



FRANK GYAN OKYERE

COMPUTER VISION METHODS FOR THE ANALYSIS OF MULTIMODAL  
AND MULTIDIMENSIONAL DATA FOR HIGH-THROUGHPUT PLANT  
PHENOTYPING

SCHOOL OF WATER, ENERGY AND ENVIRONMENT

Research Degree

Doctor of Philosophy (PhD)

Academic Year: 2019-2023

Supervisors:

Professor Fady Mohareb

Professor Malcolm Hawkesford

Dr. Daniel Simms



SCHOOL OF WATER, ENERGY AND ENVIRONMENT

Research Degree

PhD

Academic Year 2019 - 2023

FRANK GYAN OKYERE

Supervisors:

Professor Fady Mohareb

Professor Malcolm Hawkesford

Dr. Daniel Simms

2023

This thesis is submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy.

© Cranfield University 2023. All rights reserved. No part of this publication may  
be reproduced without the written permission of the copyright owner.

## **Academic integrity declaration**

I declare that:

- The thesis submitted has been written by me alone.
- The thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and secondary data, is accurate to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that knowingly submitting work in violation of the above statement will be considered by examiners as academic misconduct.

## **Abstract**

Drought and nutrient stresses substantially impact crop productivity, frequently resulting in lower yields and financial losses to farmers. Early detection and tracking of these stresses are significant for improving agricultural yield and achieving the world food challenge goals. Advanced image technologies such as the use of conventional cameras and imaging spectroscopy combined with computer vision methodologies are being leveraged for non-invasive plant phenotyping to improve crop productivity, resilience, and sustainability. This thesis seeks to combine these technologies with machine learning algorithms to develop automated phenotyping pipelines to phenotype plants at different growth stages, particularly; for nutrient and drought stress identification and quantification. With this proposal, the onset of plant nutrient and drought stresses could be detected early and at different growth stages of the plants. Three experiments (two in the glasshouse and one in the field) were conducted, and images were acquired using digital RGB and a hyperspectral camera. The first experiment (nutrient stress) was performed in the glasshouse, the second (nutrient stress) on the field and the third (drought stress) in the glasshouse. The glasshouse nutrient experiment was performed on quinoa and cowpea plants made of four treatments: high nitrogen high phosphorus (HNHP), high nitrogen low phosphorus (HNLP), low nitrogen low phosphorus (LNHP) and low nitrogen low phosphorus (LNLN). The field nutrient experiment was performed on wheat plants made of 12 Olsen phosphorus variations (approximately 3, 6, 9, 12, 15, 18, 21, 25, 30, 40, 50 and 60 ppm). The third experiment is a wheat drought analysis under variable nitrogen at selected plant growth stages. The treatments include: well-watered high-nitrogen (WWHN), well-watered low-nitrogen (WWLN), drought-stress high-nitrogen (DSHN) and drought-stress low-nitrogen (DSLN). Several image processing and machine learning techniques were employed to pre-process, post-process, and analyze plant-specific traits for tracking plant drought and nutrient stresses. Specifically, using digital imaging, a new segmentation algorithm invariant to illumination and complex background scenes was proposed to segment field and glasshouse-based images. Statistical and machine learning methods were employed to identify phenotypic traits sensitive to nutrient (nitrogen and phosphorus) deficiencies. Additionally, plant colour, morphology, and texture features were critically analyzed to assess their response to different stresses in plants. Using a hyperspectral imaging, a hybrid deep learning model



was proposed to detect and track plant nitrogen and phosphorus deficiencies. In this case , the spatial and spectral characteristics of plants were analyzed, and deep learning algorithms were combined to understand their response to nutrient and drought stress in plants. Finally, using the spectral characteristics of plants, different conventional machine learning algorithms including Random Forest, Partial Least Square Regression and Support Vector Machines were developed to model the trends and patterns of plants for drought stress detection. The research results show a link between colour and nutrient stress, while texture and colour features were highly responsive to drought stress. The short-wave infrared region of the electromagnetic spectrum was highly responsive to plant phosphorus deficiency, while the blue, red, and near-infrared regions were highly connected to plant nitrogen deficiency. Furthermore, combining a proposed vegetation indices (VIs) from the VNIR regions of the spectrum with already known VIs resulted in the easy identification of plant drought stress compared to using only the known or proposed indices.

**Keywords:** Drought, Image processing, Machine learning, Nutrient, Plant phenotyping, Plant stresses, Spectroscopy

## **Acknowledgements**

My sincere gratitude goes to my supervisors at Rothamsted Research (Professor Malcolm Hawkesford) and Cranfield University (Professor Fady Mohareb and Dr Daniel Simms), without whom I could not have completed this PhD thesis and its research publications. Their expertise, guidance, feedback and encouragement have been invaluable throughout this project, and I am very grateful to them. To Professor Hawkesford and Professor Mohareb, thank you for encouraging me to finish my PhD. Also, to Dr. Simms, thanks for always bringing on board new ideas and directing me on the right path when the going gets tough. I feel privileged to have every one of you as a mentor.

Thanks to OCP S.A. for funding the overall SAFA (Sustainable Agriculture for Africa) program, of which this PhD project is a beneficiary. A special thanks to Dr Pouria Sadeghi-Tehran, whose guidance and encouragement have been tremendously helpful throughout this project. From the beginning of my PhD, he was a steadfast support system, helping me acquire the necessary skills and experience to get through the challenges of this PhD project. Thank you also to the FP04 members from UM6P (University of Mohammed VI Polytechnic), especially Professor Bruno Gerard, Dr Manal Mhada and Dr Manal Akrouch, who assisted me in diverse ways in completing this project.

