin Transformer [59] represents a hierarchical vision transformer architecture that replaces convolution with shifted self-attention windows. This enables scalable and efficient modeling of local and global dependencies in images. Swin has shown

strong performance on general computer vision benchmarks and has proven well-suited for image regression tasks. In agricultural imaging specifically, a Swin variant achieved high accuracy in plant and disease identification benchmarks [60, 40], demonstrating its relevance for plant-related applications.

Other architectures were initially considered (such as VGG16, InceptionV3, MobileNetV2, or Xception), since these have also been successfully applied in plant disease classification and often achieve very high accuracy [57, 40]. However, we decided to focus on the three models above because they encompass both traditional CNN designs and modern transformer-based designs, and each is widely recognized in recent literature. Moreover, prior studies indicate that many state-of-the-art CNN architectures perform similarly. Expanding our experiments to every possible model would likely yield diminishing returns, especially since the goal of this study is not to provide a comprehensive tier list of all deep learning models, but rather to evaluate a representative selection of popular architectures on our dataset.

Phenotyping cereals introduces additional challenges not present in leafy plants. For instance, beetroot and similar crops exhibit radial symmetry, with leaf vectors pointing toward the plant centre [4], but oats and other grasses lean to one side and lack clear morphological distinctions between stems, leaves, and reproductive organs. These structural and visual ambiguities make segmentation, tracking, and trait extraction significantly more difficult.

Central to our approach is the requirement to segment the leaves used with the LI-600 to estimate plant metrics. Existing literature uses methods such as Mask R-CNN [61, 62] or YOLO-based detection models [63], which require large labelled datasets. We found that such models also misclassified the oat spikelet as leaves. Some approaches use 3D point clouds to facilitate leaf instance segmentation [64, 4], but 3D reconstruction of densely packed, constantly moving oat stalks outdoors remains an open problem. Recently, foundation models for segmentation such as SAM 2 [65] have presented an opportunity to achieve zero-shot leaf segmentation without the need for 3D point clouds.

2.5 Yield and Biomass Estimation

A threat to farmer's livelihoods is predation, a study in Poland [66] reported that starlings (Sturnidae) destroyed 22% of cherry crops in 1974 and 16% in 1980. In a broader farmer survey across four districts, of 1476 farmers questioned, about 46% of

respondents confirmed rook (Corvus frugilegus) damage to their fields, with additional farmers suspecting it. Reported bird-related crop losses ranged from 22-32% in corn, 10-13% in wheat, 3-18% in barley, and 8-15% in oats. While there's some work to find predation via leaf damage on leafy plants such as broccoli or detecting caterpillars, there is no existing work containing a dataset that includes visible predation as loss of grain in barley or oats, or even any cereals. A significant portion of crops in our outdoor field dataset show visible signs of predation, introducing a novel category of dataset to the field.

A few approaches were briefly tested for counting grains but ultimately further work is required to count grains accurately. A recent study [67] Introduces a parallel dual-backbone model (ResNet-50 plus an auxiliary CNN), a feature pyramid network for multi-scale features, optimized anchor boxes, and progressive non-max suppression to better separate overlapping panicles. Their enhanced two-stage detector reached 84.4% mAP (mean AP) on oat ear images, which is 13% higher than a standard ResNet-50 Faster R-CNN baseline. No code is available. To our knowledge there is currently no other model available for counting oat spikelets in real field conditions.

Another recent study [68] presents Counting Heads of Cereal Crops Net (CHCNet), which is a unified model designed for counting multiple cereal crop heads by few-shot learning, which effectively reduces labeling costs. Specifically, a refined vision encoder is developed to enhance feature embedding, where a foundation model, namely, the segment anything model (SAM), is employed to emphasize the marked crop heads while mitigating complex background effects. They tested on wheat, rise, sorghum. Code availability is indicated, but the repository link was not accessible at the time of writing.

A 2025 study [69] uses YOLOv5 to detect barley heads at different growth stages in the field, model code and dataset is available publically. MaP at 93.1%, this model warrants testing for barley counting in the future.

T-rex counting [70] is an interactive object counting model designed to first detect and then count any objects. Users can specify the object of interest in a reference image and the model will then detect all objects with a similar pattern in an input image. Applied out-of-the-box to oat panicles, T-rex Counting did not yield dependable counts, indicating a need for domain-specific fine-tuning.

Table 2.2: Additional vegetation and thermal indices used in [2] study (excluding those listed in Table 2.1).

Index (name & abbreviation)	Formula
Blue Normalized Difference Vegetation Index (BNDVI) [49]	$\frac{N - B}{N + B}$
Canopy Chlorophyll Content Index (CCCI) [50]	$\left(\frac{N - RE}{N + RE}\right)\left(\frac{N - R}{N + R}\right)$
Green-Blue Normalized Difference Vegetation Index (GBNDVI) [49]	$\frac{N - (G + B)}{N + (G + B)}$
Normalized Green-Blue Difference Index (NGBDI) [48]	$\frac{G - B}{G + B}$
Green Normalized Value (GN) [2]	$\frac{G}{B + G + R + RE + N}$
Green-Red Normalized Difference Vegetation Index (GRNDVI) [49]	$\frac{N - (G + R)}{N + (G + R)}$
Green-Red Vegetation Index (GRVI) [51]	$\frac{G - R}{G + R}$
Plant Senescence Reflectance Index (PSRI) [52]	$\frac{R - G}{N}$
Red-Blue Normalized Difference Vegetation Index (RBNDVI) [49]	$\frac{N - (R + B)}{N + (R + B)}$
Red-Blue Vegetation Index (RBVI) [53]	$\frac{R - B}{R + B}$
Red-edge Normalized Value (RE_n) [2]	$\frac{RE}{B + G + R + RE + N}$
Red Normalized Value (R_n) [2]	$\frac{R}{B + G + R + RE + N}$
Simple Ratio Pigment Index (SRPI) [54]	$\frac{B}{R}$
Structure Insensitive Pigment Index (SIPI) [54]	$\frac{N - B}{N - R}$

Chapter 3

Design

This chapter details the design of the two High-Throughput Phenotyping Systems introduced in Chapter 1. The first system, described in Section 3.1, was developed for a controlled greenhouse experiment using potted oats to evaluate multi-camera RGB imaging under managed drought treatments. The second system, outlined in Section 3.2, extends this approach to a field configuration designed for natural environmental conditions and incorporates a custom-built multispectral (RGB+RE+NIR) camera.

3.1 Greenhouse RGB High Throughput Phenotyping System

In the first experiment, we aimed to induce water stress on the oat plants and record this using the HTPS and the LI-600. 120 pots were housed in two 8x4 meter poly-carbonate greenhouses with the side walls removed below 1.5 meters. Each greenhouse was equipped with a customised gantry robot based on the Farmbot Genesis XL v1.7, which provided a 6 x 3 x 1.5 meter workspace. To improve control over the platform, we developed an open source ROS2 controller for the robot [71]. The farmbot gantry was also extended by 1 meter vertically to accommodate the height of the oats and the pivoting camera mount.

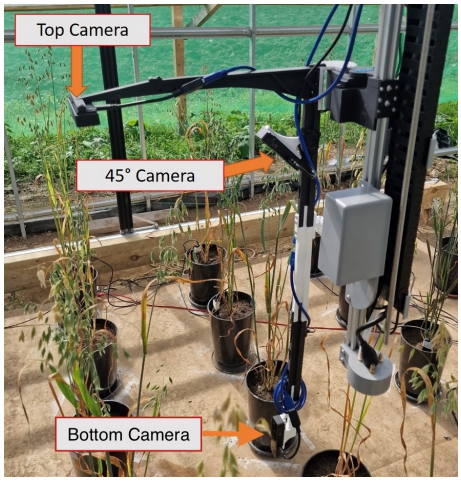

Figure 3.1: Pivoting camera arm for the imaging system.

3.1.1 Imaging Hardware

The imaging system used 3 Luxonis OAK-D Lite AF cameras. These were chosen for their wide field of view, stereo depth and the ability to manually control focus. The cameras are mounted on a 3D printed camera arm pointing top down, at 45°and the bottom camera was mounted horizontally. The 45°camera and the bottom camera were oriented in portrait mode to maximize image capture of the oat plant. The camera arm pivoted through 180°to allow for imaging on two sides of the plant. The camera rig is shown in Figure 3.1.

3.1.2 Agronomic Design

The gantry of the robot moved between the pots to image the plants along one row, and then rotated the camera arm to image the pots from the other side. The camera arm is positioned so that the top camera has a slight offset of circa 10cm between the left and right pass. We label the images top left and right and 45°left and right. The soil experiment, including the preparation of soil treatments and collection of LI-COR measurements, was designed and conducted by collaborators from the Department of Biology at Maynooth University. The experiment was carried out to study the effects of soil microbiology on plant physiology, while the imaging system and all related data processing were developed and operated as part of this thesis. The experiment was performed with oats planted in pots in order to control the growing conditions of each plant in relation to soil biology. The pots were populated with 4 types of soil: Soil - Control (SC), Autoclaved soil (A), Autoclaved soil with microbial inoculant from a conventional farm (AC), and Autoclaved soil with a microbial inoculant from

an Organic farm(AO). Additionally, each of these had a varied amount of organic carbon in the soil by the addition of Coconut Fibre (C.F.). Table 3.1 shows the breakdown of the different soil treatments in the experiment.

Table 3.1: Experimental soil treatments - Control (SC), Autoclaved soil (A), Autoclaved soil with microbial inoculant from a conventional farm (AC), and Autoclaved soil with a microbial inoculant from an Organic farm(AO) and the percentages of additional organic carbon added to the soil, and drought conditions. Note: $i \in [1:5]$.

Soil	**Normal**			**Drought**		
C.F.	**Control**	**+5%**	**+10%**	**Control**	**+5%**	**+10%**
AC	N-1-i	N-2-i	N-3-i	D-4-i	D-5-i	D-6-i
AO	N-7-i	N-8-i	N-9-i	D-10-i	D-11-i	D-12-i
SC	N-13-i	N-14-i	N-15-i	D-16-i	D-17-i	D-18-i
A	N-19-i	N-20-i	N-21-i	D-22-i	D-23-i	D-24-i

3.1.3 LI-600 Data Acquisition

The LI-600 was used to record the data of the top-most leaf. These measurements were synchronised with the RGB images for later correlation.

Data were collected over several days towards the end of the drought phase of the experiment, allowing water stress to be visible. For each date, all LI-600 measurements and camera metrics data were cleaned to remove any corrupt LI-600 measurements or pots that were missing readings.

All of these measurements are recorded and tabulated based on time, date and pot code. Each row in the data generated by the RGB metric extraction contains 9 RGB indices and 13 index statistics, resulting in 117 RGB metrics that we tested against 6 LI-600 metrics, for a total of 702 data pairs per row.

3.1.4 Dataset Structure and Analytical Design

The greenhouse RGB experiment originally comprised 120 potted oat plants (60 drought-stressed, 60 well-watered). However, due to technical and logistical constraints, the full cohort was not available for joint imaging and LI-600 analysis.

Imaging runs could only be completed in one greenhouse compartment, as the FarmBot system in the second greenhouse malfunctioned during the experimental period. Additionally, LI-600 measurements are time-intensive, limiting the number of days on which synchronized physiological recordings could be collected.

Final Regression Dataset The final dataset used for correlation and deep learning regression consisted of 46 individual plants sampled across three experimental days, one of which included two imaging sessions (morning and evening). For each plant per session, four RGB images were captured, comprising two angled views $\pm 45^{\circ}$ and two slightly offset top-down views.

After filtering for synchronised LI-600 pairings and quality control, the dataset was organised by camera viewpoint. A total of 334 image–measurement pairs were retained for each primary viewpoint:

- 334 top-view RGB images with corresponding LI-600 measurements
- 334 angled-view RGB images with corresponding LI-600 measurements

These viewpoints were analysed both independently and in combination, with only synchronised image–LI-600 pairs retained for regression and correlation analyses.

The following LI-600 metrics were used as regression targets:

- Stomatal conductance (gsw)
- Leaf vapour pressure
- Leaf vapour pressure deficit
- Leaf H_2O mole fraction
- Steady-state fluorescence
- Maximum fluorescence

Model performance for all prediction approaches was evaluated using R^2, computed on both a representative morning subset and the full experimental period. R^2 expresses the proportion of variance in the physiological trait explained by the model and enables scale-independent comparison across different LI-600 metrics. Because R^2 corresponds to the square of the Pearson correlation coefficient, it also allows direct comparison between deep learning regression and RGB index-based correlation results.

Image Inputs and Segmentation Targets The comparison between whole-plant and top-leaf segmentation directly tests whether isolating the leaf measured by the LI-600 improves predictive performance, or whether broader plant context provides additional informative signal for deep learning models.

- Whole plant segmentation
- Top-leaf-only segmentation

Similarly, the comparison between top, angled, and combined views evaluates which camera orientation provides the most informative representation for physiological inference.

- Top-down (0°)
- Angled (45°)
- Top-down and Angled Combined

RGB Index-Based Feature Structure For the correlation-based analysis, nine RGB vegetation indices were computed per image, including NGRDI, TGI, VARI, HI, a* chrominance, GLI, and raw R, G, B values.

For each index, thirteen descriptive statistics were calculated (mode, mean, median, max, min, standard deviation, variance, skewness, kurtosis, range, Q1, Q3, IQR), resulting in:

- 9 indices × 13 statistics = 117 RGB-derived features per sample
- Each evaluated against 6 LI-600 metrics
- 702 feature–physiology pairs per row

This pathway evaluates whether physiological traits can be described using interpretable colour-derived statistics alone, without end-to-end deep learning.

Segmentation and Classification Dataset Top-leaf segmentation for the deep learning pathway was supported by a labelled dataset comprising:

- 546 manually annotated instances

Figure 3.2: Drone Photograph of the outdoor-field HTPS.

- 344 SAM2-derived segmentation masks
- Total labelled instances: 890

Class distribution:

- Leaf: 606 labels
- Panicle: 1632 labels
- PanicleClump: 54 labels
- Noise: 114 labels

Data augmentation included horizontal flipping, 90° rotations, saturation (±25%), and exposure (±10%), with three augmented outputs generated per training example. The dataset was split into training, validation, and test sets using a 70/20/10 ratio plant-level split.

3.2 Outdoor Multispectral High Throughput Phenotyping System

Following the controlled greenhouse trials, we extended the HTPS for outdoor deployment to evaluate system performance under field-like variability. To better

approximate a realistic cereal crop environment, we adapted the HTPS to an outdoor-field (OF) configuration with a single overhead camera above crops planted in the soil in traditional rows with 15cm intra-row spacing and 30cm inter-row spacing. The oat plants were grown at a higher density to simulate commercial cereal cropping better. The HTPS traversed the beds in a raster pattern, capturing continuous top-down RGB video at $\tilde{5}$ frames per second (fps) using a single camera. This simplified video-only configuration removed the complexity of multi-angle imaging, while more accurately reflecting the constraints of real-world deployment. A top-down view of the outdoor-field system is shown in Figure 3.2.

3.2.1 Imaging Hardware

The raised-bed platform was based on a Farmbot Genesis XL gantry robot (6 m × 3 m workspace). Position was logged via ROS at ∼1 Hz and interpolated to each frame of image capture. This provided spatial resolution of approximately 2 cm, enabling accurate indexing of individual plants.

A Luxonis OAK-D Long Range camera was mounted overhead on the gantry. The left and right stereo sensors were retrofitted with narrow band-pass filters: a red-edge (RE) filter centred at 730 nm (40 nm FWHM) applied to the left camera and a near-infrared (NIR) filter centred at 850 nm (40 nm FWHM) applied to the right camera. The centre sensor retained its standard IR-cut filter, providing RGB output. This configuration produced triplets of RGB, RE, and NIR at ∼5 fps, streamed via ROS 2 from a Raspberry Pi 5.

3.2.2 Agronomic Design

Oats and barley were planted in a 15 × 19 grid, with 15 cm spacing between plants within rows and 30 cm spacing between rows, covering a total area of approximately 5.4 m × 2.1 m. Data were collected every two days from June 3 to June 29, 2025, with 38 plants sampled per session. A heat map visualising sampling frequency per-plant is shown in Figure 3.3.

3.2.3 LI-600 Data Acquisition

Physiological ground truth was obtained using the LI-600. Measurements were taken on the uppermost fully expanded leaf, with the same measurements as before: g_{sw},

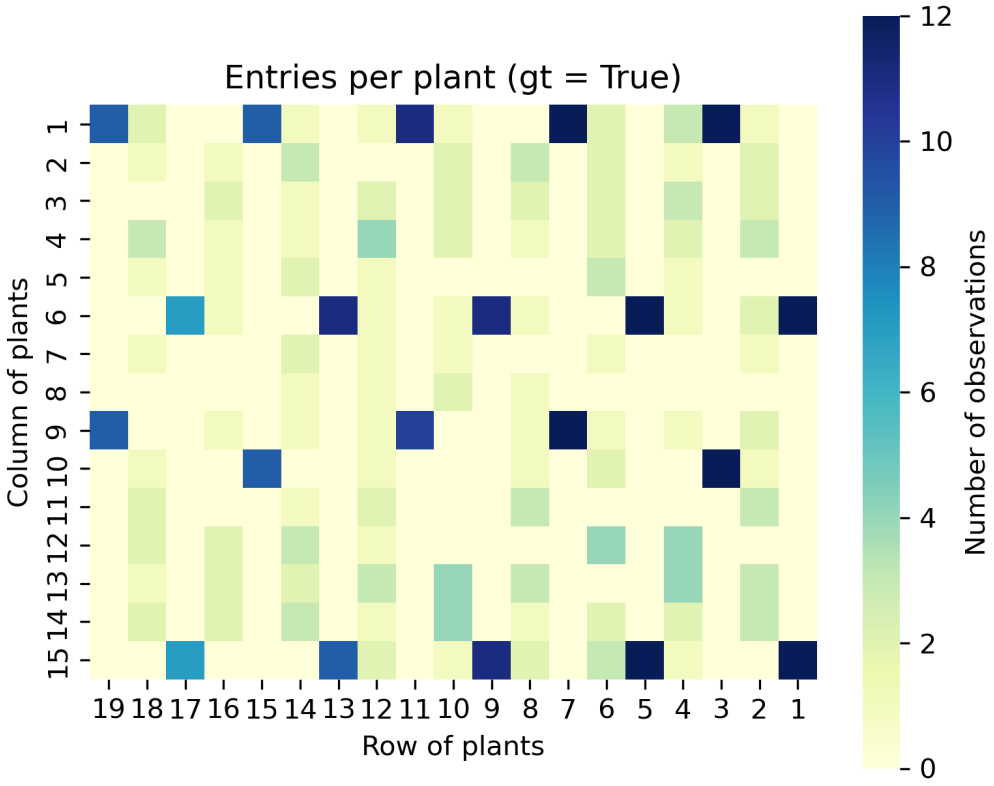

Figure 3.3: Heatmap showing LICOR LI-600 sampling frequency per plant

VPD_{leaf}, VP_{leaf}, H_2O, Fs and Fm′.