I want to thank Dr. Nicolas Virlet and Dr. Latifa Greche for their dedication to my cause, especially in writing and publishing two scientific papers from this thesis. I appreciate all the nights and chilly days they dedicated to working on my research publications and thesis. Words cannot express my gratitude to them, and I wish them the best in their personal and professional careers. Also, thanks to all the members of Professor Malcolm's team at Rothamsted who significantly affected my research and thesis, especially March Castle, without whom the glasshouse experiments would not have been performed. A big shoutout to Ian Truckell of Cranfield University, who ensured I accessed my images in real-time. Thank you to Professor Fady's group (Brix), whose monthly meeting helped shape my research and presentation skills.

To my wonderful wife, Mrs Mabel Nyadedzor, who encouraged and patiently helped me get through the challenging moments of my PhD journey, I say thank you. To my lovely son Nana Adjei Adepa Okyere-Gyan, the success of this project is dedicated to you.

## Table of Contents

|                                                                                                                                                |       |
|------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| Academic integrity declaration .....                                                                                                           | i     |
| Abstract .....                                                                                                                                 | ii    |
| Acknowledgements .....                                                                                                                         | iv    |
| Table of Contents .....                                                                                                                        | v     |
| List of Figures .....                                                                                                                          | ix    |
| List of Equations .....                                                                                                                        | xvii  |
| List of Tables .....                                                                                                                           | xviii |
| List of Abbreviations .....                                                                                                                    | xxi   |
| Executive Summary .....                                                                                                                        | xxv   |
| 1 Introduction .....                                                                                                                           | 1     |
| 1.1 Background .....                                                                                                                           | 1     |
| 1.2 Image processing for plant phenotyping .....                                                                                               | 2     |
| 1.2.1 Imaging sensors .....                                                                                                                    | 2     |
| 1.2.2 RGB visible image processing .....                                                                                                       | 4     |
| 1.2.3 Hyperspectral image processing .....                                                                                                     | 8     |
| 1.2.4 Phenotyping nutritional status of plants .....                                                                                           | 12    |
| 1.2.5 Image-based phenotyping of plant drought stress .....                                                                                    | 16    |
| 1.3 Significance of the Research Problem .....                                                                                                 | 20    |
| 1.5 Hypothesis Statement .....                                                                                                                 | 21    |
| 1.6 Statement of Objectives .....                                                                                                              | 21    |
| 1.7 References .....                                                                                                                           | 22    |
| 2 Machine Learning Methods for Automatic Segmentation of Images of Field- and<br>Glasshouse-Based Plants for High-Throughput Phenotyping ..... | 30    |
| 2.1 Abstract .....                                                                                                                             | 30    |
| 2.2 Introduction .....                                                                                                                         | 31    |
| 2.3 Materials and Methods .....                                                                                                                | 34    |
| 2.3.1 Experimental Setup and Imaging .....                                                                                                     | 35    |
| 2.3.2 White balance and colour correction of plant images .....                                                                                | 37    |
| 2.3.3 Machine learning based segmentation process .....                                                                                        | 37    |
| 2.3.4 Regression for SPAD values predictions .....                                                                                             | 44    |
| 2.3.5 Comparing segmentation performance .....                                                                                                 | 45    |
| 2.3.6 Evaluating classification model performance .....                                                                                        | 47    |
| 2.3.7 Assessment of performance of SPAD prediction models .....                                                                                | 49    |
| 2.4 Results .....                                                                                                                              | 50    |
| 2.4.1 Performance of the ML Models Based on Different Number of Training<br>Features .....                                                     | 50    |
| 2.4.2 Segmentation Accuracy Analysis .....                                                                                                     | 52    |
| 2.4.3 Performance of SPAD prediction regression models .....                                                                                   | 54    |
| 2.5 Discussion .....                                                                                                                           | 62    |
| 2.6 Conclusions .....                                                                                                                          | 65    |

|                                                                                                                                                           |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.7 References .....                                                                                                                                      | 66  |
| 2.8 Additions to Published Article .....                                                                                                                  | 73  |
| 2.8.1 Challenges with the Proposed Segmentation Algorithm .....                                                                                           | 73  |
| 2.8.2 Developing the improved segmentation tool .....                                                                                                     | 74  |
| 2.8.3 Machine learning-based segmentation process .....                                                                                                   | 75  |
| 2.8.4 Image segmentation using state-of-the-arts CNN models .....                                                                                         | 80  |
| 2.8.5 Dataset for training the model .....                                                                                                                | 81  |
| 2.8.6 Training the networks .....                                                                                                                         | 82  |
| 2.8.7 Evaluating the performance of the models .....                                                                                                      | 84  |
| 2.8.8 Results and Discussions .....                                                                                                                       | 85  |
| 2.8.9 Conclusion.....                                                                                                                                     | 90  |
| 2.8.10 Reference.....                                                                                                                                     | 90  |
| 3 Selection of Nitrogen and Drought-Responsive Plant Phenotypic Traits in Wheat<br>Using Machine Learning Methods .....                                   | 92  |
| 3.1 Abstract .....                                                                                                                                        | 92  |
| 3.2 Introduction.....                                                                                                                                     | 93  |
| 3.3 Materials and Methods .....                                                                                                                           | 95  |
| 3.3.1 Experiment setup .....                                                                                                                              | 95  |
| 3.3.2 Low throughput physiological measurement.....                                                                                                       | 96  |
| 3.3.3 Image pre-processing .....                                                                                                                          | 98  |
| 3.3.4 Handcrafted feature extraction .....                                                                                                                | 99  |
| 3.3.5 Sensitivity analysis .....                                                                                                                          | 105 |
| 3.3.6 Model development .....                                                                                                                             | 106 |
| 3.3.7 Statistical analysis and model evaluation .....                                                                                                     | 111 |
| 3.4 Results.....                                                                                                                                          | 112 |
| 3.4.1 Hand-crafted features and physiological measurement.....                                                                                            | 112 |
| 3.4.2 Sensitivity analysis results for the extracted features.....                                                                                        | 116 |
| 3.4.3 Time series analysis of sensitivity for early stress detection.....                                                                                 | 121 |
| 3.4.4 Evaluating the performance of classical ML models. ....                                                                                             | 124 |
| 3.4.5 Evaluating the performance of the proposed CNN models.....                                                                                          | 130 |
| 3.4.6 Performance of the proposed model with hand-crafted features .....                                                                                  | 132 |
| 3.4.7 Performance of the proposed model at different time points .....                                                                                    | 133 |
| 3.5 Discussion.....                                                                                                                                       | 135 |
| 3.6 Conclusions.....                                                                                                                                      | 137 |
| 3.7 References .....                                                                                                                                      | 138 |
| 4 Modelling the spatial-spectral characteristics of plants for nutrient status<br>identification using hyperspectral data and deep learning methods. .... | 142 |
| 4.1 Introduction .....                                                                                                                                    | 143 |
| 4.2 Materials and Methods .....                                                                                                                           | 146 |
| 4.2.1 Plant materials and treatments.....                                                                                                                 | 147 |
| 4.2.2 Hyperspectral imaging system and data collection.....                                                                                               | 148 |
| 4.2.3 Pre-processing of raw spectral data.....                                                                                                            | 148 |

|        |                                                                                                                                     |     |
|--------|-------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.2.4  | Data transformation.....                                                                                                            | 150 |
| 4.2.5  | Dataset for model training .....                                                                                                    | 153 |
| 4.2.6  | Developing models for spatial-spectral characterization of HSI data.....                                                            | 154 |
| 4.2.7  | Optimal wavelength selection .....                                                                                                  | 158 |
| 4.2.8  | Evaluation metrics for model performance .....                                                                                      | 159 |
| 4.3    | Results.....                                                                                                                        | 160 |
| 4.3.1  | Characteristics of spectral curves for different nutrient treatment .....                                                           | 160 |
| 4.3.2  | Model performance based on different pre-processing methods .....                                                                   | 162 |
| 4.3.3  | Effective wavelength selection .....                                                                                                | 167 |
| 4.3.4  | Results of discrimination analysis at different growth stages .....                                                                 | 169 |
| 4.4    | Discussion.....                                                                                                                     | 172 |
| 4.5    | Conclusion .....                                                                                                                    | 175 |
| 4.6    | References .....                                                                                                                    | 176 |
| 5      | Proximal Hyperspectral Imaging for Phenotyping Plant Drought Stress Using<br>Univariate and Multivariate Modelling Techniques ..... | 181 |
| 5.1    | Abstract .....                                                                                                                      | 181 |
| 5.2    | Introduction.....                                                                                                                   | 182 |
| 5.3    | Materials and Methods .....                                                                                                         | 185 |
| 5.3.1  | Experiment setup .....                                                                                                              | 187 |
| 5.3.2  | Physiological measurements of drought stress.....                                                                                   | 188 |
| 5.3.3  | Hyperspectral Data Acquisition.....                                                                                                 | 189 |
| 5.3.4  | Hyperspectral data pre-processing .....                                                                                             | 189 |
| 5.3.5  | Segmenting the hyperspectral data.....                                                                                              | 190 |
| 5.3.6  | Extracting Known Vegetation Indices .....                                                                                           | 191 |
| 5.3.7  | Wavelength selection and new drought stress indices .....                                                                           | 194 |
| 5.3.8  | Machine learning models for drought stress identification.....                                                                      | 196 |
| 5.3.9  | Multivariate analysis for stomatal conductance and photosynthetic rate<br>predictions .....                                         | 197 |
| 5.3.10 | Model Training and Testing .....                                                                                                    | 197 |
| 5.4    | Results.....                                                                                                                        | 200 |
| 5.4.1  | Reference data of gas exchange measurements .....                                                                                   | 200 |
| 5.3.2  | Spectral Reflectance Analysis .....                                                                                                 | 203 |
| 5.4.2  | Correlation between the known VIs and gas exchange measurements ( $P_n$<br>and $g_s$ ) .....                                        | 204 |
| 5.4.3  | Waveband selection and proposed indices .....                                                                                       | 206 |
| 5.4.4  | Machine learning-based drought detection.....                                                                                       | 210 |
| 5.5    | Discussion.....                                                                                                                     | 216 |
| 5.6    | Conclusion .....                                                                                                                    | 219 |
| 5.7    | Reference.....                                                                                                                      | 220 |
| 6      | General Discussions .....                                                                                                           | 228 |
| 6.1    | Plant stress detection using image processing systems.....                                                                          | 229 |
| 6.1.1  | Digital imaging for stress analysis .....                                                                                           | 229 |

|                                                                                |     |
|--------------------------------------------------------------------------------|-----|
| 6.1.2 Hyperspectral imaging for stress analysis .....                          | 232 |
| 6.2 Machine learning and statistical methods for plant phenotyping .....       | 238 |
| 6.2.1 Feature Extraction and Selection .....                                   | 239 |
| 6.2.2 Trait analysis for stress classification and prediction.....             | 240 |
| 6.2.3 Selecting ML and deep learning models for plant phenotyping .....        | 241 |
| 6.3 Research implications- novelty, impact and contribution to knowledge ..... | 242 |
| 6.4 Research Implication to Africa’s Agriculture .....                         | 243 |
| 6.5 References .....                                                           | 244 |
| 7 Conclusions and recommendations.....                                         | 248 |
| 7.1 Summary .....                                                              | 248 |
| 7.2 Key Objectives and Findings .....                                          | 248 |
| 7.3 Recommendations for Future Work .....                                      | 251 |
| 7.4 Concluding Remarks .....                                                   | 252 |
| Appendices .....                                                               | 253 |
| Appendix A Ethical approval letter .....                                       | 253 |
| Appendix B Supplementary Materials for Chapter 2 .....                         | 254 |
| Appendix C - Supplementary Materials for Chapter 3 .....                       | 261 |
| Appendix D Supplementary Materials for Chapter 4 .....                         | 265 |
| Appendix E Supplementary Materials for Chapter 5.....                          | 269 |