Sampling followed a hybrid constant/semi-random approach. Twenty plants were sampled consistently throughout the entire period to capture purely temporal changes, while the remaining eighteen were selected semi-randomly to expand dataset diversity. Two plants were chosen from each even-numbered row for constant sampling, evenly spaced along the row. In each odd-numbered row, two plants were selected semi-randomly-one from the top half and one from the bottom half-to maintain a balanced spatial distribution. The semi-random selection was performed manually rather than using a randomizer. Randomly selected samples did not repeat within the same week but could reappear between weeks. Using the gorund truth samples, readings for other plants were interpolated and used as synthetic data, described below.

3.2.4 Statistical and Machine-Learning Analysis

Dataset Structure

Data collection occurred between June 3 and June 29, 2025, at two-day intervals, after filtering missing data due to mechanical issues and data corruption, this resulted in

12 total imaging sessions. After additional data filtering and organisation described in 4.2.3 this produced:

- **390** plants with LI-600 measurements.
- **2827** plant-level RGB+RE+NIR triplets.

Split Design. Compared to the earlier RGB-only experiment, the raised-bed multispectral setup introduced three new sources of variation: spectral modality (RGB vs. RGB+MS), plant species, and the method of accessing LI-600 ground-truth data (direct vs. synthetic). To evaluate the effect of these factors while avoiding spatial leakage, four complementary dataset definitions were created:

- **Split A:** A purely ground-truth (GT) dataset using directly measured LI-600 values. A 60/20/20 train/validation/test split was enforced by partitioning Plant IDs (not rows), so all readings (dates/frames) from a given plant were kept together and never split across folds. To preserve the distribution of per-plant sample sizes, plants were first sorted by their row counts and divided into four row-stratified bins; within each bin, plants were randomly selected in proportion to the bin's total rows to meet the target row budget, yielding a split that matches the global row distribution while avoiding plant overlap between splits.
- **Split B:** A half-GT configuration where approximately 50% of the plants in the training and validation sets used ground-truth data, and the remaining half were filled with synthetic samples. Synthetic values were selected to match the per-plant row counts of the GT subset, preserving identical data distribution and ensuring no overlap between GT and synthetic plants across splits.
- **Split C:** A fully synthetic variant where all training and validation samples were replaced with synthetic values matched one-to-one to the original GT plant structure (i.e., same number of rows and plants per split). This design isolates model behaviour when trained solely on synthetic physiological data.
- **Split D:** A combined configuration where all available GT samples were retained, and an equal number of synthetic samples were added to the training and validation sets. This doubles the dataset size relative to the other splits, enabling the model to leverage both real and synthetic data.

Each of these four dataset definitions were generated not only for the combined dataset (oats and barley together) but also independently for each crop species, maintaining identical proportions and grid-level exclusion logic.

Adjacency filtering. Because each plant image may include portions of neighboring plants, training on plants spatially adjacent to the test set could lead to visual data leakage. To mitigate this, any plant directly neighboring a test plant in the raised-bed grid (8-neighborhood) was excluded from the training and validation sets for the same date. This ensured that test plants remained completely unseen from adjacent viewpoints. Applying this filter slightly adjusted the dataset balance, reducing the Split A ratio from the initial 60-20-20 to 54-21.6-24.4. All subsequent splits (B-D) reused this filtered baseline to maintain consistent exclusion across experiments. As all splits shared the same test set, no additional adjacency filtering was required beyond this initial step.

Spectral modality. Finally, each split was evaluated under both available imaging modalities: (i) standard RGB, and (ii) multimodal RGB+MS (RGB plus co-registered multispectral bands).

Total variants. This systematic expansion yielded 24 dataset variants in total: four dataset definitions $\times$ three species configurations $\times$ two imaging modalities $=$ 24.

Chapter 4

Implementation

This chapter details the practical implementation of the HTPS described in Chapter 3.

4.1 Greenhouse RGB HTPS

4.1.1 Design to Implementation differences

Beyond the seedling stage the bottom camera was no longer able to see the top leaf so for this study we used only the top and 45°camera. The Oak-D Lite has two monochrome cameras with on-camera disparity map calculation. However, we observed significant errors from the built-in disparity map. The next method for depth calculation tested was depth from focus but wind motion between frames corrupted the data. We therefore adopted the zero-shot Depth Anything approach [72].

Upon testing, we found that the camera's focus has a significant impact on the accuracy of depth estimation, and as a result, we disabled auto focus and manually set the camera's focus to different depth points. The camera lenses were manually calibrated using a UR5 robotic arm and a checkerboard target to achieve consistent focus across depth stacks. Calibration involved incrementally adjusting the camera's focus for optimal sharpness at distances ranging from 10 cm to 70 cm, resulting in a set of five discrete focal distances. The best focal distance was selected by evaluating both the sharpness of the RGB image and the quality of the depth map generated by Depth Anything [72]. Although a distance of 55 cm produced clearer depth maps, the sharpest RGB image-prioritized for accurate leaf imaging-was obtained at 70 cm.

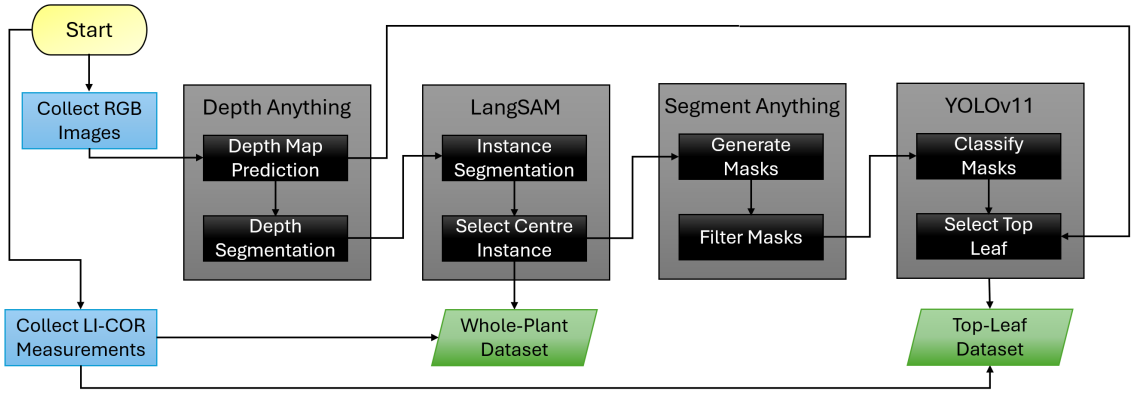

Figure 4.1: Pipeline for whole-plant image analysis. Depth estimation with Depth Anything and depth-based segmentation isolates the plants from the background. LangSAM is used for instance segmentation, the centre plant is selected as the target. This is used to create the Whole-Plant dataset. SAM2 then generates segmentation masks, which are classified into plant organs using a fine-tuned YOLOv11 model. The top leaf, defined as the leaf closest to the camera, is identified to enable comparison between whole-plant metrics and single-leaf physiological measurements obtained with the LI-600 sensor.

4.1.2 Image Processing

This pipeline first uses depth estimation with Depth Anything [72] and depth-based segmentation to isolate the plants from the background. To further refine the region of interest, we employ Lang-Segment-Anything (LangSAM) [73], a language-driven segmentation framework that combines SAM [74] with the grounding capabilities of GroundingDINO [75]. Using the prompt "oat plant", this model can segment instances of oat plants with high reliability. Subsequently, SAM2 [65] is applied to generate detailed organ-level segmentation masks, which are classified using a fine-tuned YOLOv11 classifier [63]. The pipeline flow chart is shown in Figure 4.1 Each step of the pipeline is visualised in Figure 4.2, and described below.

Background removal via Depth-Anything

The depth segmentation step was challenging due to variability in the absolute depth measurements from Depth Anything, making fixed depth thresholds unreliable. A custom tool was developed to set thresholds dynamically.

Plant Isolation via LangSAM

Each RGB image is first passed through the LangSAM framework [73]. Using the prompt "oat plant", LangSAM identifies all candidate masks, and the mask whose

centroid lies closest to the image centre is selected as the target plant. Occasionally, the mask produced by LangSAM is the inverse of the desired mask (ie, it masks the entire background minus the plant itself). To counter this, the overlap between the mask and the RGB content is computed. If less than 10% of the mask overlaps with the nonzero pixel regions of the image, the mask is rejected and the original image is retained for downstream segmentation (This bug only occurred on images containing 1 plant). Otherwise, the selected mask is applied to the image to isolate the central plant from neighbouring pots and background. The outputs of this step are used as the whole-plant dataset and are also fed into the next stage of the pipeline for detecting the top leaf.

Organ Segmentation with SAM2

To identify meaningful substructures of the isolated plant, we apply SAM2 [65]. The SAM2 model is `sam2.1_hiera_s`, configured with `points_per_side=64`, `pred_iou_thresh=0.80`, `stability_score_thresh=0.90`, and a minimum region area of 100 pixels to suppress noise. Since SAM2 may also segment soil or background clutter, we apply a color-based filter in the HSV color space to retain only masks corresponding to green tissue. Specifically, masks are retained if the mean hue lies in the range [5, 85], and either the value channel exceeds 90 or both saturation and value exceed 15 and 45, respectively. Overlapping masks are post-processed to eliminate duplicates by subtracting smaller masks from larger masks, then deleting masks with more than 90% area loss.

Organ Classification with YOLOv11 (Training and Inference)

YOLO was selected as the classification backbone due to its favourable trade-off between accuracy and efficiency, and strong tooling support. At the time of development, YOLOv11 was the most recent stable release and was reported to offer improved efficiency and performance over earlier versions [63].

Training data and augmentation. An initial dataset of 546 instances of leaves and spikelets (labelled as `Panicle`) was annotated manually in Roboflow. The dataset was split using a 70/20/10 train/validation/test ratio and expanded using three augmented outputs per training example with randomized transformations: horizontal flip, 90° clockwise and counter-clockwise rotations, saturation adjustment in the range ±25%, and exposure adjustment in the range ±10%. A YOLOv11x

model was trained on this dataset.

During integration with the segmentation pipeline, systematic domain mismatch was observed between hand-drawn annotations and SAM2-generated masks: SAM2 occasionally grouped multiple spikelets into a single region (`PanicleClump`), and the pipeline also produced occasional non-plant masks (e.g., pot segments) that were not represented in the initial training set. To address these failure modes, a further 344 masks were curated directly from SAM2 outputs, targeting spikelet clumps and background artifacts, and used to refine the classifier. After this second-stage training, the final test accuracy was 0.943.

Inference on SAM2 masks. At runtime, each filtered SAM2 mask is mapped back to the RGB frame, cropped to its bounding box, and resized (with padding) to 640×640 for classification. The classifier assigns one of four classes: `Leaf`, `Panicle`, `PanicleClump`, or `Noise`. If a region is predicted as `Leaf` with confidence below 0.99, it is reassigned as `Panicle` to reduce under-confident leaf assignments.

Across the labelled SAM2-derived evaluation set, the class distribution was: `Leaf` (606), `Panicle` (1632), `PanicleClump` (54), and `Noise` (114). Based on the confusion matrix, per-class precision and recall on the test set were: `Leaf` (precision = 0.93, recall = 0.93), `Panicle` (precision = 0.97, recall = 0.99), `PanicleClump` (precision = 1.00, recall = 0.71), and `Noise` (precision = 0.91, recall = 0.77).

Top Leaf Selection using Depth

To identify the topmost leaf, we analyze depth images aligned to the RGB frame. For each mask labeled as `Leaf`, we compute the maximum pixel intensity in the masked depth region (where brighter pixels are closer to the camera). We retain the top 10% of leaf masks based on this maximum depth value and compute the area of each. From the three largest masks, the one with the highest maximum intensity is selected as the top leaf. This helps filter out spikelets that were misclassified as leaves and would be taller than the top leaf.

The resulting masks were used to generate the whole-plant and top-leaf image subsets for downstream regression and correlation experiments.

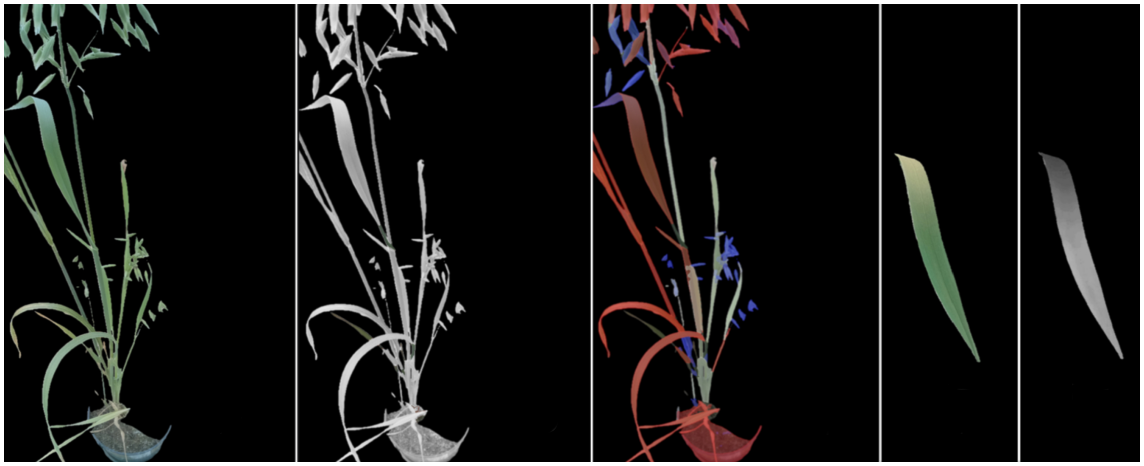

Figure 4.2: Visualization of top-leaf segmentation and RGB-index retrieval.(A) Background removal and centre plant selection using Depth Anything and LangSAM; (B) Organ segmentation with SAM2; (C) Classification of segmented regions as leaf or spikelet (labelled as `Panicle` in YOLOv11); (D) Top-leaf selection based on depth values from Depth Anything; (E) HI-index extraction.

4.1.3 Image Regression and RGB analysis

Both the segmented top leaf and the whole plant targets were used as inputs for evaluating the performance of multiple image regression models and RGB metric analysis. In addition to the pipeline outputs, images from a single day were manually segmented by a human for both the whole-plant and top-leaf targets to validate the effectiveness of the segmentation pipeline. We chose nine standard RGB metrics: TGI, NGRDI, VARI, HI, a* and GLI, along with the raw red, green and blue values. We also calculated thirteen descriptive statistics: Mode, Mean, Median, Max and Min to summarise the central tendency and extreme values of the data, standard deviation and variance to capture the spread and variation, skewness and kurtosis to measure the shape of the distribution, range depicts the spread in its simplest form, the first quartile and third quartile provide insight into the distribution's spread around the median and finally the inter-quartile range is useful for depicting the spread while being more robust to outliers.

4.2 Outdoor Multispectral HTPS

Building on the controlled greenhouse trials, the next stage extended the HTPS to a natural outdoor environment. This transition required both mechanical and computational adaptations to handle denser canopies, variable illumination, and larger spatial scales.

4.2.1 Design to Implementation Differences

Initial Split Design

Earlier interpolation-based experiments used four broader splits that included all interpolated LI-600 data in training, validation, and testing. While plants were correctly separated by ID and adjacency-filtered, this setup exposed the model to a nearly complete spatial map of interpolated values. Because interpolation creates smooth gradients across the raised bed, models trained under this configuration could implicitly infer spatial position rather than learning physiological relationships, leading to unrealistically high R^2 values. The revised split structure adopted here addresses this by restricting access to only the same sparse ground-truth locations used in the LI-600 measurements (approximately 15% of the full dataset). This ensures that synthetic data reflect the limited sampling density of the original physiological measurements and prevents spatial overfitting.

Multispectral camera implementation

To ensure geometric consistency across sensors, a full stereo calibration was performed using a chessboard procedure, rasterising the cameras onto a common epipolar plane and removing fisheye distortion. All three sensors were manually focused to infinity, yielding a minimum focus distance of approximately 50 cm.

Because the OAK-D "mono" mode introduced a visible screen-door artefact, all sensors were initialized as colour cameras. The RE and NIR streams were then converted to grayscale. Standard OpenCV conversion (`cv2.cvtColor(..., COLOR_BGR2GRAY)`) applies fixed channel weights ($0.299R + 0.587G + 0.114B$), which produced a dim image since the green and blue pixels contributed only noise under the fitted filters. Instead, the red channel was isolated directly, providing a cleaner signal. It should be noted, however, that the Bayer red filter has a broad spectral tail and therefore behaves as a non-ideal band-pass, with partial leakage into the near-infrared. This spectral response may attenuate or distort true RE and NIR signals, meaning that raw index values should not be interpreted at face value. In practice, this can be addressed through gain calibration or by training models directly on the raw responses, which is the approach adopted in this study.

One limitation of the system was that the three camera streams were transmitted asynchronously as ROS topics, leading to frame-to-frame de synchronisation. This made stereo depth reconstruction unreliable. While not critical for the present work

(which relied on per-frame multispectral indices), future iterations should implement hardware-level synchronisation to enable robust depth estimation.

4.2.2 Image Processing Pipeline Adaptations

For the outdoor row crops, the greenhouse segmentation pipeline (Depth Anything + LangSAM + SAM2 + YOLO) proved unreliable due to the significantly denser canopy. The overlapping leaves and lack of clear plant boundaries caused instance segmentation to fail frequently, producing merged or incomplete masks. To ensure a consistent visual input, the segmentation step was replaced with a simple window cropping: for each selected frame, a fixed-size crop centered on the plant position was extracted, including any background details. This method guaranteed that the target plant was fully visible while maintaining consistent framing across all samples. These cropped images formed the input for both spectral index analysis and the end-to-end learning experiments. In contrast to the greenhouse HTPS. The pipeline for this system focuses on data cleaning and organization rather than segmentation. The flow chart for creating the finished dataset is shown in Figure 4.3

4.2.3 Data Organization

Coordinate synchronization and frame-plant association. Initially, An automated alignment procedure was implemented to synchronize the Farmbot's 1 Hz position logs with the imaging system's 5 Hz frame rate. Cubic splines were fitted to the raw (x, y) coordinate stream to interpolate robot positions at the exact timestamps of each frame, while a constant spatial offset and velocity-dependent correction compensated for camera mounting displacement and reporting lag. Using these corrected coordinates, each frame was automatically associated with the nearest plant in a predefined reference map of the raised bed. Typically, two to five frames were captured per plant, with the central frame selected as representative for analysis. Each image was also tagged with its corresponding crop species (Husky oat or Adalina barley), enabling species-specific comparisons in later experiments.