## List of Figures

- Figure 1-1 Phenotyping pipeline from image acquisition to data analysis (Virlet et al., 2017). This involves data acquisition using an aerial (eg. UAV) or ground based systems (e.g. field scanalyzers) to acquire images. The sensors may include visible RGB camera, hyperspectral, thermal, 3D laser sensors etc. The data are then stored in a data management or cloud server. The stored data can be pre-processed, post processed and analyzed using statistical or machine learning methods..... 4
- Figure 1-2 A typical image processing pipeline for plant phenotyping. This mainly involves four steps; (a) image acquisition using visible RGB cameras, (b) image pre-processing which involves cropping, image enhancement and colour calibration. (c) The images are then segmented to remove background from foreground or isolate the region of interest. (d) Feature extraction follows this step where important traits of plants such as colour and texture are extracted for further analysis. .... 5
- Figure 1-3** Example of a hyperspectral image of plant leaf canopy. (a) A stack of narrow band sub-images from 394 nm to 2500 nm covering the visible and near infrared regions as well as the short wave infrared regions. This forms a 3-D hypercube made of the two spatial dimensions (height (h) x width(w) and one spectral dimension ( $\lambda$ ); (b) a spectral signature of a plant pixel spanning from 390 nm to 2500 nm ..... 8
- Figure 1-4** A spectral reflectance curve showing the reflectance properties of a typical green vegetation across different wavelengths of light (Kuska et al., 2018). Machine learning for image-based plant phenotyping. The region around 400-500 nm corresponds to chlorophyll, indicating strong absorption by chlorophyll pigments while the regions around 700 nm -1000 nm shows a high reflectance due to proteins which might be indicative of structural compounds in the plants. The SWIR around 1200 nm -1500 nm is marked for cellulose and lignin while a dip in the the region between 1500 nm-1800 nm shows the responsiveness to changes in water components in the plant canopy. A feature near 2300 nm is attributed to sugar, likely a reference to carbohydrate content. .... 9
- Figure 2-1 Diagrammatic representation of the proposed method for glasshouse and field image, using MLP, multilayer perceptron. The steps involved in this method include: image acquisition and preprocessing, patch extraction (for background (BG) and foreground (FG), colour transformation, feature selection and training a MLP classifier using the annotated pixel features as input. .... 35
- Figure 2-2 Annotation of images into foreground and background patches for feature extraction. (a) Wheat image annotation and (b) Cowpea image annotation. The FG represents the foreground annotation and the BG is the background annotation. Patches were extracted from two crops (wheat from the field and cowpea from the glasshouse)..... 40
- Figure 2-3 Feature-selection process involving correlation analysis and feature ranking based on importance score. (a) is the heatmap of all extracted features and (b) is a plot of feature importance for feature selection. Abbreviations: R, G, B (R = red, G = green and B = blue channels); H, S, V (H = hue, S = saturation and V = value); ybr

(y = luma, b = blue component and r = red component); Lab (L = lightness and a and b = chromaticity); YUV (Y = luma or brightness, U = blue projection and V = red projection); Luv (L = luminance, u = blue axis and v = red axis); hls (h = hue, l = lightness and s = saturation); and XYZ (X and Z = spectral weighting curves and Y = luminance). ..... 41

Figure 2-4 Examples of images obtained from the glasshouse and field, segmented as reference images. The manually segmented images were obtained using Adobe Photoshop software (a) Original image, (b) manually segmented image and (c) binary image. Reference images were randomly selected from the dataset with different nutrient content and illumination and at variable growth stages..... 47

Figure 2-5 Classification accuracy scores for the three models (multilayer perceptron (MLP), random forest (RF) and support vector machines (SVMs). This accuracy score analysis was based on the type of dataset used for the training the model. The datasets include; Selected Features dataset (SF) and All Features (AF) dataset. The SF dataset was obtained after feature selection using feature importance based selection technique while the AF dataset comprise all the extracted features obtained earlier. .... 51

Figure 2-6 Examples of glasshouse and field-segmented plants using the proposed method and selected segmentation methods. (a) Original wheat image, (b) ExG wheat segmented image, (c) proposed method (MLP) segmented image, (d) original cowpea image, (e) ExG cowpea and (f) MLP segmented image. The machine learning based methods including the proposed MLP were trained with the SF datasets while for the ExG and ExGR techniques, established colour indices together with Otsu thresholding were used..... 52

**Figure 2-7** Mean segmentation accuracy rate comparison for quality assessment of five different segmentation methods for glasshouse-based images; Qseg measures the segmentation consistency on a pixel-by-pixel basis, Sr measures the consistency of plant pixels between image regions and Es measures the rate of pixel misclassification. These were applied on the five segmentation methods: multilayer perceptron (MLP), support vector machines (SVMs), random forest (RF), excess green (ExG) and excess green–red (ExGR). ..... 53

Figure 2-8 Comparison of segmentation accuracy rate (Qseg-, Sr, Es) for quality assessment of five different segmentation methods; multilayer perceptron (MLP), support vector machines (SVMs), random forest (RF), excess green (ExG) and excess green–red (ExGR) for field-based images. Qseg measures the segmentation consistency on a pixel-by-pixel basis, Sr measures the consistency of plant pixels between image regions and Es measures the rate of pixel misclassification. These were applied to the five segmentation methods. .... 54

Figure 2-9 Scatterplots of predicted and observed SPAD values for the glasshouse plants. MLP-GR, RF-GR, SVM-GR, ExGR-GR and ExG-GR represent the glasshouse-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green–red segmentation methods, respectively. All the regressions methods were performed using the XGBoost regression technique and evaluated using the  $R^2$ , MAE and RMSE metrics..... 58



|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 2-10 Scatterplots of predicted and observed SPAD values for the field plants. MLP-FR, RF-FR, SVM-FR, ExGR-FR and ExG-FR are the field-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green and excess green and excess green–red segmentation methods, respectively. All the regressions methods were performed using the XGBoost regression technique and evaluated using the $R^2$ , MAE and RMSE metrics. .... | 62 |
| Figure 2-11 Plant image with stress visual colour similar to the background colour ....                                                                                                                                                                                                                                                                                                                                                                                                         | 74 |
| Figure 2-12 Workflow of the improved segmentation algorithm .....                                                                                                                                                                                                                                                                                                                                                                                                                               | 75 |
| Figure 2-13 Diagrammatic explanation of the LBP feature extraction .....                                                                                                                                                                                                                                                                                                                                                                                                                        | 77 |
| Figure 2-14 Plot showing feature combinations and resulting output (the short-dashed lines show the optimum number of features relevant for the binary classification of pixels .....                                                                                                                                                                                                                                                                                                           | 79 |
| Figure 2-15 A plot of feature importance for feature selection using the decision-tree-based recursive feature.....                                                                                                                                                                                                                                                                                                                                                                             | 80 |
| Figure 2-16 Sample dataset images (A) Glasshouse quinoa images with different lightening conditions and superimposed leaves (b) Field images (wheat) with different non-plant matter (sand, rocks, etc.) (C) Sugar beet datasets with different background matter and leaves colouration.....                                                                                                                                                                                                   | 82 |
| Figure 2-17 Comparison of MLP <sub>p</sub> and MLP <sub>i</sub> models based on accuracy, F-score, and recall. Note: MLP models were trained with 13 features based on only the colour information of plants, while the MLP <sub>i</sub> were trained with 15 features based on plants' colour and texture information. Error bars showing only the positive component correspond to one standard deviation.....                                                                                | 85 |
| <b>Figure 2-18</b> Test results on selected glasshouse images (A) original images (B) Corresponding ground truth images (C) MLP <sub>i</sub> predicted masks (D) U-Net predicted masks (E) VGG-16 segmented images (masks). ....                                                                                                                                                                                                                                                                | 87 |
| Figure 2-19 Test results on selected field images (wheat) (A) original images (B) Corresponding ground truth images (C) MLP <sub>i</sub> predicted masks (D) U-Net predicted masks (E) VGG-16 segmented images (masks). ....                                                                                                                                                                                                                                                                    | 88 |
| Figure 2-20 . Test results on selected sugar beet images (A) original images (B) Corresponding ground truth images (C) MLP <sub>i</sub> predicted masks (D) U-Net predicted masks (E) VGG-16 segmented images (masks). ....                                                                                                                                                                                                                                                                     | 89 |
| Figure 3-1 Visualization of image samples representing the top and side views of the plants. The top view images were taken at 0° to the axis of the camera and 3 m from away from the camera whilst the side view images were taken at 90° from the camera axis and 1.5 m away from the ground. A colour checker was placed beside the plants during each image acquisition for colour correction using a custom made algorithm. ....                                                          | 98 |

- Figure 3-2 Illustration of example of Gray Level Co-occurrence Matrix (GLCM) Calculations form an image matrix. The image is represented by a grid of pixel values, with each number representing a gray-level intensity ranging from 1 to 8. The GLCM matrix is constructed by counting the frequency of a specific pixel intensity value is next to a pixel with a different intensity value (e.g., 2). This sample appears to depict a 0-degree angle (horizontal adjacency). Each GLCM element represents the number of times a given intensity pair (for example, 1 and 2) occurs in the specified direction. .... 101
- Figure 3-3 A representation of the proposed convolutional neural network (CNN) architecture used for this study. This is made of two sections; (A) the CNN section where plant images serves as input to the model and (B) an 11-element handcrafted feature vector extracted from the plants are integrated with the CNN based features to train the fully connected layers. .... 111
- Figure 3-4 Measured physiological traits of plants: (a) nitrogen content, (b) stomatal conductance, (c) photosynthetic net assimilation. Groups represented by bars with the same letter (e.g., a, b, c) are not statistically different from each other at a significance level of 5% ( $p = 0.05$ ). Bars with different letters indicate statistically significant differences between groups. .... 113
- Figure 3-5 Pearson correlation coefficients of extracted features with (a) stomatal conductance, (b) nitrogen content and (c) CO<sub>2</sub> net assimilation. The extracted features include ASM: Angular second moment; CONT: contour area; DISS: Dissimilarity; PAR: Plant arial, HULL: Hull area, PAD: Plant area density; SIFT: Scale-invariant feature transform; HOG: Histogram of oriented gradient; GPB: Green Blue; RGBVI: Red, green blue vegetation index; DGCI: dark green colour index. .... 115
- Figure 3-6 The biplot of the principal component analysis (PCA) of the hand-crafted features (ASM: Angular second moment; CONT: contrast ; DISS: Dissimilarity; PAR: Plant arial ratio, HULL: Hull area, PAD: Plant arial density; SIFT: Scale-invariant feature transform; HOG: Histogram of oriented gradient; GMB: Green Blue; RGBVI: Red, green blue vegetation index; DGCI: dark green colour index) (a) is the biplot showing distribution loading vectors (b) trait importance depicting the contribution of variance to the principal component PC1 vs PC2 accounting for 74.9% of the variability. .... 116
- Figure 3-7 Sensitivity analysis of the input variables on the three levels of stresses (a. nitrogen stress, b. drought stress, c. combined stress) using RF based on the OAT method. The features of each model are instances where those features were not part of the training dataset, and the ALL is the combined features used for training the model. The short dashed red, yellow and blue lines indicate the baseline for comparing the different models using the  $R^2$ , MAE AND RMSE metrics. DGCI: dark green colour index, GPB: green blue colour Index, RGBVI: red, green blue vegetation index, ASM: angular second moment, CONT: contrast, DISS: dissimilarity. PAR: plant aspect ratio, PAD: plant area density, HULL: convex hull area, SIFT: scale invariant feature transform. HOG: histogram of oriented gradient. All: all the features combined. .... 119