Outdoor dataset challenges. However, this approach proved insufficient for the outdoor raised-bed experiment. Upon closer inspection, the corrected coordinates still exhibited irregular jitter and positional drift. This behaviour was traced to intermittent packet loss and timing jitter in the Farmbot's communication stream, likely caused by the weak Wi-Fi connection between the outdoor Farmbot and the

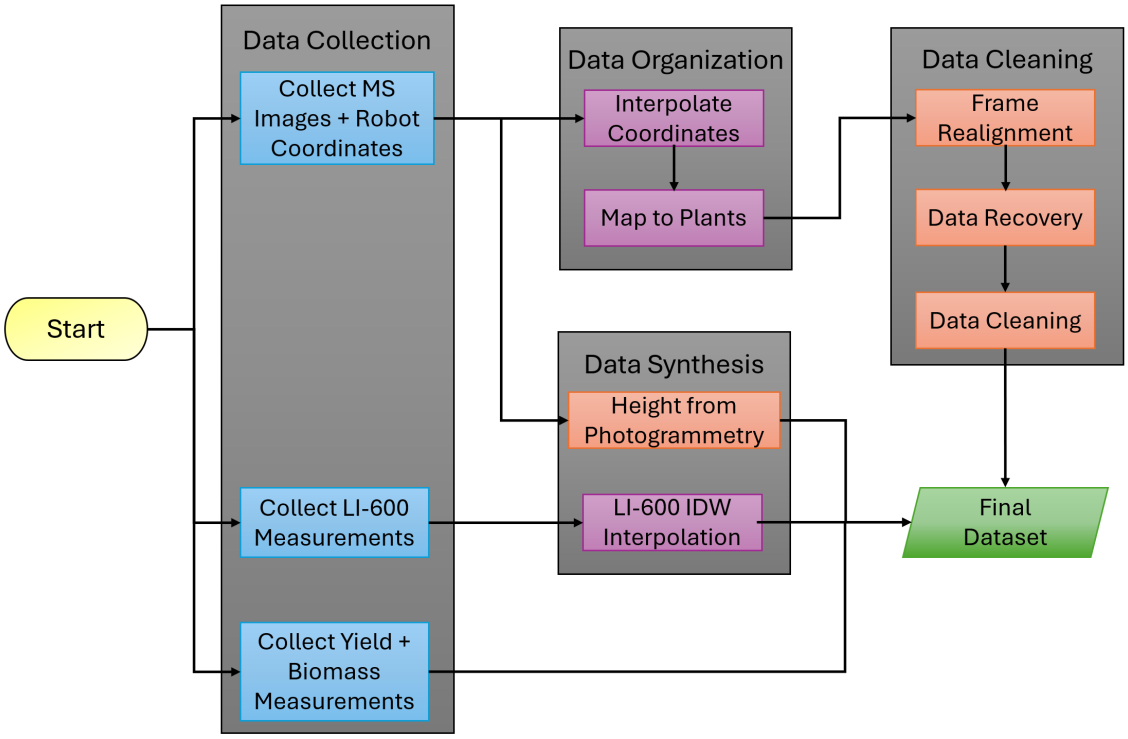

Figure 4.3: Pipeline for Multispectral Outdoor System. Multispectral images and robot coordinates are first collected and interpolated to align each frame with the corresponding plant. Manual realignment, recovery of missing sessions, and cleaning steps ensure spatial and temporal consistency. Additional data synthesis includes photogrammetric height reconstruction and LI-600 interpolation using inverse distance weighting (IDW). The final dataset combines validated imaging, physiological, and yield measurements for downstream analysis.

base station. Unlike the indoor HTPS system, where the router was located in the same greenhouse, the outdoor setup required the Farmbot to transmit over a weaker network.

Network reliability and mitigation. The signal remained adequate near the gantry's home position at coordinates (0,0) but degraded substantially as the Farmbot moved up to 6 m away, leading to delayed or dropped packets. Although the issue mostly began to manifest beyond the midpoint of the bed, it was not evident during the initial full run. The Farmbot completed its planned trajectory without visible interruption, and a brief review of the footage confirmed full spatial coverage. However, subtle frame buffering caused by intermittent packet delays occurred throughout the dataset. While not visually noticeable in real time, this resulted in timestamp misalignments that only became apparent during later data processing, when image frames and reported coordinates failed to align consistently.

In addition, the Farmbot occasionally skipped waypoints in its planned path, sometimes moving diagonally between non-adjacent targets or missing entire plant rows. These deviations were strongly correlated with distance from the greenhouse router, confirming network signal degradation as the underlying cause.

This experience highlights a key limitation of network-dependent imaging in field robotics: real-time streaming introduces reliability and synchronization challenges in weak-signal environments. Future systems should favor on-board storage for image capture, with data offloaded to a NAS or cloud server post-operation, eliminating the dependency on continuous high-bandwidth connectivity.

Manual correction. In the final dataset, these issues were mitigated through extensive manual correction and cleaning. This involved frame-by-frame visual realignment of images to their corresponding plants, the recovery and integration of new imaging dates (06-05, 06-07, 06-25) where Farmbot coordinate logs were corrupted, and the removal of incomplete or faulty entries. The resulting dataset represents a fully validated and spatially consistent version of the raised-bed experiment, with reliable correspondence between imaging data and physiological measurements across all plants.

Band registration. Finally, lightweight 2D registration was applied to align the red-edge (RE) and near-infrared (NIR) bands with the RGB reference. This was done by simply applying constant horizontal pixel offsets of –100 px for the RE band and

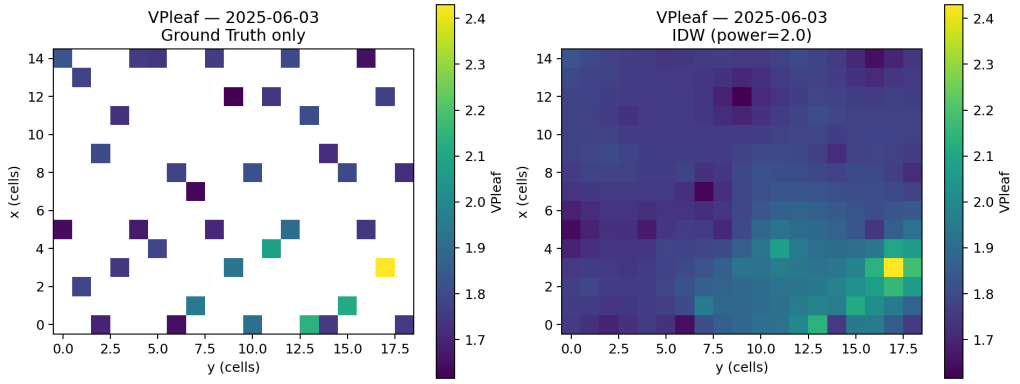

Figure 4.4: Example of IDW LI-600 parameter interpolation for VPleaf - Left) Ground truth readings, Right) Interpolated values

+200 px for the NIR band. The offsets were determined using a calibration approach by visually aligning overlapping image regions, corresponding approximately to the 5 cm and 10 cm standoffs between cameras at an imaging height of about 1 m. Other methods such as SIFT-based feature matching were tested but proved ineffective due to asynchronous frame capture between the cameras. Traditional stereo alignment tools were also unhelpful, as the asynchronous triggering introduced a sporadic vertical misalignment between frames that was non-trivial to correct. Therefore, a coarse alignment approach was chosen, as it was sufficient for global channel-mean spectral calculation, effectively mimicking a single-pixel response from a remote sensing system.

4.2.4 Data Processing

LI-600 Data Interpolation

After associating images with plant positions, missing LI-600 readings were interpolated. For data interpolation inverse distance weighting (IDW) was used across the grid. Looking at the sampled values such as gsw as a scalar value in each pixel, we assign a value to every empty pixel that is the weighted average of all sampled values, with weights proportional to $1/\text{distance}^2$. Initially some other approaches were tested such as ordinary kriging, spatio-temporal kriging and neighbour propagation, but IDW seemed to return the most consistent results. An example of IDW interpolation is show in Figure 4.4

Height Data

In theory, the stereo camera setup should have been capable of retrieving metric depth information from the plant canopy. However, based on experience with the Greenhouse RGB HTPS, stereo depth quality is known to degrade significantly at these imaging distances, and this effect would likely be exacerbated by the wider baseline of the LR stereo pair used here. The original plan was to employ NVIDIA's feature-stereo pipeline to combine AI-based disparity estimation with traditional stereo depth, thereby producing smooth, high-resolution metric depth maps similar to those generated by Depth Anything but with true scale. Unfortunately, this approach proved infeasible because asynchronous frame retrieval during motion prevented consistent stereo matching.

To obtain any usable depth information for analysis, a photogrammetric approach was adopted using Meshroom. All images from the final-day raster scan were processed to generate a sparse 3D point cloud of the raised bed. This reconstruction was imported into CloudCompare, where it was projected along the z-axis to create a depth map. The resulting height field was downsampled into a 15×19 grid to yield a single representative canopy height per plant. While this method is far from ideal, it serves as a practical substitute for true metric depth, included primarily to illustrate the potential value of canopy height information in downstream phenotyping analyses.

Yield and Biomass Data

Yield and biomass were sampled per-plant, by harvesting the grain and straw of each plant, and placing into labelled envelopes one by one. After drying in a 70 degree oven for 72 hours, each envelope was weighed and a final dry weight of biomass and yield labelled for each plant. The thousand grain weight (TGW) was calculated by counting out 1000 grains and weighing them, which came out to a hulled TGW of 34g for oats. Visualisations of biomass yield and height are shown in Figure 4.5

4.2.5 Statistical & Machine-Learning Analysis

Extracting RGB + NIR + RE Colour and texture Metrics

Building on the greenhouse RGB-index pipeline, the outdoor analysis retained the core RGB indices (e.g., NGRDI, TGI, VARI, GLI) but omitted HI and a^* chrominance, and introduced multispectral indices enabled by the RE and NIR bands (NDVI,

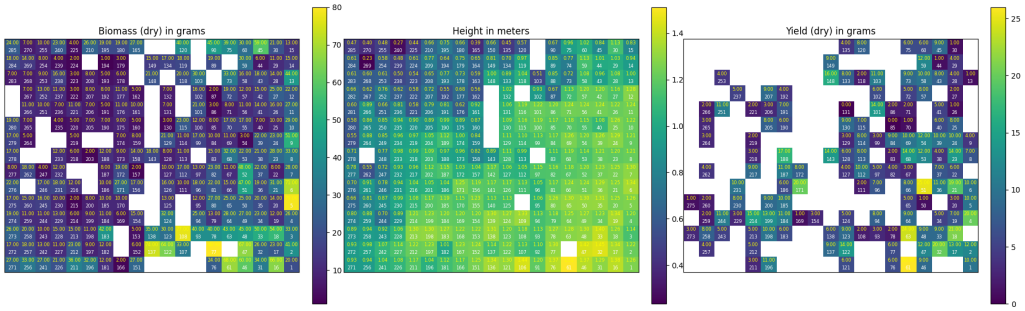

Figure 4.5: Visualisation of collected Biomass, Height and Yield data in the outdoor raised bed.

NDRE, GNDVI, and CI_{green}). The complete RGB and multispectral index lists are provided in Tables 4.1 and 4.2, respectively.

A key methodological change in this experiment was that indices were computed using only the *global channel mean* (one value per band per image), rather than the full suite of descriptive statistics previously extracted. This effectively reduces each plant frame to a single "remote-sensing style" pixel response. This choice was motivated by three factors: (i) it reduces sensitivity to residual band misalignment artefacts that can introduce hard edges and distort higher-order statistics; (ii) it provides a sanity check on the multispectral hardware, since at this coarse representation the RE/NIR bands should provide clearer physiological signal than RGB alone; and (iii) prior greenhouse results suggested that higher-order statistics (e.g., standard deviation, skewness and kurtosis) rarely added predictive value compared to just the mean.

In addition to vegetation indices, eight standard texture metrics were extracted from the gray-level co-occurrence matrix (GLCM) for both the green and NIR channels: contrast, dissimilarity, homogeneity, energy, angular second moment, correlation, entropy, and variance. These were computed using the scikit-image implementation with distances of 1, 2, 4, and 8 pixels and four directional angles (0°, 45°, 90°, 135°). The matrices were normalised and symmetrised before feature extraction. For each metric, values were calculated separately at each distance and then averaged across all distances to produce a single representative texture descriptor per channel. Together, these features capture both the spectral and structural characteristics of the plant canopy and were used as inputs to the regression models (linear regression and random forest). For each plant, the multispectral dataset included eight RGB

indices, 22 multispectral indices, and eight texture metrics, all paired with six physiological traits measured by the LI-600. This produced $38 \times 6 = 228$ pairs of numeric feature–target combinations that are associated with each cropped image.

End-to-End Learning

Three convolutional architectures-ResNet-50, EfficientNet-B5, and Swin-Large (patch4-window12-384)-were trained across all 24 dataset variants in both RGB (3-channel) and RGB+MS (5-channel) configurations. Each model was implemented as a multivariate regressor jointly predicting seven physiological and agronomic targets (*gsw*, *VPleaf*, *VPDleaf*, *Fs*, *Fm′*, *yield*, and *biomass*).. Images were centre-cropped to 384×384,px. This resolution matched the input size of the largest Swin Transformer model and comfortably framed the target plant, leaving enough padding to capture the entire canopy while erring on the side of including a few neighbouring leaves. Random horizontal and vertical flips, as well as rotations of up to $\pm 15°$, were applied during training. Inputs were normalized per channel (mean = 0.5, std = 0.25).

Training employed a masked, variance-weighted RMSE loss that ignored missing targets and weighted each term by the inverse of its population variance. The AdamW optimizer (weight decay 10^{-4}) was used with fixed learning rates determined empirically for each model: 3×10^{-3} (EfficientNet), 7×10^{-4} (ResNet), and 2×10^{-5} (Swin). A dynamic batch-size finder automatically maximized GPU memory utilization, except for Swin, which was capped at half the discovered limit to prevent memory exhaustion. Each model was trained for 75 epochs, with the lowest validation loss checkpoint used for testing.

Evaluation metrics included per-target RMSE and R^2, as well as the mean R^2 across all nine outputs. A looping script automated the full experimental matrix (4 dataset definitions $\times$ 3 species configurations $\times$ 2 imaging modalities $\times$ 3 models), yielding 72 training runs in total. All results were logged to TensorBoard and exported as structured CSV summaries for subsequent analysis.

Table 4.1: Summary of RGB-based vegetation indices (VIs) used in this study.

Index (name & abbreviation)	**Formula**
Normalized Green–Red Difference Index (NGRDI) [33]	$\frac{G-R}{G+R}$
Green Leaf Index (GLI) [34]	$\frac{2G-R-B}{2G+R+B}$
Triangular Greenness Index (TGI) [34]	$-0.5[(640-480)(R-G) -(640-550)(R-B)]$
Visible Atmospherically Resistant Index (VARI) [34]	$\frac{G-R}{G+R-B}$
Red–Blue Vegetation Index (RBVI) [53]	$\frac{R-B}{R+B}$
Normalized Green-Blue Difference Index (NGBDI) [48]	$\frac{G-B}{G+B}$
Red–Blue Normalized Difference Vegetation Index (RBNDVI)	$\frac{N-(R+B)}{N+(R+B)}$
Simple Ratio Pigment Index (SRPI)	$\frac{B}{R}$
RGB Vegetation Index (RGBVI) [76]	$\frac{G^2-RB}{G^2+RB}$
Green Normalized Value (GN) [2]	$\frac{G}{B+G+R+RE+N}$

Table 4.2: Summary of multispectral vegetation indices (VIs) used in this study.

Index (name & abbreviation)	**Formula**
Normalized Difference Vegetation Index (NDVI) [41]	$\frac{N-R}{N+R}$
Normalized Difference Red Edge Index (NDRE) [46]	$\frac{N-RE}{N+RE}$
Green Normalized Difference Vegetation Index (GNDVI) [44]	$\frac{N-G}{N+G}$
Blue Normalized Difference Vegetation Index (BNDVI) [49]	$\frac{N-B}{N+B}$
Green–Blue Normalized Difference Vegetation Index (GBNDVI) [49]	$\frac{N-(G+B)}{N+(G+B)}$
Green–Red Normalized Difference Vegetation Index (GRNDVI) [49]	$\frac{N-(G+R)}{N+(G+R)}$
Canopy Chlorophyll Content Index (CCCI) [50]	$\left(\frac{N-RE}{N+RE}\right)\left(\frac{N-R}{N+R}\right)$
Ratio Vegetation Index (RVI) [47]	$\frac{N}{R}$
Chlorophyll Vegetation Index (CVI) [42]	$\frac{N \cdot R}{G^2}$
Red-edge Normalized Value (RE_n) [2]	$\frac{RE}{B+G+R+RE+N}$
Red Normalized Value (R_n) [2]	$\frac{R}{B+G+R+RE+N}$
Plant Senescence Reflectance Index ($PSRI_{alt}$) [54]	$\frac{R-G}{N}$
Structure Insensitive Pigment Index (SIPI) [54]	$\frac{N-B}{N-R}$

Chapter 5

System Validation

This chapter evaluates the performance and reliability of both High-Throughput Phenotyping Systems (HTPS) described in Chapter 4. Validation focused on three main aspects: (i) segmentation accuracy and consistency in the greenhouse dataset, (ii) colour correction and its influence on downstream prediction performance, and (iii) verification of multispectral sensor behaviour in the outdoor setup. Together, these tests establish the accuracy limits and practical reliability of the imaging and processing pipelines.

5.1 Greenhouse RGB HTPS

5.1.1 Segmentation accuracy validation

Top leaf selection is an ambiguous task, as the potted plants contained multiple stems with their own top leaves, which were similar in size and height to each other and whose relative height could not be distinguished by a human with 100% certainty. To quantify both human agreement and automated-pipeline performance we built an interactive annotation tool that allows an annotator to click one or more seed points on the target object (either the whole plant or the top leaf). The points are passed to `SAM2`, which returns a candidate mask. If the annotator is satisfied, the mask is accepted and the next image is loaded; otherwise a new set of points can be supplied. This procedure produces masks that are consistent with automated pipeline segmentations, because every mask ultimately comes from the same segmentation backend.

Two independent raters annotated the *08-29-Morning* image set, which comprises

192 frames in total (96 top-down and 96 angled views). Because the whole-plant outline is unambiguous in these greenhouse images, we only measured inter-rater reliability for the top-leaf target, where multiple leaves can confuse the "highest" leaf definition when viewed in RGB. This established a baseline for human agreement, allowing direct comparison with the automated segmentation pipeline. Following common practice in instance segmentation the masks were evaluated using mean average precision at an intersection-over-union (IoU) threshold of 0.50 (mAP@50). For inter-rater reliability the ground-truth for rater A is the prediction for rater B; for automated evaluation the human mask is treated as the ground-truth and the pipeline mask as the prediction. For whole-plant segmentation the depth-guided LangSAM routine achieved nearly perfect masks: mAP@50=89.6–100% depending on view configuration. Qualitative inspection confirmed that the few failures were caused by inclusion of the plant pot, not by missed plant material.

In contrast, top-leaf segmentation proved much harder.The two human annotators achieved an mAP@50 of 60.4% on the top-leaf task (Table 5.1). This moderate agreement reflects the genuine ambiguity: depending on camera angle, a single oat culm can present two or three leaves at comparable heights. This ceiling establishes the upper bound we can reasonably expect from any automated method that relies only on monocular RGB cues. On the *08-29-Morning* set the pipeline obtained only mAP@50=15.6% for angled views and 18.8% for top-down views. Visual diagnostics showed four recurring error modes: (i)A spikelet was misclassified as a leaf; (ii) SAM2 occasionally grouped the target leaf with an adjacent leaf or spikelet, forming a merged mask whose IoU with the single-leaf ground-truth falls below 0.5; (iii) depth errors led to selection of the second-highest rather than the true apex leaf; (iv) When a leaf curved or folded over itself, both the front and back surfaces were often segmented as separate leaves by SAM2, due to the strong boundary line along the fold. Although poor in a strict pixel sense, these masks still capture useful morphological context (multiple leaves, partial stem, spikelet), which we believe explains why the downstream image regressors performed *better* with the "noisy" pipeline masks than with the crisply-cropped human single-leaf masks. Section 6.1.2 compares the effectiveness of both the pipeline and human segmentation modes for both top-leaf and whole-plant targets for licor metric prediction.

These findings define the achievable segmentation accuracy in greenhouse conditions and contextualize the limits of automated leaf detection for subsequent regression analyses.

Table 5.1: Segmentation accuracy (mAP@50) and inter-rater agreement.

Target	Comparison	View(s)	Dataset	mAP@50
Top leaf	Human vs. Human	top+angled	08-29-Morning	60.4 %
Whole	Pipeline vs. Human	angled	08-29-Morning	100.0 %
Whole	Pipeline vs. Human	top	08-29-Morning	89.6 %
Whole	Pipeline vs. Human	top+angled	08-29-Morning	94.8 %
Top leaf	Pipeline vs. Human	angled	08-29-Morning	15.6 %
Top leaf	Pipeline vs. Human	top	08-29-Morning	18.8 %
Top leaf	Pipeline vs. Human	top+angled	08-29-Morning	17.2 %

5.1.2 Colour correction

Beyond segmentation accuracy, colour consistency across the dataset can significantly affect both visual indices and learned features. We therefore validated and compared several white-balancing and colour-normalisation approaches.

While the Luxonis cameras come with reliable auto white balance, there turned out to be a need for white balancing across the whole dataset. This was made challenging without an inclusion of a white balancing target in the imaging data. A few approaches were tested, such as matching rectangles of the background plywood between adjacent images that have some overlap using sliding window template matching using SSIM [77]. This method, while able to match regions in the background plywood reliably, introduced a lot of noise attempting to match the lighting to a reference frame recursively. What proved the most reliable was segmenting the entire background plywood in each frame using SAM2 point selection tool from inter-rater reliability testing and using it as a global colour matching template for standard RGB gain. This normalized the images so they appear to be taken in the same lighting. For RGB indices, the correlations dropped significantly, with the highest R^2 for GSW dropping from 0.42 to 0.185. VPD_{leaf} becomes the most correlated LI-600 reading, dropping to 0.28. This is likely due to a dampening of biological colour variation, or perhaps speaks to the plywood not being an optimal white balancing target. On the image regression side, performance post-normalization doesn't change more than margin of error. Due to the ambiguity of the effectiveness of the white balancing in this instance, analysis will be done without white balancing for this dataset.

To quantify the effect of illumination normalisation, image-domain statistics were evaluated across all greenhouse subsets. Histogram intersection between original and corrected images remained extremely high (mean ≈ 0.99 across channels), indicating that global colour distributions changed only minimally after normalisation. However, calibration consistency was low when comparing corrected calibration

frames to the fixed reference target, with a mean structural similarity index (SSIM) of 0.30, suggesting that the reference-based correction did not reliably reproduce a consistent illumination state. Channel variance shifts were also predominantly negative, indicating mild dynamic range compression rather than stable colour alignment. Together, these results confirm that the greenhouse colour normalisation produced small but inconsistent transformations, supporting the decision to use uncorrected images for subsequent analyses.

The colour-balanced datasets were used both to train the three deep learning models and to compute RGB correlations, as shown in Table 5.2. Neutral-grey balancing yielded slightly better performance than white balancing, although the overall differences were minor. Given the ambiguity of the greenhouse results and the risk of introducing artificial colour noise without a stable calibration reference, colour normalisation was not applied to the greenhouse dataset, while outdoor preprocessing decisions are discussed separately in the following section.

Table 5.2: Model performance (R^2) under original, white-balanced, and grey-balanced preprocessing for the Greenhouse dataset. Bold indicates the best value in each column.

Condition	Model	gsw	VP_{leaf}	VPD_{leaf}	H_2O_{leaf}	F_s	Fm'
Original	ResNet	**0.24**	0.66	0.50	0.65	-0.30	-0.01
	EfficientNet	0.19	**0.70**	**0.60**	0.69	0.20	0.31
	Swin	0.15	0.67	0.53	0.68	-0.10	0.20
	RGB Metric	0.10	0.14	0.14	0.15	0.13	0.09
White Balanced	ResNet	0.09	0.59	0.57	0.61	0.10	0.34
	EfficientNet	0.12	0.62	0.44	0.64	-0.18	0.31
	Swin	0.11	**0.70**	0.58	0.70	0.04	0.33
	RGB Metric	0.08	0.14	0.14	0.14	0.12	0.11
Grey Balanced	ResNet	0.12	0.58	0.56	0.60	0.24	0.40
	EfficientNet	0.08	0.64	0.46	0.61	0.22	**0.42**
	Swin	0.10	0.69	0.56	**0.71**	**0.22**	0.34
	RGB Metric	0.08	0.15	0.14	0.15	0.10	0.09

5.2 Outdoor Multispectral HTPS

5.2.1 Multispectral Camera Validation

To validate the suitability of the modified OAK-D LR sensor for computing vegetation indices, we performed a set of post-hoc sanity checks based on established practice in low-cost multispectral imaging [78, 79]. These tests do not provide full radiometric calibration but are sufficient to confirm that the sensor captures the expected spectral

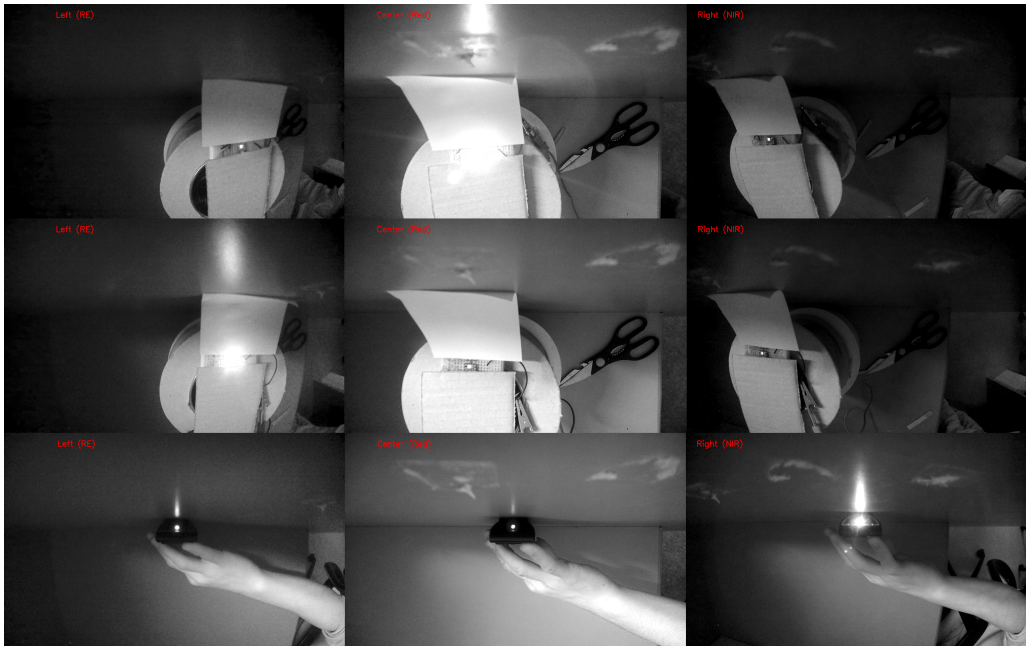

Figure 5.1: Each column shows the output of one of the three cameras: Red Edge (left), Visual Red (centre), and Near-Infrared (right). Each row corresponds to a different LED illumination: Visual Red (top, narrowband), Red Edge (middle, narrowband), and Near-Infrared (bottom, broadband).

contrasts required for NIR and red-edge-based analysis.

First, spectral separation was assessed using 3 LEDs, a visual red LED, RE-narrowband LED and a NIR TV-remote LED. Each was imaged with all three sensors to test if responses are focused on the correct bands. Figure 5.1 shows the result of this experiment side by side. The VIS red LED is very brightly lit in only the visual red image, a tiny amount leaks through to RE and NIR sensors, this is expected and normal, as the LED is very bright and is a direct light source, so even a fraction of a percentage leak would still show up. Next, a similar effect is seen with the narrowband RE LED, with a bright light visible with the RE sensor and dim dots leaking through to the other channels. In the absence of a narrowband NIR LED on hand, a TV remote was used as a broadband NIR light source. TV remotes emit predominantly NIR light, but with a wider full width halfmax, this can be seen by the relatively larger fractions leaking through to the RE and VIS red sensors. It should be noted that the TV remote light was invisible to the naked eye, so what we're seeing is RE light leaking through the factory IR-cut filter on the Luxonis LR camera.

Near infrared imaging is used in the medical field for effective visualisation of subsurface blood vessels [80]. To test if our multispectral camera behaves as expected in this context, an image of an arm was taken, shown in Figure 5.2. RGB and VIS

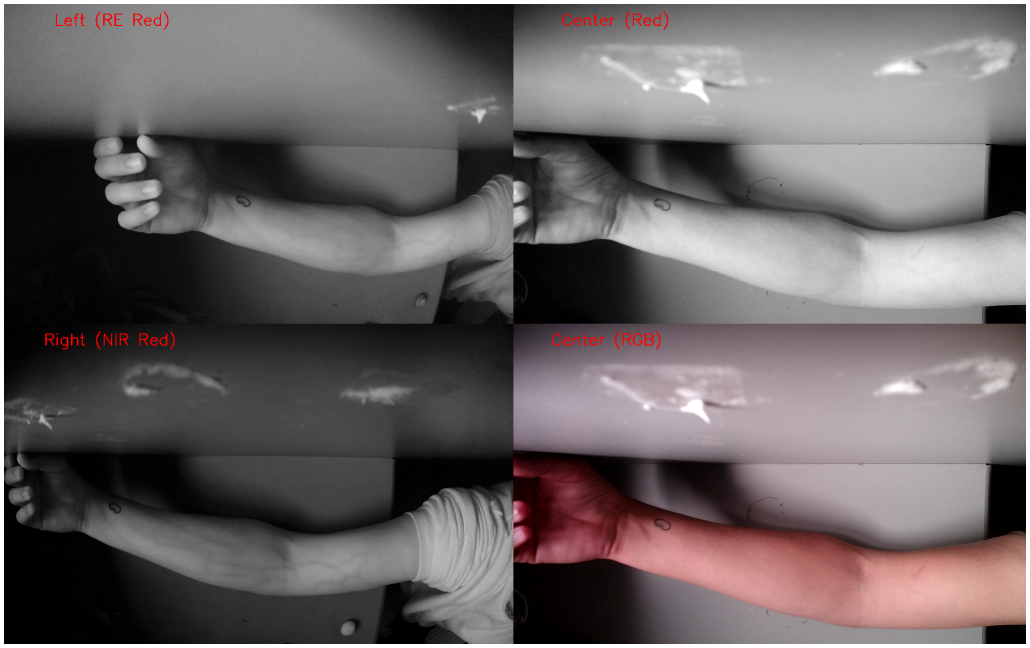

Figure 5.2: a) RE image showing visible veins, b) VIS red image showing barely visible veins, c) NIR image showing very clearly visible veins, d) RGB image showing almost no veins

red images show very faint veins, as can be seen with the naked eye. The NIR sensor shows significantly better visualisation of subsurface veins. What was faintly visible before is now extremely clear, and smaller veins that weren't visible at all in VIS red are now clear. The RE image lies somewhere between the VIS red and NIR images. This is the expected behaviour for the 3 sensors and is a point towards the custom MS camera being a valid alternative to more expensive cameras.

An important characteristic of near-infrared (NIR) imaging is that water attenuates NIR light more strongly than visible or red-edge wavelengths. To illustrate this effect, an experiment was conducted as shown in Figure 5.4. A glass of water was imaged under four conditions-empty, 1 cm depth, half-full, and full-using the three camera modules (red-edge, visual red, and NIR). A reflective metal object was placed at the bottom of the glass to visualize how water depth affected apparent brightness.

This experiment demonstrates that the NIR camera successfully detects the expected attenuation of near-infrared light by water. As shown in Figure 5.3, mean pixel intensity at the bottom of the glass decreased from 65.1 (empty) to 38.6 (full), corresponding to a 40.7 % reduction. In contrast, the visible camera brightness increased due to additional surface reflections. The RE image showed no change, this is in line with expectations as water should absorb a very small amount of RE light. This confirms that the multispectral sensor is sensitive to the characteristic absorption behaviour of water in the near-infrared range.

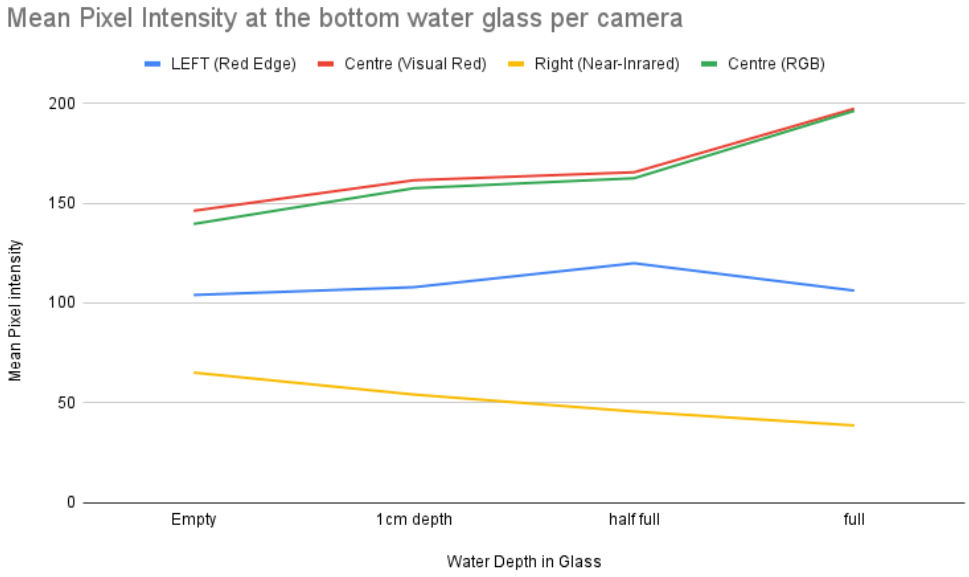

Figure 5.3: Visualisation of pixel intensity as water depth increases

5.2.2 Colour correction

White Balancing

In the open-field (OF) dataset, a dedicated white-balancing procedure was implemented. The gantry beam along the x-axis was painted with a titanium-dioxide-based white paint. As the gantry traversed the bed in a raster pattern, the beam became visible at the start and end of each y-axis movement, providing a consistent reference target for colour calibration.

White balancing was applied by iterating over the full dataset and, for each frame, locating the most recent "raster limit" frame (where the gantry's y-coordinate was either zero or its maximum value). These limit frames always contained the painted beam, which served as the reference target.

The algorithm then extracted the balancing region in two stages. First, the reference frame was converted to a "whiteness" map by averaging across RGB channels, followed by Otsu thresholding and contour detection. The largest contour was selected, providing a rough mask of the painted beam (including shaded portions and some background noise). An axis-aligned largest inscribed rectangle was then computed to crop this region tightly.

Afterwards, within this cropped rectangle, the same process was repeated - Otsu thresholding is applied, and after a second contour-finding step, the largest inscribed

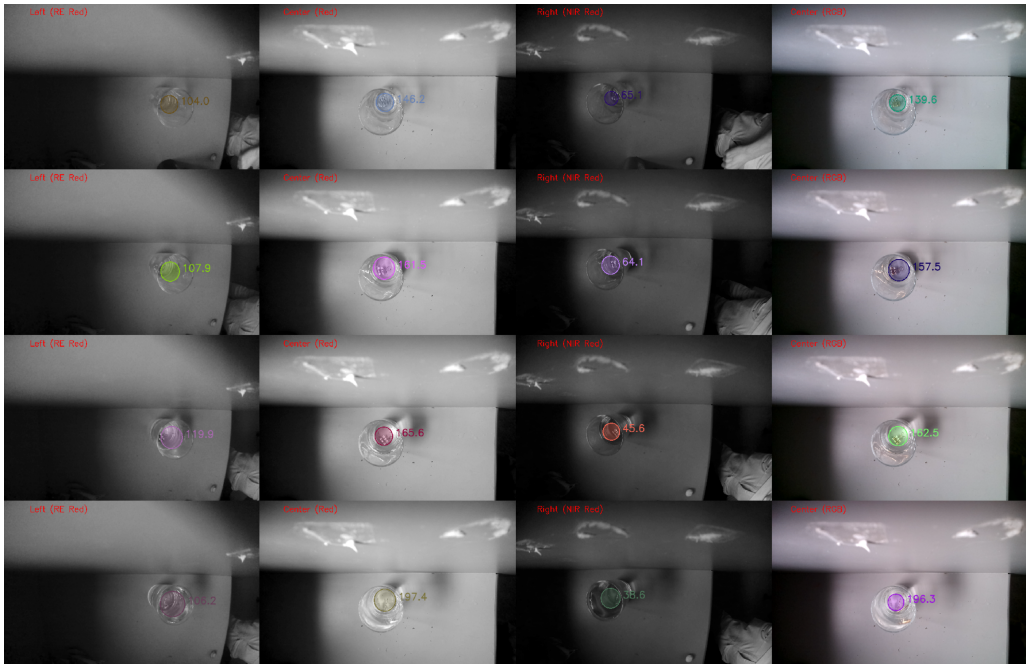

Figure 5.4: Water Experiment

rectangle was computed again. This ensured that the final calibration region corresponded to the most uniformly bright part of the white target, minimizing the influence of dirt, shadows, or paint irregularities.