Figure 3-8 Morris based method for sensitivity analysis of plants under different stress conditions: (a) Nitrogen sensitivity, (b) Drought sensitivity, and (c) Combined Stress sensitivity. Each plot uses two variables on its axes:  $\mu^*$  (x-axis): represents the mean sensitivity while  $\sigma$  (y-axis) represents the standard deviation or variability in sensitivity. DGCI: dark green colour index, GPB: green blue colour Index, RGBVI: red, green blue vegetation index, ASM: angular second moment, CONT: contrast, DISS: dissimilarity. PAR: plant aspect ratio, PAD: plant area density, HULL: convex hull area, SIFT: scale invariant feature transform. HOG: histogram of oriented gradient ..... 120

Figure 3-9 Time series analysis of the ASM feature from 0 DADS to 15 DADS. This was done to understand the dynamic effect of the feature as when plants were subjected to water stress under varying nitrogen conditions. The significant tests are conducted at 3-day and 6-day intervals. Bars followed by the same letter do not differ statistically by Tukey's test at  $p=0.05$ ..... 123

Figure 3-10 Plot of model performances (accuracy) for top and side views of plant features. Accuracy range is from 0-100% with models close to 100% indicating high accuracy and close to 0% indicating low accuracy. DGCI: dark green colour index, GPB: green blue colour Index, RGBVI: red, green blue vegetation index, ASM: angular second moment, CONT: contrast, DISS: dissimilarity. PAR: plant aspect ratio, PAD: plant area density, HULL: convex hull area, SIFT: scale invariant feature transform. HOG: histogram of oriented gradient. MORP features comprise all the morphology features (PAD, PAR and HULL) whereas COL represents all the colour features (DGCI, GPB and RGBVI). TEX represents the texture features (CONT, DIS and ASM) whilst LD is the local image descriptors (HOG and SIFT). Finally, the All features represents all the features combined. .... 127

Figure 3-11 Confusion matrices depicting classification results of the three models (the proposed CNN, ResNet and DenseNet). This shows the accuracy of the models in classifying plants based on their stress levels. The true class (ground truth) are on the y-axis, while the predicted class are on the x-axis. Each matrix shows the percentage and count of correctly and incorrectly classified samples for four different classes: WWHN WWLN, DSHN and DSLN. (a) shows results for non-segmented plant images while (b) shows results for segmented plant images ..... 132

Figure 3-12 A confusion matrix depicting the classification results for the proposed CNN combined with hand-crafted features. This shows the accuracy of the models in classifying plants based on their stress levels. The true class (ground truth) are on the y-axis, whilst the predicted class are on the x-axis. Each matrix shows the percentage and count of correctly and incorrectly classified samples for four different classes: WWHN WWLN, DSHN and DSLN. (a) shows results for non-segmented plant images while (b) shows results for segmented plant images. .... 133

Figure 3-13 Confusion matrix depicting the classification of the four treatments (WWHN, WWLN, DSHN, DSLN) at selected drought stress stages (0, 9, and 15 DADS). 135

Figure 4-1 Flowchart outlining the process of building the proposed 3D-2D hybrid model and comparing it to the state-of-the-art models for plant nutrient status classification using hyperspectral imaging data. The diagram is divided into two main sections.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Part 1 describes the development and training of different models using hypercube patches for plant nutrient status classification. Part 2 describes the building of different models using the best training datasets for plant nutrient status classification at different growth stages.....                                                                                                                                                                                                                                                        | 147 |
| Figure 4-2 Schematic representation of the proposed 3D-2D hybridCNN model frameworks with the full-spectrum bands as inputs. The green, yellow, red and violet boxes represents the hypercube patches extracted from the HNLP, HNHP, LNHP and LNLP classes respectively. The 3D block has stacks of 3D convolutional layers interspersed with batch normalization and max pooling layers. The 2D component is seperable in nature with max-pooling layers. The dense layers transform the feature maps into a vector suitable for classification. .... | 156 |
| Figure 4-3 Average reflectance spectra for cowpea under different N and P levels obtained at growth stage III. The x-axis represents the wavelengths ranging from approximately 400 nm to 2500 nm, covering the visible to near-infrared (NIR) spectrum whiles the y-axis represents reflectance, indicating how much light is reflected by the plant material at each wavelength. Each line represents the average spectral signature for five plants from each treatment of the experiment.....                                                      | 161 |
| Figure 4-4 Average reflectance spectra for quinoa under different N and P levels obtained at growth stage III. The x-axis represents the wavelengths ranging from approximately 400 nm to 2500 nm, covering the visible to near-infrared (NIR) spectrum whilst the y-axis represents reflectance, indicating how much light is reflected by the plant material at each wavelength. ....                                                                                                                                                                | 162 |
| Figure 4-5 Box and whisker plot of classification accuracy scores for cowpea for the four models: Hybrid-CNN, 3DCNN, 2D CNN and HybridSN. (A) Raw dataset, (B) SNV dataset, (C) LDA dataset, (D) Hybrid CNN dataset .....                                                                                                                                                                                                                                                                                                                              | 164 |
| Figure 4-6 Box and whisker plot of classification accuracy scores for quinoa for the four algorithms: Hybrid-CNN, 3DCNN, 2D CNN and HybridSN. (A) Raw dataset, (B) SNV dataset, (C) LDA dataset, (D) Hybrid CNN dataset .....                                                                                                                                                                                                                                                                                                                          | 164 |
| Figure 4-7 Plot displaying the second derivate reflectance values over a wavelength range from approximately 390 nm to 2500 nm. The vertical black dashed lines are marked at specific wavelength positions, indicating selected wavelengths for cowpea species. The wavelength selection was done using CFS feature selection method. ....                                                                                                                                                                                                            | 168 |
| Figure 4-8 Plot displaying the second derivate reflectance values over a wavelength range from approximately 390 nm to 2500 nm. The vertical black dashed lines are marked at specific wavelength positions, indicating selected wavelengths for quinoa species. The wavelength selection was done using CFS feature selection method.....                                                                                                                                                                                                             | 168 |
| Figure 5-1 The schematic diagram of the methodology for analysing spectral images for drought stress identification (A) is the preprocessing step (involving data calibration, denoising, desampling and segmentation, (B) drought stress-related physiological measurements ( $g_s$ and $P_n$ ), (C) the extraction of known VIs (D) ensemble learning model for selecting sensitive spectral wavelengths (E) the development of                                                                                                                      |     |