Finally, per-channel gains were computed so that the mean RGB value of this refined region mapped to pure white $(255, 255, 255)$, and these gains were applied to the original plant image. In addition, a *neutral-grey balancing* variant was implemented. Instead of forcing the region to pure white, the per-channel gains were chosen so that $R = G = B$, equalising the three channels to a common grey level. This process is visualised in Figure 5.5 Together, these procedures provided consistent colour normalisation under variable outdoor lighting conditions. Grey balancing produced more natural corrections, this was probably caused by the white balancing target exposure peaking the sensor in some instances and not others. For this reason, only grey balancing was implemented in the full processing.

To quantify the stability of both correction strategies, image-domain statistics were evaluated across the dataset. Histogram intersection between original and corrected images averaged 0.94 for grey balancing and 0.87 for white balancing, indicating that grey balancing preserved the original colour distribution more closely. Calibration consistency was assessed using the SSIM between the corrected beam region and an ideal uniform target (neutral grey for grey balancing and pure white for white balancing). Both yielded comparable, high scores (grey: 0.95, white: 0.96), confirming that the painted beam provides a reliable calibration reference. This

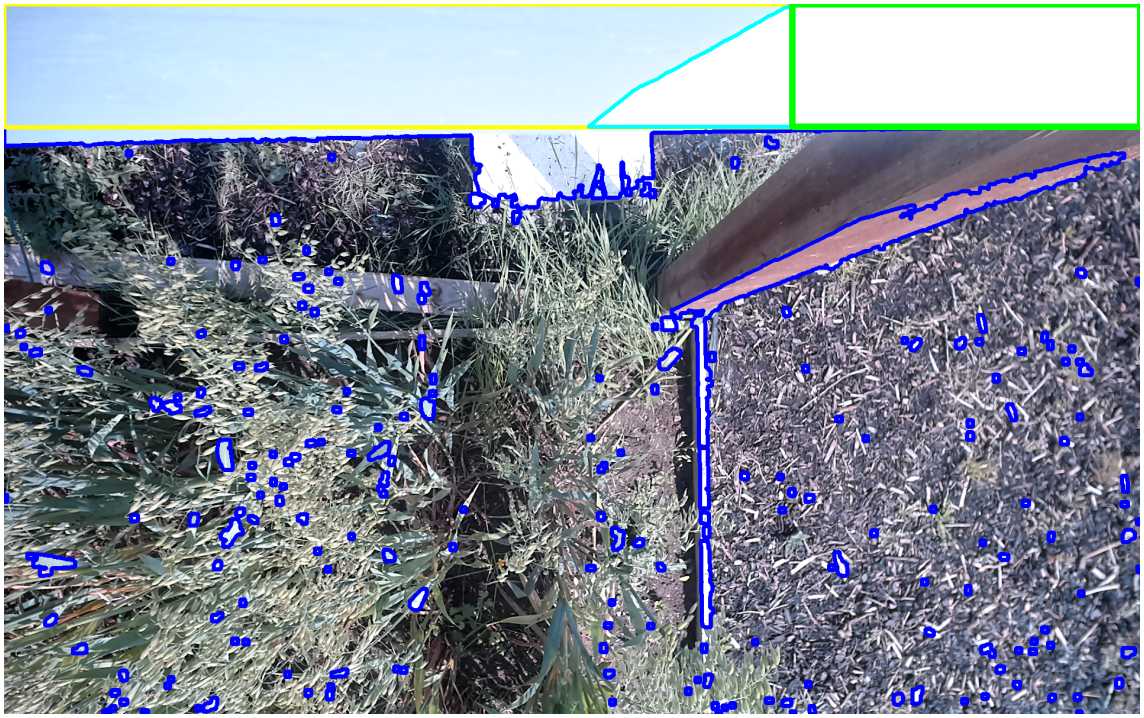

Figure 5.5: White balancing target selection. Blue - initial contours of whitest region in the image; Yellow - Largest axis-aligned inscribed rectangle in contour; Cyan - second round of contours of whitest region in yellow rectangle; Green - Largest axis-alligned rectangle in cyan contours, final white-balance target selection.

analysis supports that white balancing yielded more aggressive, less consistent results than grey balancing.

Effect of Grey-Balancing on Model Performance

To evaluate whether grey-balance correction improved prediction accuracy, models were trained on both the unaltered and grey balanced versions of the outdoor multispectral dataset. Table 5.3 compares the best-performing models across the four data splits for each condition.

Across the four splits, grey-balanced preprocessing produced marginal gains for most LI-600 physiological metrics but noticeably stronger performance in yield and biomass prediction. In particular, Split A showed a substantial improvement in yield ($R^2 = 0.42$ vs. –0.37) and biomass ($R^2 = 0.39$ vs. –0.06). These results suggest that while grey-balancing results in minimal change for direct physiological trait estimation, it meaningfully enhances predictions of aggregate agronomic traits such as yield and biomass.

Table 5.3: Model performance (R^2) under grey-balanced and unaltered preprocessing for the outdoor field dataset. Bold indicates the best value in each column.

Condition	Split	gsw	$\mathbf{VP_{leaf}}$	$\mathbf{VPD_{leaf}}$	$\mathbf{F_s}$	$\mathbf{Fm'}$	yield	biomass
Grey	A	0.04	0.69	0.58	0.23	0.20	0.42	0.39
Balanced	B	0.26	0.66	0.52	0.27	0.25	**0.49**	0.21
	C	**0.27**	0.72	0.62	**0.42**	**0.31**	-1.05	0.12
	D	0.23	**0.77**	0.60	0.19	0.23	0.04	**0.45**
Unaltered	A	0.25	0.62	0.46	0.27	0.27	-0.37	-0.06
	B	**0.27**	0.67	0.55	0.23	0.26	-0.87	0.19
	C	**0.27**	0.76	**0.63**	0.31	0.29	-0.76	0.17
	D	0.23	0.75	0.60	0.16	0.21	0.23	0.40

5.2.3 HSV filtering

HSV-based color filtering was explored to remove background elements from the images. Both normalized and absolute HSV color spaces were tested by manually labeling plant and background pixels across a representative image set. The resulting 3D point clouds of labeled colours showed very similar distributions between the two approaches, so the absolute HSV values were used for simplicity. To reduce noise, regions in HSV space containing only background points were excluded, while mixed regions containing both plant and background samples were retained to avoid removing plant pixels. The remaining plant colour region was converted into a 3D binary colour-space volume, downsampled to a 32×32×32 grid, and dilated and eroded to form a smooth continuous representation of likely plant colours. However, when this mask was applied to the images, it preserved non-green plant material, such as heavily shaded and red-yellow leaves, while also including large portions of lighter soil and bark-mulch regions. The distinction between soil and plant pixels was highly inconsistent due to variable lighting conditions, shadows, and color overlap (particularly in senescing leaves and yellow barley). Consequently, HSV-based background filtering was not implemented for this dataset, though it remains a potentially valuable preprocessing step for achieving cleaner segmentation and analysis in future work under more controlled illumination conditions.

Chapter 6

Results

6.1 RGB Experiment Analysis

This section includes analysis done on the greenhouse dataset, compared against RGB data from the outdoor field dataset.

6.1.1 RGB Index Analysis

To evaluate how well RGB indices capture physiological variation measured by the LI-600, we first analysed the GH dataset. Each pot was imaged from four viewpoints (top and angled orientations on both sides). For every RGB index–statistic pair (e.g. NGRDI mean, HI standard deviation), the median value across the four views was taken, reducing the influence of outliers and producing a single representative feature list per pot per day. Data from the first and last imaging days were then concatenated, yielding 90 rows of paired RGB features and LI-600 measurements for analysis.

Pearson correlations were then computed between each LI-600 metric and each camera metric. To calculate statistical significance, the Bonferroni correction was applied, which adjusts the p-values by multiplying them by the number of tests to control for false positives in multiple comparisons. The p-value represents the probability that the observed correlation could have occurred by chance; a lower p-value suggests a stronger indication that the correlation is real. Typically, a p-value below 0.05 is considered significant, meaning there is less than a 5% chance the result is due to random variation. Since the Bonferroni correction multiplies the p-values by the total number of comparisons made, in this case 702, it is a highly conservative

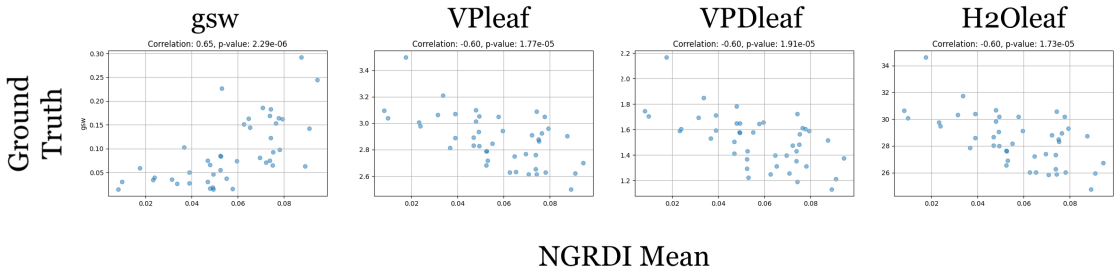

Figure 6.1

method, meaning that any correlations deemed significant using this method are very unlikely to be false positives.

Limiting the RGB correlations to the top camera view with the whole plant as the target, after adjusting the p-value using the Bonferroni correction, a total of 43 significant correlations were found, with 10 unique camera metrics exhibiting at least one significant correlation. Significant correlations were found with 4 of the 6 LI-600 metrics- gsw, VPleaf, VPDleaf and H2Oleaf, The mean value of NGRDI was the most correlated RGB metric to all of the LI-600 metrics, as shown in Table 6.1. No significant correlations were found for Fs or Fm'. A scatter plot visualising these correlations is shown in figure 6.1

Table 6.1: Correlation and R^2 between LI-600 physiological trait and Camera Metric for greenhouse experiment.

LI-600 Metric	**Camera Metric**	**Correlation**	**R^2**
GSW	NGRDI Mean	0.648	0.416
VP_{leaf}	NGRDI Mean	-0.598	0.358
VPD_{leaf}	NGRDI Mean	-0.597	0.356
H_2O_{leaf}	NGRDI Mean	-0.599	0.359

The RGB correlations were significantly lower for the outdoor-field experiment, with R^2 values ranging from 0.15 to 0.09. However, unlike the GH experiment, significant correlations were found for all LI-600 metrics, as shown in Table 6.2.

6.1.2 Statistical and Deep Learning Analysis

We also trained end-to-end learning models (ResNet50, Swin Transformer, and EfficientNet-B5) to predict LI-600 physiological traits, both individually and jointly. Unlike index-based approaches, these models learn discriminative features directly from the images, without requiring prior assumptions about which visual cues are relevant.

Table 6.2: Correlation and R^2 between LI-600 physiological trait and Camera Metric for outdoor-field experiment.

LI-600 Metric	**Camera Metric**	**Correlation**	**R^2**
GSW	a* channel Q1	-0.315	0.099
VP_{leaf}	BLUE Variance	0.380	0.144
VPD_{leaf}	BLUE Variance	0.370	0.137
H_2O_{leaf}	BLUE Variance	0.383	0.147
Fs	a* channel Q1	-0.367	0.133
Fm'	GLI Q3	0.305	0.093

Table 6.3: Comparison of segmentation targets and methods. Human inter-rater agreement (Human A vs Human B) and automated segmentation accuracy (automated method vs Human A) are reported separately.

Target	**Segmentation method**	**Segmentation accuracy (%)**	**Swin R^2**	**RGB R^2**
Whole	Human	100.00	0.32	0.325
Whole	Automated	89.58	0.41	0.42
Leaf	Human	53.12	-0.64	0.142
Leaf	Automated	18.75	0.14	0.181

For gsw prediction on the GH dataset, angled and top-down views produced very similar results (RMSE = 0.115 and 0.114, respectively), while combining both reduced the error slightly further (RMSE = 0.105). This small gain, however, does not justify the added hardware complexity, so top-down views were chosen for the remainder of the analysis and for future experiments conducted using an open-field mobile robot. This choice is also practical: in dense canopies angled imaging is not feasible, as cameras cannot easily be lowered between plants. These results confirm that relying on top-down views alone provides sufficient predictive accuracy.

Table 6.3 compares the performance of different segmentation targets and methods. Segmentation accuracy is reported as mAP@50, where a prediction is considered correct if the Intersection-over-Union (IoU) with the reference mask exceeds 50%. For human-based segmentation, this corresponds to inter-rater reliability (IRR), which represents the maximum agreement that can be achieved between annotators. Automated segmentation methods yielded higher predictive performance than human segmentation, likely because the pipeline tends to capture additional background context that indirectly aids prediction. Furthermore, using the whole plant as the segmentation target outperformed using only the top leaf, supporting the hypothesis that including more visual information in the target region improves prediction accuracy.

Table 6.4 summarizes the performance of the three deep learning models alongside

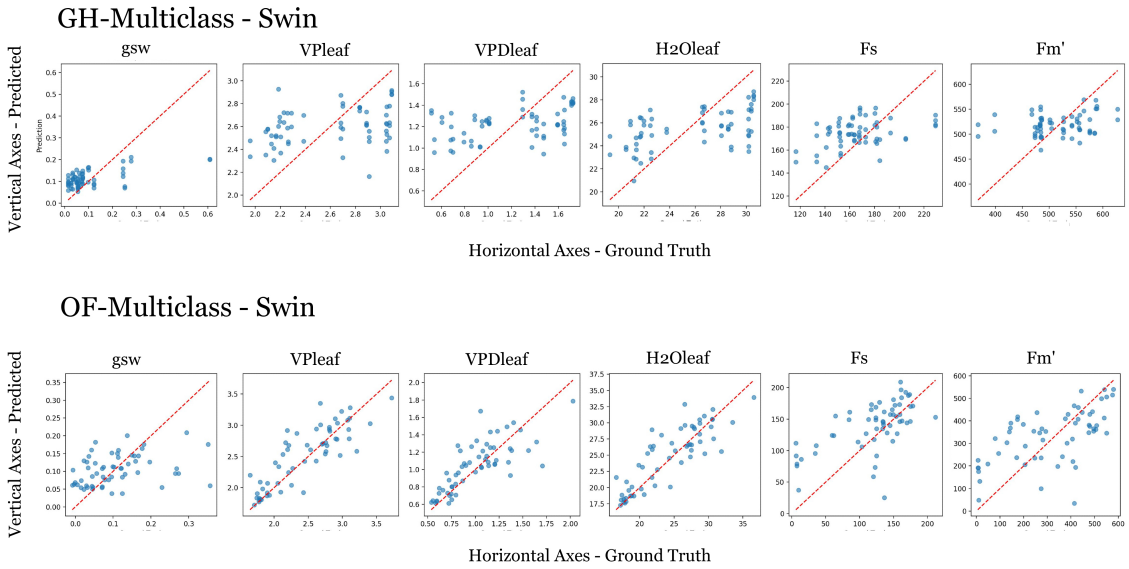

Figure 6.2

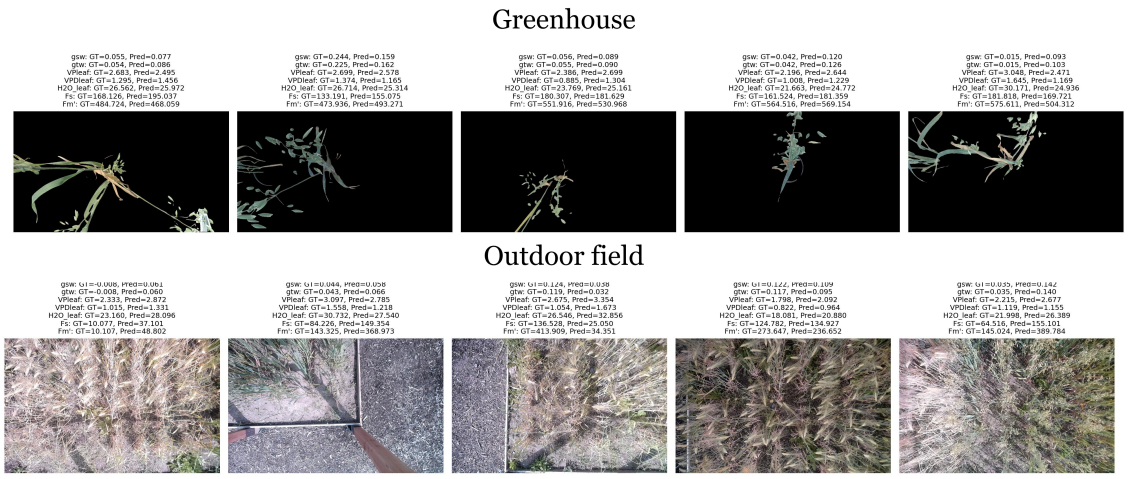

Figure 6.3

RGB correlation (R^2) values. A scatter plot visualisation of the best performing multiclass models for the greenhouse and outdoor field are shown in figure 6.2. A visualisation of a few example inference runs (randomly selected from the test set) is shown in figure 6.3. Training was performed using top-down camera views and the cleaned dataset. For the GH data, the entire plant was used as the segmentation target. In contrast, for the outdoor field (OF) dataset, a centre crop was applied to capture the target plant while still including some neighbouring plants. Multiclass models were trained to predict all LI-600 metrics simultaneously, while singleclass models were trained separately for each metric.

In the greenhouse experiment, the Swin model achieved the best overall performance. However, the RGB correlation results indicate that the NGRDI may outperform

Table 6.4: Model performance (R^2) predicting physiological traits from RGB images. Bold indicates the best value in each column (per experiment) (ties bolded). "NS" = not significant.

Data	Model	gsw	$\mathbf{VP_{leaf}}$	$\mathbf{VPD_{leaf}}$	$\mathbf{H_2O_{leaf}}$	$\mathbf{F_s}$	$\mathbf{Fm'}$
GH-Multiclass							
	ResNet	0.35	0.14	0.13	0.13	0.10	-0.56
	EfficientNet	0.33	-0.09	-0.17	-0.08	-0.08	-0.08
	Swn	0.40	0.28	0.22	0.23	-0.19	-0.10
GH-Singleclass							
	ResNet	0.38	0.11	0.31	0.21	0.17	0.06
	EfficientNet	**0.47**	0.22	0.29	0.27	0.04	-0.11
	Swin	0.41	0.27	**0.40**	0.35	**0.24**	0.02
	RGB Metric	0.42	**0.35**	0.36	**0.36**	NS	NS
OF-Multiclass							
	ResNet	**0.24**	0.66	0.50	0.65	-0.30	-0.01
	EfficientNet	0.19	0.70	0.60	0.69	0.20	0.31
	Swin	0.15	0.67	0.53	0.68	-0.10	0.20
OF-Singleclass							
	ResNet	0.11	**0.82**	**0.68**	**0.84**	**0.24**	0.33
	EfficientNet	0.13	0.81	0.60	0.82	0.23	0.33
	Swin	0.18	**0.82**	0.55	0.83	0.12	**0.34**
	RGB Metric	0.10	0.14	0.14	0.15	0.13	0.09

deep learning for the potted dataset.

Results differ substantially for the outdoor field experiment. Here, g_{sw} proved difficult to predict, with R^2 values below 0.25, whereas VPleaf, VPDleaf, and H2Oleaf achieved strong predictions ($R^2 = 0.82$, 0.68, and 0.84, respectively). Heatmap analysis of LI-600 measurements using kriging suggests that this difference arises because g_{sw} values appear spatially "random," while VPleaf, VPDleaf, and H2Oleaf exhibit smoother spatial patterns, with neighbouring plants showing similar values. Since the centre-cropping approach for this dataset includes parts of neighbouring plants, their visual features may have reinforced these smoother traits, leading to higher R^2 values. In contrast, the lack of spatial coherence in g_{sw} measurements likely limited predictive performance.

6.1.3 Drought Prediction

Drought prediction was only evaluated on the greenhouse (GH) dataset, as plants in the outdoor field experiment were not subjected to any artificial drought stress. A drought prediction model was trained using a random forest with 200 decision trees and gsw, VPleaf, VPDleaf, and H2Oleaf as the 4 inputs. Fs and Fm' were ignored as

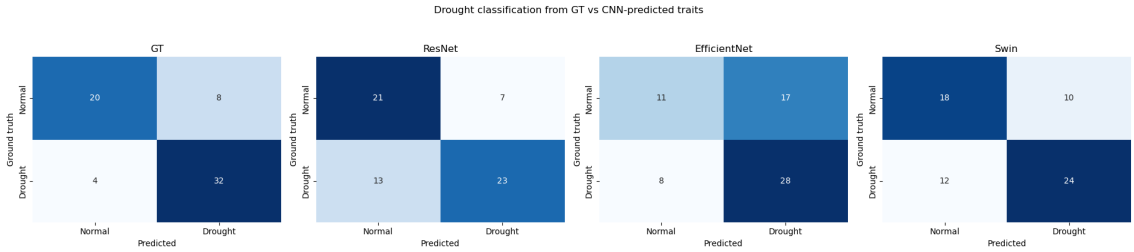

Figure 6.4: Confusion matrices for drought prediction using ground-truth features and predictions from EfficientNet-B5, ResNet50, and Swin models. Only Swin achieved a level of agreement with ground truth, while ResNet50 and EfficientNet-B5 exhibited large systematic errors.

both had low R^2 values. The model was trained, validated and tested using the same splits as the deep learning models. To test prediction performance, the combined outputs of single-class models were fed into the drought classifier. Although drought was imposed as a binary treatment, effective plant water status likely varied due to soil composition and microbial effects. By the time of imaging late in the drought period, however, the visual separation between drought-treated and normally watered plants was clear, making binary labelling a reasonable approximation of end-state stress under realistic agronomic variability.

The results are summarized in Figure 6.4 and Table 6.5. Relative to the LI-600 baseline (Balanced Accuracy (BA) = 0.80), the CNN-based models recover a substantial proportion of the attainable classification performance. Swin achieves a BA of 0.655 with a well-balanced precision–recall profile (F1-score = 0.686), indicating that the model captures meaningful physiological differences between drought and control plants. ResNet50 performs similarly (BA = 0.694), exhibiting the highest precision among the CNNs (0.767) but slightly lower recall. EfficientNet-B5 attains a lower BA (0.585), though it maintains strong sensitivity to drought cases (recall = 0.778). Because the dataset is moderately imbalanced (60% drought vs. 40% normal), BA is reported to ensure equal weighting of sensitivity and specificity across models.

Table 6.5: Classification performance of drought prediction models using Random Forest classification on predicted physiological traits.

Model	Accuracy	BA	Precision	Recall	F1-score
Ground truth	0.812	0.802	0.800	0.889	0.842
ResNet50	0.688	0.694	0.767	0.639	0.697
EfficientNet-B5	0.609	0.585	0.622	0.778	0.691
Swin	0.656	0.655	0.706	0.667	0.686

Figure 6.5 compares the distributions of normalized physiological traits predicted by Swin against the ground-truth LI-600 measurements, focusing on gsw and VPleaf,

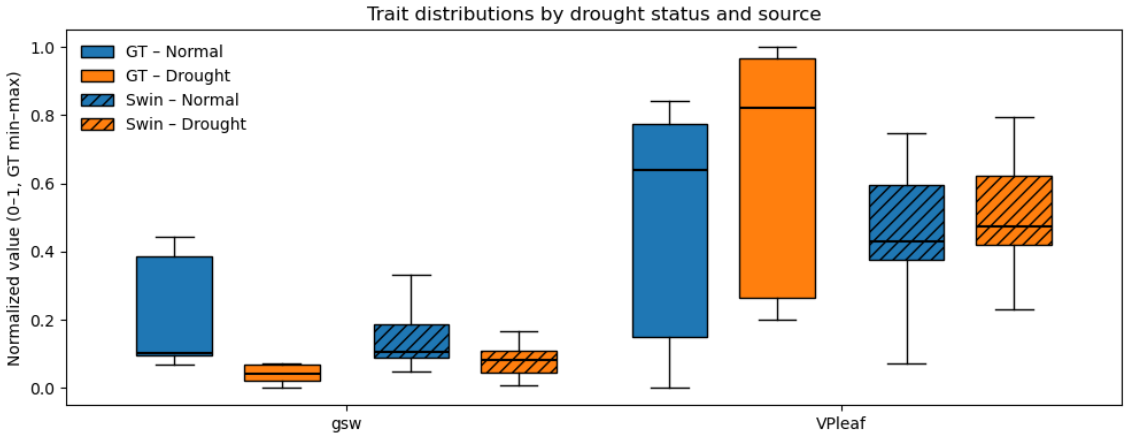

Figure 6.5

as VPDleaf and H2Oleaf were found to be strongly collinear with VPleaf. The LI-600 measurements show a clear separation between drought and normal plants: drought plants exhibit substantially lower gsw and higher VPleaf, reflecting stomatal closure and increased leaf vapor pressure under water stress. Swin captures these drought–normal offsets well for both traits. For gsw, the ground truth displays an almost binary separation with minimal overlap between groups, whereas the Swin distributions reproduce the same directional shift but with narrower distributions and partial overlap. For VP_{leaf}, Swin again recovers the correct offset in mean values, although the predicted distributions are noticeably narrower than the LI-600 measurements and do not express the full dynamic range of the sensor. These differences in spread, rather than direction of shift, explain the remaining misclassifications in the confusion matrix: Swin correctly encodes the drought trend but compresses the magnitude of the physiological response. Overall, the model reflects the underlying drought physiology, recovering the major directional changes associated with water stress.

6.2 Outdoor Experiment Analysis

6.2.1 Multispectral Vegetation Index Analysis

To evaluate how well RGB and multispectral (MS) indices capture physiological variation measured by the LI-600, a similar procedure was followed for the outdoor field data. The Pearson correlations between each index and each LI-600 metric across all plants and imaging days were computed. Bonferroni correction was applied to all p-values to account for multiple comparisons, ensuring that significant correlations

were unlikely to be false positives.

Indices were instead calculated by cropping a centre rectangle of size 384px by 384px, the same as what is used in the deep learning analysis. The mean of each channel is calculated to compute the vegetation indices. This both avoids mis-registration artifacts while also acting as a comparison to remote sensing, with the whole image simulating a single pixel with a GSD of approximately 345 mm at canopy height. This is opposed to GSD of approximately 0.9 mm at canopy height from the full images.

To compare RGB-only and MS-derived features, we ranked all indices by their mean absolute correlation across LI-600 traits (excluding yield, which had limited sample size). This balanced averaging approach avoids overweighting any single LI-600 trait and provides a more general assessment of index performance. The top 10 indices are shown in Table 6.6, alongside whether they require multispectral bands or are RGB-only.

Table 6.6: Top 10 vegetation indices (VI) ranked by averaged mean coefficient of determination (R^2) across LI-600 metrics. The modality column refers to whether the VI can be calculated using RGB data only or requires multispectral bands.

Index	**Mean R^2**	**Modality**
BLUE	0.246	RGB
GREEN	0.241	RGB
RED	0.224	RGB
BNDVI	0.205	MS
GN	0.203	MS
GBNDVI	0.201	MS
RBNDVI	0.196	MS
GNDVI	0.195	MS
GRNDVI	0.188	MS
GBVI	0.184	RGB

Interestingly, the strongest correlations in this dataset were obtained with individual colour bands. This behaviour has been noted in prior studies [81], especially at lower altitudes, but such single-band correlations are less stable across different imaging conditions. Vegetation indices based on normalized band combinations, while in this instance weaker in absolute correlation, provide more robust performance. For this reason, we focus below on comparing the best *index-based* features rather than raw band values. It should be noted however how NIR (Mean $R^2 = 0.066$) and RE (Mean $R^2 = 0.053$) don't come up in that table. The underlying cause for this warrants further investigation.

Table 6.7 directly compares the strongest RGB-only index; GBVI against the best

multispectral index; BNDVI across all LI-600 traits. BNDVI outperformed GBVI for every trait except biomass, where the RGB index retained an advantage. This shows the importance of multispectral information for most physiological traits at remote sensing resolution. This is the expected result based on existing literature.

Table 6.7: Comparison of best RGB-only index (GBVI) versus best multispectral index (BNDVI) across LI-600 metrics.

Metric	**GBVI** (R^2)	**BNDVI** (R^2)
Fm'	0.128	0.207
Fs	0.018	0.032
VPDleaf	0.425	0.435
VPleaf	0.357	0.399
biomass	0.095	0.068
gsw	0.042	0.046

Averaged across all indices, BNDVI achieved the highest overall R^2, confirming that MS indices consistently capture physiological variation more effectively than RGB-only metrics. At the per-metric level (Table 6.8), MS indices outperformed RGB indices for four of the six unique LI-600 parameters, particularly for Fm′ and *biomass*, where they explained 17% and 7% more variance, respectively. Meanwhile, RGB indices retained value for certain parameters. The Simple Ratio Pigment Index (SRPI) closely matched the performance of BNDVI for leaf vapour pressure-related traits (VPDleaf, and VPleaf), with differences below 0.01 R^2. Finally, the Triangular Greenness Index (TGI), derived solely from RGB reflectance slightly outperformed multispectral indices for *gsw*, demonstrating that visible-band indices can still be effective proxies for stomatal conductance under these conditions. These results show that at coarse GSD, access to multispectral information is essential extracting plant physiological parameters as accurately as possible.

Table 6.8: Best multispectral and RGB indices per LI-600 metric (excluding raw colour bands).

Metric	**Best MS Index**	R^2	**Best RGB Index**	R^2
Fm'	R n	0.307	RGBVI	0.137
Fs	R n	0.133	TGI	0.108
VPDleaf	BNDVI	0.435	SRPI	0.441
VPleaf	BNDVI	0.399	SRPI	0.393
biomass	RE n	0.170	SRPI	0.100
gsw	R n	0.073	TGI	0.104

6.2.2 Texture Analysis

Alongside the VI calculation for each image, 8 gray level co-occurrence matrix textures features were extracted from both the green and NIR channels. Table 6.9 shows the mean R^2 for each texture metric, indicating that textures derived from the green channel are more powerful than textures derived from the NIR channel, being ahead eight out of eight times. For the GLCM feature calculations, four cardinal directions were used with a pixel offset of one. Other offsets (2, 4, 8 pixels, and the average across all distances) were evaluated, but a distance of one produced the best results.

Table 6.9: Mean coefficient of determination (R^2) for Green-channel and Near-Infrared channel texture metrics.

Texture Metric	**Mean R^2 (Green)**	**Mean R^2 (NIR)**
Angular Second Moment	0.068	0.067
Contrast	0.234	0.081
Correlation	0.253	0.201
Dissimilarity	0.237	0.071
Energy	0.082	0.044
Entropy	0.160	0.010
Homogeneity	0.156	0.002
Variance	0.121	0.044

Much like Table 6.8, Table 6.10 compares the predictive power of multispectral (MS) and RGB texture features, specifically those derived from the Near-Infrared (NIR) and Green channels, respectively. A similar trend is observed: Green-channel texture outperformed NIR texture for five of the seven LI-600 metrics, achieving higher R^2 values across most physiological parameters. However, NIR texture maintained a slight advantage for fluorescence-related traits (Fm' and Fs), exceeding Green texture by only 0.01–0.03 R^2. In contrast, for vapour pressure-related traits ($VPDleaf$ and $VPleaf$) and *biomass*, Green texture explained 20–30% more variance, indicating substantially greater predictive power. These results suggest that at proximal imaging scales, with fine GSD (approx 0.9 mm) spatial information encoded in texture features becomes more informative than additional spectral bands. At this resolution, subtle differences in canopy structure, leaf orientation, and surface roughness dominate the signal, diminishing the relative benefit of multispectral information.

6.2.3 Deep Learning Analysis

Table 6.11 shows the best training results for multi-class training on RGB and RGB+MS data, all of which are RGB. It also shows that the introduction of

Table 6.10: Best Green-channel and Near-Infrared texture metrics per LI-600 measurement.

Metric	Best Green Texture	R^2	Best NIR Texture	R^2
Fm'	homogeneity	0.236	correlation	0.244
Fs	correlation	0.084	correlation	0.113
VPDleaf	correlation	0.527	correlation	0.409
VPleaf	contrast	0.567	correlation	0.372
biomass	correlation	0.211	correlation	0.148
gsw	correlation	0.066	correlation	0.064
yield	correlation	0.032	correlation	0.020

Table 6.11: Comparison of best-performing model configurations for each dataset split with multi-class model outputs (combined oats and barley). Performance is measured as R^2

Dataset	Mode	Model	*gsw*	*VPleaf*	*VPDleaf*	*Fs*	*Fm′*
Split_A	RGB	ResNet	0.037	0.687	0.575	0.230	0.201
Split_B	RGB	ResNet	0.256	0.659	0.521	0.273	0.252
Split_C	RGB	ResNet	0.268	0.718	0.624	0.423	0.312
Split_D	RGB	Swin	0.228	0.769	0.603	0.192	0.228

synthetic data leads to generally higher model R^2 on the unseen test set, with Split D achieving the strongest overall performance across all LI-600 traits. As described in 3.2.4, Splits A–D represent increasing levels of synthetic data incorporation: Split A contains only ground-truth data, Split B is a 50/50 mix, Split C includes only synthetic data, and Split D combines both full sets. The RGB modality outperformed the multispectral configuration (RGB + Red Edge + NIR) in most cases, suggesting that, at the proximal imaging scale, visible-spectrum and texture cues may already capture much of the physiologically relevant variation. However, this trend may also be influenced by residual spatial misalignment between the multispectral channels, introduced by asynchronous frame capture during data acquisition. Such registration artifacts can create artificial edges that the model may misinterpret as meaningful features, thereby masking true spectral signals. Future work should address this through improved calibration and temporal synchronization to fully leverage the additional multispectral information.

ResNet demonstrated superior and more stable performance compared to Swin Transformers in this configuration. Once again, vapour-pressure-related traits proved the easiest to predict, with R^2 values ranging from 0.69 to 0.77. Fluorescence-related parameters (Fs and Fm') were moderately well-predicted ($R^2 = 0.19$–0.42), whereas stomatal conductance (gsw) remained the most challenging, reaching a maximum of $R^2 = 0.27$. Overall, these findings suggest that the inclusion of synthetic training data improves model generalization to unseen samples. Notably, Split C often produced higher R^2 values than Split A. Further investigation is needed to confirm

the cause of this trend, though it is likely related to the lower variance and smoother spatial gradients present in the synthetic data, which make it easier for the model to learn consistent predictive relationships compared to the more variable ground-truth LI-COR readings.

Table 6.12 presents the results of single-variable training, illustrating that focusing on individual LI-600 traits generally leads to higher model performance compared to multi-output prediction. The effects of synthetic data and spectral modality observed previously were reinforced in these single-variable results. However, unlike in the multi-target experiments, the Swin Transformer outperformed ResNet in most single-variable cases. However, this gain came at a considerable computational cost: to achieve superior performance, multiple Swin models had to be trained and aggregated, whereas ResNet achieved comparable accuracy as a single, faster, and more lightweight network. Overall, while single-output training provides a modest accuracy advantage, it also illustrates diminishing returns relative to the added training complexity and computational expense.

Table 6.12: Single-variable model performance across dataset splits. Performance is measured as R^2

Dataset	Mode	Model	*gsw*	*VPleaf*	*VPDleaf*	*Fs*	*Fm′*
Split A	RGB	Swin	0.168	0.776	0.602	0.113	0.192
Split B	RGB	ResNet	0.290	0.688	0.557	0.328	0.309
Split C	RGB	Swin	0.267	0.715	0.582	0.334	0.328
Split D	MS	Swin	0.285	0.784	0.667	0.084	0.322

Table 6.13 compares single-variable model performance across species for Split D, which combines the full ground-truth dataset with an equal amount of synthetic data, effectively doubling the training set size. The RGB modality was used for all runs, as it generally yielded the best overall performance in prior experiments. Although Split D (RGB) was not always the absolute top-performing configuration for every species, the overall performance patterns and trends remain consistent even when substituting the best split for each species.

The combined dataset included both oat and barley plants, with training, validation, and testing samples drawn randomly from the unified pool. As described in Section 3.2.4, the combined dataset contains a slightly larger proportion of oat plants due to their greater representation in the experimental trials. For comparison, the single-species datasets were split independently using the same proportional structure as the combined dataset to ensure fairness.

Overall, model performance was comparable across all species configurations, indi-

cating strong transferability of the learned features. The combined model performed on par with the species-specific models for most traits, with only minor variations. Fluorescence parameter *Fs* showed notably higher R^2 in the single-species models, while *VPleaf* slightly exceeded both individual species in the combined case-though not by a statistically meaningful margin. The remaining metrics (*gsw*, *VPDleaf*, and *Fm′*) fell within the range defined by the single-species models.

These results suggest that, within this methodological setup, it is feasible to train a species-agnostic model capable of accurately predicting key physiological traits across both oats and barley. This supports the broader goal of developing generalized phenotyping systems applicable across various cereal species.