classification and regression models for identification of drought stress and prediction of gas exchange traits ( $P_n$  and  $g_s$ ) respectively..... 186

**Figure 5-2** Flowchart outlining the feature selection process applied to hyperspectral data. This workflow is based on an ensemble feature selection technique comprising three base models: correlation feature selection (CFS), chi-square and ReliefF methods. While each based model produces top ten selected features, the Boruta SHAP algorithm selects the final 10 wavelengths. .... 195

**Figure 5-3** Photosynthetic rates from 0 to 15 DADS for four different treatments: WWHN, WWLN, DSHN, and DSLN. The WWHN and WWLN are the water well-watered plants with high and low nitrogen, respectively, while the DSHN and DSLN are the drought-stressed plants with high and low N levels, respectively. The results presented are mean and standard deviations from the original data; the dissimilar lower-case group (a, b, and c) represents a significant difference with  $p < 0.05$ . . 202

**Figure 5-4** Stomatal conductance for the four treatments: WWHN, WWLN, DSHN, and DSLN from 0 to 15 DADS. The WWHN and WWLN are the water well-watered plants with high and low nitrogen, respectively, while the DSHN and DSLN are the drought-stressed plants with high and low N levels, respectively. The results presented are mean and standard deviations from the original data; the dissimilar lower-case group (a, b, and c) represents a significant difference with  $p < 0.05$ . 202

**Figure 5-5** Spectral reflectance of the averaged DSHN, DSLN, WWHN, and WWLN treatments for 0 DADS (a), 6 DADS (b), and 15 DADS (c). Spectral values are shown as mean  $\pm$  standard deviation. The WWHN and WWLN are the water well-watered plants with high and low nitrogen, respectively, while the DSHN and DSLN are the drought-stressed plants with high and low N levels, respectively ..... 204

**Figure 5-6** Pearson correlation between the extracted features and the gas exchange measurements ( $P_n$  and  $g_s$ ). VIs with a correlation of more than 0.5 were selected for further analysis. The heatmap highlights how different vegetation indices correlate with the four treatments (WWHN, WWLN, DSHN, DSLN). Strongly positive correlations (green cells) suggest that these indices are good predictors or indicators for the treatments, whereas negative correlations (red cells) indicate the opposite. .... 206

**Figure 5-7** Spectral correlation matrix of hyperspectral wavelengths represented as a colourmap image of the correlation between all pairs of spectral features from 394 nm to 1015 nm. The colour scale on the right ranges from 0.1 (black) to 1.0 (light peach), with 1.0 indicating a very strong positive correlation (meaning the two wavelengths have very similar reflectance patterns across the data). The Lower values (dark purple/black) represent weaker correlations, (meaning the reflectance values at these wavelengths differ significantly) ..... 208

**Figure 5-8** Correlations heatmap showing the relationships between the proposed indices(RDI, DDI and NDDI) and the two gas exchange measurements ( $P_n$  and  $g_s$ ). The correlation coefficients are between 1 to -1 with the colour gradient ranging from dark teal (high positive correlation) to brown/orange (high negative correlation). .... 210

Figure 5-9 Confusion matrices depicting the performance of the SVM, RF and DNN classifiers trained with (a) known VIs, (b) proposed VIs, (c) combined VIs and (d) PCA transformed features..... 214

**Figure 5-10.** Prediction of plant  $g_s$  using four models (RF, SVR, PLSR, and PR). All the models were trained with the combined VIs except the PLSR, which were trained with the whole spectrum. .... 215

**Figure 5-11** Prediction of plant  $g_s$  using four models (Random Forest regression (RF), Support vector regression (SVR), Partial least square regression (PLSR) and Polynomial regression (PR)). All the models were trained with the combined VIs except the PLSR, which was trained with the whole spectra..... 215