Table 6.13: Comparison of single-variable model performance across species and dataset variants. Performance is measured using the coefficient of determination (R^2).

Dataset	Model	*gsw*	*VPleaf*	*VPDleaf*	*Fs*	*Fm′*
Combined	Swin	0.257	0.799	0.630	0.119	0.259
Barley	Swin	0.343	0.752	0.646	0.145	0.104
Oat	Swin	0.256	0.812	0.588	0.263	0.451

6.2.4 Yield and Biomass Prediction

Ground-truth yield and biomass data were obtained following the procedures outlined in Section 3.2.2. To evaluate model performance on these agronomic traits, the best-performing configurations identified for each corresponding task were selected for analysis. Table 6.14 summarizes the yield prediction performance for oats, the bets performer for which was resnet traiend on multispectral data. Yield prediction achieved a moderately strong R^2 of 0.52, which is encouraging given the aforementioned inconsistencies between image timing and final harvest weights. This result suggests that visual yield cues-such as panicle size, colouration, and canopy density-were effectively captured by the imaging system despite post-imaging changes caused by predation and delayed harvesting. The oat yield dataset contained 101 samples, with a standard deviation of 5.48 and a range of 26.0, indicating moderate variability across plants. Barley data were excluded from yield analysis because of severe post-imaging predation and less frequent imaging intervals, leaving insufficient valid data to train a reliable model.

To complement the yield analysis, biomass prediction was also examined as part of the broader agronomic evaluation of model performance (Table 6.15) Barley achieved the highest predictive accuracy ($R^2 = 0.74$, MS, Swin), followed by the combined dataset

Table 6.14: Best-performing model for oat yield prediction (resnet , multispectral data). The table also reports the number of samples (n), standard deviation, and range of ground-truth yield values.

Species	n	**Std.**	**Range**	***yield*** R^2
Oats	101	5.48	26.0	0.524

($R^2 = 0.56$, MS, Swin) and oats ($R^2 = 0.39$, MS, ResNet). This reversal relative to yield performance may be explained by the narrower range and lower variability of biomass in barley (standard deviation 7.52; range 41.0; $n = 141$) compared to oats (standard deviation 17.53; range 78.0; $n = 127$) Visual inspection of the dataset supports this interpretation-barley tended to grow in more uniform clumps with consistent tiller numbers, while oat "plants" often ranged from single stems to dense multi-tiller clusters, complicating the relationship between visual canopy area and total biomass. While physiological traits exhibited strong inter-species transferability, the prediction of biomass and yield remains challenging to generalize, reflecting the higher structural and developmental variability inherent to these traits.

Table 6.15: Best-performing models for biomass prediction across species and dataset configurations. The table also reports the number of samples (n), standard deviation, and range of ground-truth biomass values for each species.

Species	n	**Std.**	**Range**	***biomass*** R^2
Combined	268	14.96	79.0	0.561
Barley	141	7.52	41.0	0.739
Oats	127	17.53	78.0	0.387

Chapter 7

Conclusions and Future Work

This thesis investigated the design and implementation of two cost-conscious robotic High-Throughput Phenotyping Systems - a greenhouse multi-camera RGB system and an outdoor single-camera multispectral system - to evaluate whether proximal imaging can serve as a reliable proxy for LI-COR LI-600 physiological measurements. These systems aim to inform future development of AURA's field robot prototypes and to assess the feasibility of low-cost, scalable imaging approaches for drought-related physiological traits.

7.1 Discussion of results

Across all experiments, models predicting moisture related LI-600 parameters - namely leaf vapour pressure (VP_{leaf}), vapour pressure deficit (VPD_{leaf}), and leaf water content (H_2O_{leaf}) - achieved the highest predictive performance. These traits are physiologically interconnected and express strongly in both spectral and textural domains, which explains their reliable estimation from RGB and multispectral imagery. Biomass and yield, in contrast, were more variable: they depend on complex structural and developmental cues that are not well captured from single imaging sessions.

The models' ability to predict LI-COR traits generalised robustly across crop species, with comparable performance between oats and barley. However, yield and biomass prediction did not generalise, reflecting their stronger dependence on physical structure rather than more general health indicators.

7.1.1 Yield and Biomass Prediction

ResNet50 achieved $R^2 = 0.52$ for oat yield and $R^2 = 0.56$ for biomass across oats and barley combined. These results, while moderate, demonstrate potential for trait estimation beyond physiology but also highlight sensitivity to environmental and logistical factors - especially post-imaging predation and asynchronous harvesting. Future studies that tighten the temporal link between imaging and harvest, or introduce explicit spikelet detection and counting, could improve these outcomes.

7.2 Summary of Research Objectives

This thesis answers the following key research questions:

1. **Can proximal RGB imaging predict LI-COR metrics for oat plants in a greenhouse setting?**
   Yes. In the controlled greenhouse experiment using automated whole-plant segmentation targets (Table 6.4), the Swin Transformer achieved a prediction accuracy of $R^2 = 0.27$ for leaf vapour pressure, $R^2 = 0.40$ for vapour pressure deficit, and $R^2 = 0.41$ for stomatal conductance, demonstrating that RGB imaging can approximate drought-related physiological traits.

2. **Can proximal multispectral imaging estimate LI-COR metrics, yield and biomass for oats and barley in an open field setting?**
   Partly. In the outdoor raised-bed multispectral HTPS, the Swin Transformer achieved $R^2 = 0.784$ for leaf vapour pressure, $R^2 = 0.667$ for vapour pressure deficit, and $R^2 = 0.285$ for stomatal conductance using a five-channel RGB+MS input (Table 6.12). Multispectral inputs only outperformed RGB in the most data-rich configuration (Split D). For agronomic traits, multispectral data proved more advantageous: ResNet50 achieved $R^2 = 0.52$ for oat yield (Table 6.14) and $R^2 = 0.56$ for biomass across oats and barley (Table 6.15). Overall, models generalised well between species for LI-COR physiological traits, but yield and biomass predictions did not transfer as reliably.

In addition, two new datasets were produced and will be publicly released, including the first known dataset documenting visible spikelet predation in cereal crops. A modular image-processing pipeline for plant and organ segmentation was developed and validated, and the potential of LI-600 measurement interpolation for synthetic training and validation data generation was demonstrated.

7.3 Key takeaways

The following points summarise the main experimental insights regarding model behaviour, cross-species generalisation, and system design trade-offs.

- **Whole-plant context matters.** End-to-end regressors trained on whole-plant masks consistently outperformed top-leaf models (e.g. *gsw* $R^2 = 0.41$ vs. 0.14 for whole vs. leaf targets). Broader morphological context outweighed segmentation precision, demonstrating that information density matters more than exact boundaries.
- **Multispectral advantage is limited at proximal scale.** The addition of red-edge and near-infrared channels provided only marginal gains, suggesting that for short-range imaging, RGB's higher spatial and textural fidelity is often sufficient.
- **Swin vs. ResNet performance is context-dependent.** Both architectures performed similarly overall, but distinct trends emerged: Swin performed more reliably when trained on interpolated LI-600 data and single-variable regressions, whereas ResNet performed better with direct ground-truth data and multi-target predictions. This pattern suggests that Swin benefits from structured, information-dense representations, while ResNet remains more tolerant of noise and mixed signal sources.
- **Physiological traits transfer across species; yield does not.** Models maintained strong predictive power for LI-600 traits across oats and barley, but yield and biomass estimation failed to generalise. Further analysis of the underlying reasons is important.
- **Performance vs. deployment complexity.** Combining top-down and angled views improved RMSE only marginally ($0.1139 \rightarrow 0.1048$) while adding significant hardware complexity. Future systems will prioritise simpler top-down arrays that maintain accuracy while reducing calibration and maintenance requirements.

7.3.1 Model complexity and colour-based baselines

In several cases, particularly for greenhouse-grown potted plants, classical RGB indices such as NGRDI matched or exceeded the performance of deep learning

regressors. This indicates that low-order colour statistics capture a large proportion of the observable variation in RGB imagery that correlates with LI-600 measurements, and that increased model complexity alone does not guarantee improved predictive power in proximal phenotyping settings. This phenomenon is likely analogous to observations in vegetation-index literature, where raw colour channels often show strong correlations with plant health under homogeneous conditions and within a narrow developmental window. In such settings, broad colour bands can absorb much of the variation associated with physiological status. However, these relationships are largely context-dependent, making them difficult to translate into robust or scalable health metrics. In contrast, deep learning models offer a pathway to incorporate shape, texture, and spatial context beyond colour alone, which may be essential for retrieving more grounded and transferable health indicators in real world field environments.

7.4 Limitations

Colour calibration in uncontrolled field conditions

The colour normalisation strategies explored in this study were deliberately grounded in existing agricultural and remote-sensing practice. Prior work has shown that effective colour normalisation in outdoor settings often relies on pragmatic reference surfaces-such as painted panels [82] or invariant scene elements [83]. This is because laboratory-style calibration targets are difficult to deploy and maintain in the field. Our approach similarly sought to use a persistent, structurally integrated reference that did not interfere with plant imaging or restrict scalability.

In practice, this approach proved less effective than anticipated for several reasons that highlight the constraints of long-term, close-range field robotics. First, unlike studies that perform calibration relative to a single short imaging campaign, our systems operated across multiple days under varying illumination, during which the surfaces gradually accumulated debris like dirt and bird droppings. Second, colour balancing was applied post hoc rather than used to drive real-time exposure control, limiting its ability to compensate for dynamic lighting changes [84, 85]. Third, the reference surface was imaged at close range; at this scale, texture and wood grain beneath the paint introduced additional high-frequency variation that would likely be imperceptible from aerial platforms. Collectively, these factors reduced the effectiveness of the normalisation.

Alternative calibration strategies-such as placing reference panels at canopy or plant level-have also been reported as problematic due to occlusion by foliage, shadowing, and mixed illumination, and are difficult to scale across an entire field [85]. Mounting a dedicated calibration target above the canopy on the robot itself may mitigate some of these issues, but introduces additional mechanical complexity, vibration, and a risk of occluding the imaging field of view. Although post-hoc colour correction has been effective in other field phenotyping studies [84], our results indicate that its utility diminishes when imaging is performed over extended periods in uncontrolled environments. This aligns with reports that even commercially manufactured colour calibration targets can degrade or become unreliable under prolonged natural illumination [86]. More broadly, the use of non-ideal reference surfaces is well established in remote sensing, where roads, sand, or concrete are routinely treated as invariant calibration proxies; however, our findings suggest that such approximations are less reliable at proximal scales. Overall, colour calibration remains a key limitation of the current pipeline and an important area for future refinement.

Dataset size and generalisability

While the total number of plants in each experiment is modest relative to large-scale breeding trials, the dataset size is consistent with many prior high-throughput phenotyping studies that pair imaging data with proximal or biochemical ground-truth measurements. Comparable studies have evaluated segmentation and regression models using on the order of 50–150 plants or images, particularly when annotation or sensor measurements are costly or time-consuming [87, 88, 89].

In this work, all train/validation/test splits were performed at the *plant level*, ensuring that no images from the same plant appeared in more than one subset. Data augmentation was applied during training to further reduce overfitting. The strong agreement between simple RGB index correlations and end-to-end learning models suggests that the observed relationships reflect genuine physiological signal rather than dataset memorisation.

This study should therefore be interpreted as a pilot-scale evaluation of a low-cost, RGB-only HTPS architecture, intended to assess signal feasibility and system trade-offs rather than to establish population-level predictive models. Larger multi-season and multi-genotype deployments are identified as a key direction for future work.

Additional limitations

Additional limitations relate to sensing accuracy and deployment constraints. Depth-based segmentation relies on monocular predictions from Depth Anything, which can degrade under wind-induced motion blur. Because depth is only recovered up to scale, metric traits such as plant height or leaf inclination cannot be reliably estimated, and overlapping leaves may be incorrectly ordered, limiting the range of extractable 3D phenotypes. From a systems perspective, the current pipeline processes data on a single high-end GPU; real-time embedded deployment will require model optimisation or distillation. Finally, LI-600 measurements were treated as single ground-truth samples per plant, and repeated sampling across culms or time points would allow measurement noise to be better quantified in future studies.

- **Depth estimation.** Accurate metric depth was not recovered in either the greenhouse or outdoor experiments. Monocular methods provided relative but not absolute scale, limiting derivation of traits such as plant height and leaf inclination.
- **Multispectral misalignment and network issues.** Frame registration drifted due to asynchronous spectral band capture. Additionally, network latency occasionally caused FarmBot pathing interruptions, inconsistent frame rates, and coordinate reporting lag, all of which reduced temporal synchronisation accuracy.
- **Throughput constraints.** The pipeline currently runs on a desktop GPU. Real-time operation on embedded hardware will require optimisation, quantisation, or model distillation.
- **LI-600 sampling noise.** Each pot was measured once per session. Repeated within-pot measurements could quantify natural variability and improve model calibration.

7.5 Future work

1. **Field deployment on Callan.** The system will be mounted on the AURA field robot to evaluate scalability under natural outdoor conditions.
2. **Metric depth via stereo+learning.** Combining stereo disparity with neural depth estimation will enable metric-level canopy height and leaf angle prediction.

3. **Controlled water stress outdoors.** Replicating the greenhouse drought design in open-field conditions will further validate real-world use.

4. **Night-time imaging or dark-hood enclosure.** Capturing images under uniform artificial lighting at night or using a light-sealed imaging hood will mitigate illumination variability and remove dependence on unreliable calibration targets.

5. **Spectral calibration.** Use of NDVI or grey reference panels for periodic spectral validation will improve long-term consistency.

6. **Spikelet detection and counting.** Future work could incorporate spikelet detection and counting to identify signs of predation, for example by detecting steady-state losses in visible grain number over time.

7. **Generalisation studies.** Extending this approach to additional species, soil conditions, and stress treatments will test scalability and robustness.

8. **Further Validation.** Further Validation of the multispectral camera with instruments such as precise vegetation index calibration boards or diffusion boards for linearity testing would be beneficial.

7.6 Concluding remarks

The developed High-Throughput Phenotyping Systems show that low-cost RGB imaging at proximal range can approximate LI-COR physiological readings with high fidelity and cross-species generalisability, *without* requiring multispectral bands or complex segmentation pipelines. In practice, simple top-down imaging and whole-plant context (via coarse windowing) proved sufficient and more robust than organ-level segmentation. By releasing hardware designs, datasets, and code, this work promotes reproducible and affordable phenotyping. The outdoor raised-bed system provides a critical bridge toward fully autonomous deployment on *Callan*, enabling large-scale, data-driven field phenotyping for sustainable and climate-resilient cereal production.

Bibliography

[1] S. F. H. K.-K. S.-L. G. A. Brosinsky, T. Kuester. (2019) Principles of imaging spectroscopy – EO college. [Online]. Available: https://eo-college.org/resource/principles-of-imaging-spectroscopy/

[2] N. Guimarães, J. J. Sousa, P. Couto, A. Bento, and L. Pádua, "Combining UAV-based multispectral and thermal infrared data with regression modeling and SHAP analysis for predicting stomatal conductance in almond orchards," vol. 16, no. 13, p. 2467, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/16/13/2467

[3] G. Roggiolani, M. Sodano, T. Guadagnino, F. Magistri, J. Behley, and C. Stachniss, "Hierarchical Approach for Joint Semantic, Plant Instance, and Leaf Instance Segmentation in the Agricultural Domain," Jun. 2023, arXiv:2210.07879 [cs]. [Online]. Available: http://arxiv.org/abs/2210.07879

[4] E. Marks, M. Sodano, F. Magistri, L. Wiesmann, D. Desai, R. Marcuzzi, J. Behley, and C. Stachniss, "High Precision Leaf Instance Segmentation for Phenotyping in Point Clouds Obtained Under Real Field Conditions," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4791–4798, Aug. 2023, conference Name: IEEE Robotics and Automation Letters. [Online]. Available: https://ieeexplore.ieee.org/document/10158793

[5] T. T. H. Giang and Y.-J. Ryoo, "Pruning Points Detection of Sweet Pepper Plants Using 3D Point Clouds and Semantic Segmentation Neural Network," *Sensors*, vol. 23, no. 8, p. 4040, Apr. 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/8/4040

[6] M. Presten, R. Parikh, S. Aeron, S. Mukherjee, S. Adebola, S. Sharma, M. Theis, W. Teitelbaum, and K. Goldberg, "Automated Pruning of Polyculture Plants," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*. Mexico City, Mexico: IEEE, Aug. 2022, pp. 242–249. [Online]. Available: https://ieeexplore.ieee.org/document/9926632/

[7] A. Kicherer, K. Herzog, N. Bendel, H.-C. Klück, A. Backhaus, M. Wieland, J. C. Rose, L. Klingbeil, T. Läbe, C. Hohl, W. Petry, H. Kuhlmann, U. Seiffert, and R. Töpfer, "Phenoliner: A new field phenotyping platform for grapevine research," vol. 17, no. 7, p. 1625, 2017.

[8] J. Angeja, "Automated Pruning of Greenhouse Indeterminate Tomato Plants," Jun. 2018.

[9] FAO, *The State of Food and Agriculture 2021*. FAO, Nov. 2021. [Online]. Available: http://www.fao.org/documents/card/en/c/cb4476en

[10] R. U. Hameed, C. Meade, and G. Lacey, "Technology advancements and the needs of farmers: Mapping gaps and opportunities in row crop farming," vol. 15, no. 15, p. 1664, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2077-0472/15/15/1664

[11] AURA - maynooth university. [Online]. Available: https://aura-mu.github.io/

[12] Q. Tang, R. Zhang, L. Chen, G. Xu, W. Deng, C. Ding, M. Xu, T. Yi, Y. Wen, and L. Li, "High-accuracy, high-resolution downwash flow field measurements of an unmanned helicopter for precision agriculture," vol. 173, p. 105390. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169919314814

[13] R. Jiang, P. Wang, Y. Xu, Z. Zhou, X. Luo, Y. Lan, G. Zhao, A. Sanchez-Azofeifa, and K. Laakso, "Assessing the operation parameters of a low-altitude UAV for the collection of NDVI values over a paddy rice field," vol. 12, no. 11, p. 1850, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/12/11/1850

[14] R. Van De Vijver, K. Mertens, K. Heungens, D. Nuyttens, J. Wieme, W. H. Maes, J. Van Beek, B. Somers, and W. Saeys, "Ultra-high-resolution UAV-based detection of alternaria solani infections in potato fields," vol. 14, no. 24, p. 6232, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/14/24/6232

[15] User manual for DJI devices. [Online]. Available: https://support.dji.com/help/content?customId=en-us03400008333&spaceId=34&re=US&lang=en&documentType=artical&paperDocType=paper

[16] Annual climate statement for 2024 - met Éireann - the irish meteorological service. [Online]. Available: https://www.met.ie/annual-climate-statement-for-2024

[17] C. D. S. Ecosystem. Sentinel-2 | copernicus data space ecosystem. [Online]. Available: https://dataspace.copernicus.eu/data-collections/copernicus-sentinel-data/sentinel-2

[18] licor. (2025) Li-600 porometer flourometer. [Online]. Available: https://www.ecosearch.info/sites/default/files/prodotti_documentazione/LI-600.pdf

[19] C. Kubota. (2023) Vpdleaf vs vpdair - two different ways to determine vpd. [Online]. Available: https://e-gro.org/pdf/e816.pdf

[20] J. Xie, Y. Chen, Z. Yu, J. Wang, G. Liang, P. Gao, D. Sun, W. Wang, Z. Shu, D. Yin, and J. Li, "Estimating stomatal conductance of citrus under water stress based on multispectral imagery and machine learning methods," vol. 14, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1054587/full

[21] Y. Zhou, C. Lao, Y. Yang, Z. Zhang, H. Chen, Y. Chen, J. Chen, J. Ning, and N. Yang, "Diagnosis of winter-wheat water stress based on UAV-borne multispectral image texture and vegetation indices," vol. 256, p. 107076. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378377421003413

[22] M. Jia, D. Li, R. Colombo, Y. Wang, X. Wang, T. Cheng, Y. Zhu, X. Yao, C. Xu, G. Ouer, H. Li, and C. Zhang, "Quantifying chlorophyll fluorescence parameters from hyperspectral reflectance at the leaf scale under various nitrogen treatment regimes in winter wheat," vol. 11, no. 23, p. 2838, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/11/23/2838

[23] S. Rasti, C. J. Bleakley, N. Holden, R. Whetton, D. Langton, and G. O,ÄôHare, "A survey of high resolution image processing techniques for cereal crop growth monitoring," *Information Processing in Agriculture*, vol. 9, no. 2, pp. 300–315, Jun. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2214317321000226

[24] P. Song, J. Wang, X. Guo, W. Yang, and C. Zhao, "High-throughput phenotyping: Breaking through the bottleneck in future crop breeding," *The Crop Journal*, vol. 9, no. 3, pp. 633–645, Jun. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2214514121000829

[25] A. Balabantaray, S. Behera, C. Liew, N. Chamara, M. Singh, A. J. Jhala, and S. Pitla, "Targeted weed management of palmer amaranth using robotics

and deep learning (YOLOv7)," vol. 11, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2024.1441371/full

[26] X. Zhou, X. Wang, L. Ji, S. Daggubati, K. Shen, and V. M. Whitaker, "Deep learning for strawberry runner detection integrating ground and aerial imaging," vol. 12, p. 101290. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375525005210

[27] D. Guri, M. Lee, O. Kroemer, and G. Kantor, "Hefty: A modular reconfigurable robot for advancing robot manipulation in agriculture." [Online]. Available: http://arxiv.org/abs/2402.18710

[28] N. Friedman, A. Mehta, K. Love, A. Bremers, A. Ahmed, and W. Ju, "A utility belt for an agricultural robot: reflection-in-action for applied design research." [Online]. Available: http://arxiv.org/abs/2312.14358

[29] L. Grimstad, C. D. Pham, H. T. Phan, and P. J. From, "On the design of a low-cost, light-weight, and highly versatile agricultural robot," in *2015 IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pp. 1–6, ISSN: 2162-7576. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7428210

[30] L. Grimstad and P. From, "The thorvald II agricultural robotic system," vol. 6.

[31] M. R. Barbosa Júnior, R. G. d. Santos, L. d. A. Sales, and L. P. d. Oliveira, "Advancements in agricultural ground robots for specialty crops: An overview of innovations, challenges, and prospects," vol. 13, no. 23, p. 3372, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2223-7747/13/23/3372

[32] Small Robot Company. (2023, October) Small Robot Co launches Tom V4: Next generation monitoring robot. Medium. [Accessed: 29-Aug-2025]. [Online]. Available: https://medium.com/smallrobotcompany/small-robot-co-launches-tom-v4-next-generation-monitoring-robot-56b2f224d247

[33] S. Osuna-Caballero, T. Olivoto, M. A. Jiménez-Vaquero, D. Rubiales, and N. Rispail, "RGB image-based method for phenotyping rust disease progress in pea leaves using R," *Plant Methods*, vol. 19, no. 1, p. 86, Aug. 2023. [Online]. Available: https://doi.org/10.1186/s13007-023-01069-z

[34] E. R. Hunt Jr., C. S. T. Daughtry, J. U. H. Eitel, and D. S. Long, "Remote Sensing Leaf Chlorophyll Content Using a Visible Band

Index," *Agronomy Journal*, vol. 103, no. 4, pp. 1090–1099, 2011, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2134/agronj2010.0395. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.2134/agronj2010.0395

[35] D. Kumar, S. Kushwaha, C. Delvento, Liatukas, V. Vivekanand, J. T. Svensson, T. Henriksson, G. Brazauskas, and A. Chawade, "Affordable Phenotyping of Winter Wheat under Field and Controlled Conditions for Drought Tolerance," *Agronomy*, vol. 10, no. 6, p. 882, Jun. 2020. [Online]. Available: https://www.mdpi.com/2073-4395/10/6/882

[36] S. Mertens, L. Verbraeken, H. Sprenger, S. De Meyer, K. Demuynck, B. Cannoot, J. Merchie, J. De Block, J. T. Vogel, W. Bruce, H. Nelissen, S. Maere, D. Inz©, and N. Wuyts, "Monitoring of drought stress and transpiration rate using proximal thermal and hyperspectral imaging in an indoor automated plant phenotyping platform," *Plant Methods*, vol. 19, no. 1, p. 132, Nov. 2023. [Online]. Available: https://plantmethods.biomedcentral.com/articles/10.1186/s13007-023-01102-1

[37] J. Zhuang, Q. Wang, G. Song, and J. Jin, "Validating and Developing Hyperspectral Indices for Tracing Leaf Chlorophyll Fluorescence Parameters under Varying Light Conditions," *Remote Sensing*, vol. 15, p. 4890, Oct. 2023.

[38] W. M. Laprade, P. Pieta, S. Kutuzova, J. C. Westergaard, M. Nielsen, S. Christensen, and A. B. Dahl, "HyperLeaf2024 – a hyperspectral imaging dataset for classification and regression of wheat leaves," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2024, pp. 1234–1243. [Online]. Available: https://ieeexplore.ieee.org/document/10678607/

[39] Abanoublamie. (2021) Wild oats. [Online]. Available: https://www.kaggle.com/datasets/abanoublamie/wild-oats

[40] A. Hassan, R. Mumtaz, Z. Mahmood, M. Fayyaz, and M. K. Naeem, "Wheat leaf localization and segmentation for yellow rust disease detection in complex natural backgrounds," *Alexandria Engineering Journal*, vol. 107, pp. 786–798, Nov. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016824010329

[41] F. Kriegler, W. Malila, R. Nalepka, and W. Richardson, "Preprocessing transformations and their effects on multispectral recognition." [Online]. Available: https://www.semanticscholar.org/

paper/Preprocessing-Transformations-and-Their-Effects-on-Kriegler-Malila/aa0f092bba168a62cc913d027a4a892cac3e8e94

[42] M. Vincini, E. Frazzi, and P. D'Alessio, "A broad-band leaf chlorophyll vegetation index at the canopy scale," vol. 9, pp. 303–319, publisher: Springer. [Online]. Available: https://publires.unicatt.it/en/publications/a-broad-band-leaf-chlorophyll-vegetation-index-at-the-canopy-scal-8

[43] A. A. Gitelson, Y. Gritz †, and M. N. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," vol. 160, no. 3, pp. 271–282. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0176161704704034

[44] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," vol. 58, no. 3, pp. 289–298. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425796000727

[45] D. M. Woebbecke, G. E. Meyer, K. Von Bargen, and D. A. Mortensen, "Color indices for weed identification under various soil, residue, and lighting conditions," vol. 38, no. 1, pp. 259–269. [Online]. Available: https://www.scopus.com/pages/publications/0029110322

[46] E. M. Barnes, T. R. Clarke, and S. E. Richards, "COINCIDENT DETECTION OF CROP WATER STRESS, NITROGEN STATUS AND CANOPY DENSITY USING GROUND-BASED MULTISPECTRAL DATA."

[47] C. F. Jordan, "Derivation of leaf-area index from quality of light on the forest floor," vol. 50, no. 4, pp. 663–666, _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/1936256. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.2307/1936256

[48] M. Du and N. Noguchi, "Monitoring of wheat growth status and mapping of wheat yield's within-field spatial variations using color images acquired from UAV-camera system," vol. 9, no. 3, p. 289, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/9/3/289

[49] J.-f. H. X.-z. W. Fu-Min Wang, Yan-Lin Tang, "(PDF) new vegetation index and its application in estimating leaf area index of rice." [Online]. Available: https://www.researchgate.net/publication/240446353_New_Vegetation_Index_and_Its_Application_in_Estimating_Leaf_Area_Index_of_Rice

[50] D. M. El-Shikha, E. M. Barnes, T. R. Clarke, D. J. Hunsaker, J. A. Haberland, P. J. Pinter Jr., P. M. Waller, and T. L. Thompson, "Remote sensing of cotton nitrogen status using the canopy chlorophyll content index (CCCI)," vol. 51, no. 1, pp. 73–82. [Online]. Available: http://elibrary.asabe.org/abstract.asp??JID=3&AID=24228&CID=t2008&v=51&i=1&T=1

[51] T. Motohka, K. N. Nasahara, H. Oguma, and S. Tsuchida, "Applicability of green-red vegetation index for remote sensing of vegetation phenology," vol. 2, no. 10, pp. 2369–2387, publisher: Molecular Diversity Preservation International. [Online]. Available: https://www.mdpi.com/2072-4292/2/10/2369

[52] M. N. Merzlyak, A. A. Gitelson, O. B. Chivkunova, and V. Y. Rakitin, "Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening," vol. 106, no. 1, pp. 135–141, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1034/j.1399-3054.1999.106119.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1399-3054.1999.106119.x

[53] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," vol. 8, no. 2, pp. 127–150. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0034425779900130

[54] B. F. Penuelas, J. and I. Filella, "(PDF) semi-empirical indices to assess carotenoids/chlorophyll-a ratio from leaf spectral reflectance." [Online]. Available: https://www.researchgate.net/publication/229084513_Semi-Empirical_Indices_to_Assess_CarotenoidsChlorophyll-a_Ratio_from_Leaf_Spectral_Reflectance

[55] J. Zdrazil, L. Kong, P. Klimeš, F. I. Jasso-Robles, I. Saiz-Fernández, F. Güder, L. Spíchal, V. Snášel, and N. D. Diego, "Decrypting the complex phenotyping traits of plants by machine learning," Nov. 2024, pages: 2024.11.14.623623 Section: New Results. [Online]. Available: https://www.biorxiv.org/content/10.1101/2024.11.14.623623v1

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, ISSN: 1063-6919. [Online]. Available: https://ieeexplore.ieee.org/document/7780459

[57] K. Shaheed, I. Qureshi, F. Abbas, S. Jabbar, Q. Abbas, H. Ahmad, and M. Z. Sajid, "EfficientRMT-net—an efficient ResNet-50 and vision transformers approach for classifying potato plant leaf diseases," vol. 23, no. 23, p. 9516,

publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1424-8220/23/23/9516

[58] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2020. [Online]. Available: http://arxiv.org/abs/1905.11946

[59] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9992–10 002, ISSN: 2380-7504. [Online]. Available: https://ieeexplore.ieee.org/document/9710580

[60] F. J. R.-R. Borja Espejo García, Evangelos Anastasiou and H. Panoutsopoulos, "Top-tuning on transformers and data augmentation transferring for boosting the performance of weed identification | request PDF." [Online]. Available: https://www.researchgate.net/publication/372317665_Top-tuning_on_Transformers_and_Data_Augmentation_Transferring_for_boosting_the_performance_of_Weed_Identification

[61] A. Triki, B. Bouaziz, J. Gaikwad, and W. Mahdi, "Deep leaf: Mask R-CNN based leaf detection and segmentation from digitized herbarium specimen images," *Pattern Recognition Letters*, vol. 150, pp. 76–83, Oct. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167865521002361

[62] J. Weyler, A. Milioto, T. Falck, J. Behley, and C. Stachniss, "Joint plant instance detection and leaf count estimation for in-field plant phenotyping," vol. 6, no. 2, pp. 3599–3606, 2021, conference Name: IEEE Robotics and Automation Letters. [Online]. Available: https://ieeexplore.ieee.org/document/9359471

[63] R. Khanam and M. Hussain, "YOLOv11: An overview of the key architectural enhancements," 2024. [Online]. Available: http://arxiv.org/abs/2410.17725

[64] G. Roggiolani, F. Magistri, T. Guadagnino, J. Behley, and C. Stachniss, "Unsupervised pre-training for 3d leaf instance segmentation," vol. 8, no. 11, pp. 7448–7455, 2023. [Online]. Available: http://arxiv.org/abs/2401.08720

[65] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "SAM 2: Segment anything in images and videos," 2024. [Online]. Available: http://arxiv.org/abs/2408.00714

[66] J. Pinowski and R. Zając, "Damages to crops caused by birds in central europe," in *Proceedings of the General Meetings of the Working Group on Granivorous Birds*, 1990, pp. 333–347.

[67] C. Tian, J. Wang, D. Zheng, Y. Li, and X. Zhang, "Oat ears detection and counting model in natural environment based on improved faster r-CNN," vol. 15, no. 3, p. 536, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2073-4395/15/3/536

[68] W. Q, F. X, Z. Z, T. T, J. S, H. H, and Y. Q, "One to all: Toward a unified model for counting cereal crop heads based on few-shot learning," vol. 6, publisher: Plant Phenomics. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/39678648/

[69] M. Cai, H. Deng, J. Cai, W. Guo, Z. Hu, D. Yu, and H. Zhang, "Lightweight highland barley detection based on improved YOLOv5," vol. 21, no. 1, p. 42. [Online]. Available: https://doi.org/10.1186/s13007-025-01353-0

[70] Q. Jiang, F. Li, T. Ren, S. Liu, Z. Zeng, K. Yu, and L. Zhang, "T-rex: Counting by visual prompting." [Online]. Available: http://arxiv.org/abs/2311.13596

[71] J. F. Petrie, "Farmbot ROS2 Controller," urlhttps://github.com/PetriJF/FarmBot$_{R}OS2$, 2024, [$Online; accessed 19 - July - 2025$].

[72] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data," Apr. 2024, arXiv:2401.10891 [cs]. [Online]. Available: http://arxiv.org/abs/2401.10891

[73] L. Medeiros, "Lang-segment-anything: Sam with text prompt," https://github.com/luca-medeiros/lang-segment-anything, 2023, accessed: 2025-06-04.

[74] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment Anything," Apr. 2023, arXiv:2304.02643 [cs]. [Online]. Available: http://arxiv.org/abs/2304.02643

[75] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, Q. Jiang, C. Li, J. Yang, H. Su, J. Zhu, and L. Zhang, "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection," Jul. 2024, arXiv:2303.05499 [cs]. [Online]. Available: http://arxiv.org/abs/2303.05499

[76] C. Chen, X. Yuan, S. Gan, W. Luo, R. Bi, R. Li, and S. Gao, "A new vegetation index based on UAV for extracting plateau vegetation information," vol. 128, p. 103668. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569843224000220

[77] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. [Online]. Available: https://ieeexplore.ieee.org/document/1284395

[78] V. Lebourgeois, A. Bégué, S. Labbé, B. Mallavan, L. Prévot, and B. Roux, "Can commercial digital cameras be used as multispectral sensors? a crop monitoring test," vol. 8, no. 11, p. 7300. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3787446/

[79] J. D. Stamford, S. Vialet-Chabrand, I. Cameron, and T. Lawson, "Development of an accurate low cost NDVI imaging system for assessing plant health," vol. 19, no. 1, p. 9. [Online]. Available: https://doi.org/10.1186/s13007-023-00981-8

[80] N. J. Cuper, J. H. G. Klaessens, J. E. N. Jaspers, R. de Roode, H. J. Noordmans, J. C. de Graaff, and R. M. Verdaasdonk, "The use of near-infrared light for safe and effective visualization of subsurface blood vessels to facilitate blood withdrawal in children," vol. 35, no. 4, pp. 433–440. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350453312001427

[81] Y. Liu, K. Hatou, T. Aihara, S. Kurose, T. Akiyama, Y. Kohno, S. Lu, and K. Omasa, "A robust vegetation index based on different UAV RGB images to estimate SPAD values of naked barley leaves," vol. 13, no. 4, p. 686, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2072-4292/13/4/686

[82] M. M. Conley, A. L. Thompson, and R. Hejl, "Proximal active optical sensing operational improvement for research using the CropCircle ACS-470, implications for measurement of normalized difference vegetation index (NDVI)," vol. 23, no. 11, p. 5044, 2023, publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1424-8220/23/11/5044

[83] S. Sunoj, C. Igathinathane, N. Saliendra, J. Hendrickson, and D. Archer, "Color calibration of digital images for agriculture and other applications," vol. 146, pp. 221–234, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271618302600

[84] A. Haghighattalab, L. González Pérez, S. Mondal, D. Singh, D. Schinstock, J. Rutkoski, I. Ortiz-Monasterio, R. P. Singh, D. Goodin, and J. Poland, "Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries," vol. 12, no. 1, p. 35, 2016. [Online]. Available: https://doi.org/10.1186/s13007-016-0134-6

[85] D. Williams, A. Britten, S. McCallum, H. Jones, M. Aitkenhead, A. Karley, K. Loades, A. Prashar, and J. Graham, "A method for automatic segmentation and splitting of hyperspectral images of raspberry plants collected in field conditions," vol. 13, no. 1, p. 74, 2017. [Online]. Available: https://doi.org/10.1186/s13007-017-0226-y

[86] M. Tausen, M. Clausen, S. Moeskjær, A. S. M. Shihavuddin, A. B. Dahl, L. Janss, and S. U. Andersen, "Greenotyper: Image-based plant phenotyping using distributed computing and deep learning," vol. 11, 2020, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2020.01181/full

[87] M. P. Pound, J. A. Atkinson, D. M. Wells, T. P. Pridmore, and A. P. French, "Deep learning for multi-task plant phenotyping," 2017.

[88] K. Seidenthal, K. Panjvani, R. Chandnani, L. Kochian, and M. Eramian, "Iterative image segmentation of plant roots for high-throughput phenotyping," vol. 12, no. 1, p. 16563, 2022, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-022-19754-9

[89] C. Kim and M. W. van Iersel, "Image-based phenotyping to estimate anthocyanin concentrations in lettuce," vol. 14, 2023, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1155722/full