## List of Equations

|                                                                                                                                                                                       |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Equation 2.1 .....                                                                                                                                                                    | 43  |
| Equation 2.2 where $W_u^*$ is the updated weight, $W_o$ is the previous weight, $a$ is learning rate and $\partial Error$ is the derivative of error with respect to the weight. .... | 43  |
| Equation 2.3 .....                                                                                                                                                                    | 46  |
| Equation 2.4 .....                                                                                                                                                                    | 46  |
| Equation 2.5 .....                                                                                                                                                                    | 46  |
| Equation 2.6 .....                                                                                                                                                                    | 48  |
| Equation 2.7 .....                                                                                                                                                                    | 48  |
| Equation 2.8 .....                                                                                                                                                                    | 48  |
| Equation 2.9 .....                                                                                                                                                                    | 76  |
| Equation 2.10 .....                                                                                                                                                                   | 76  |
| Equation 2.11 .....                                                                                                                                                                   | 77  |
| Equation 2.12 .....                                                                                                                                                                   | 84  |
| Equation 2.13 .....                                                                                                                                                                   | 84  |
| Equation 2.14 .....                                                                                                                                                                   | 84  |
| Equation 3.1 .....                                                                                                                                                                    | 99  |
| Equation 3.2 .....                                                                                                                                                                    | 100 |
| Equation 3.3 .....                                                                                                                                                                    | 100 |
| Equation 3.4 .....                                                                                                                                                                    | 102 |
| Equation 3.5 .....                                                                                                                                                                    | 102 |
| Equation 3.6 .....                                                                                                                                                                    | 102 |
| Equation 3.7 .....                                                                                                                                                                    | 103 |
| Equation 3.8 .....                                                                                                                                                                    | 103 |
| Equation 3.9 .....                                                                                                                                                                    | 104 |
| Equation 3.10 .....                                                                                                                                                                   | 104 |
| EEi = $[Yx_1, x_2, \dots, x_i - 1, x_i + \Delta, \dots, x_k - Y(x_1, x_2, \dots, x_k)\Delta]$<br>Equation 3.11                                                                        | 106 |
| Equation 3.12 .....                                                                                                                                                                   | 107 |
| Equation 3.13 .....                                                                                                                                                                   | 108 |

|                     |     |
|---------------------|-----|
| Equation 3.14 ..... | 108 |
| Equation 3.15 ..... | 109 |
| Equation 3.16 ..... | 111 |
| Equation 3.17 ..... | 112 |
| Equation 3.18 ..... | 112 |
| Equation 3.19 ..... | 112 |
| Equation 3.20 ..... | 112 |
| Equation 3.21 ..... | 112 |
| Equation 4.1 .....  | 149 |
| Equation 4.2 .....  | 151 |
| Equation 4.3 .....  | 154 |
| Equation 4.4 .....  | 155 |
| Equation 4.5 .....  | 159 |
| Equation 4.6 .....  | 160 |
| Equation 4.7 .....  | 160 |
| Equation 5.1 .....  | 189 |
| Equation 5.2 .....  | 196 |
| Equation 5.3 .....  | 196 |
| Equation 5.4 .....  | 196 |
| Equation 5.5 .....  | 198 |
| Equation 5.6 .....  | 198 |
| Equation 5.7 .....  | 198 |
| Equation 5.8 .....  | 198 |
| Equation 5.9 .....  | 199 |
| Equation 5.10 ..... | 199 |

## List of Tables

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| Table 2-1 Optimal hyperparameters for the different classification models ..... | 48 |
|---------------------------------------------------------------------------------|----|



|                                                                                                                                                                                                                                                                                                                                                                         |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 2-2 Classification model performance assessment. MLP, RF and SVM are the multilayer perceptron, random forest and support vector machine segmentation models, respectively.....                                                                                                                                                                                   | 51  |
| Table 2-3 Correlation analysis of colour vegetation indices (CVIs) and SPAD readings for cowpea. R, G and B are the average red, green and blue values. The ExG, ExGR, MLP, SVM and RF are the excess green, excess green–red, multilayer perceptron, support vector machines and random forest segmentation methods, respectively. 56                                  |     |
| Table 2-4 Performance of the regression models during training and testing for glasshouse data. MLP-GR, RF-GR, SVM-FR, ExGR-GR and ExG-GR are the glasshouse-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green-red segmentation methods, respectively.57                                      |     |
| Table 2-5 Correlation analysis of colour vegetation indices (CVIs) and measured SPAD readings for wheat. R, G and B are the average red, green and blue values. MLP, RF, SVM, ExGR and ExG are the segmentation methods using multilayer perceptron, random forest, support vector machines, excess green and excess green–red segmentation methods, respectively. .... | 60  |
| Table 2-6 Performance evaluation of the regression models. MLP-FR, RF-FR, SVM-FR, ExGR-FR and ExG-FR represent the field-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green–red methods, respectively. ....                                                                                    | 61  |
| Table 2-7 Performance of MLPi, U-Net and VGG-16 models for the quinoa, wheat and sorghum dataset segmentation .....                                                                                                                                                                                                                                                     | 87  |
| Table 3-1 A Performance of the three classical Machine Learning techniques for stress classification (average accuracy (AA), F-score (FS), Recall score (RS), Time (T)) .....                                                                                                                                                                                           | 128 |
| Table 3-2 Performance of the three classical Machine Learning techniques for stress classification. (average accuracy (AA), F-score (FS), Recall score (RS), Time (T)) .....                                                                                                                                                                                            | 129 |
| Table 3-3 Performance analysis for proposed CNN, ResNet-18 and DenseNet.....                                                                                                                                                                                                                                                                                            | 131 |
| Table 3-4 Performance analysis for proposed CNN, model trained with hand-selected features for segmented and non-segmented images.....                                                                                                                                                                                                                                  | 132 |
| Table 3-5 Performance of the proposed CNN model for each drought stress stage ....                                                                                                                                                                                                                                                                                      | 134 |
| Table 4-1 The number of training, validation and testing samples .....                                                                                                                                                                                                                                                                                                  | 154 |
| Table 4-2 The proposed hybrid-CNN algorithm structure. ....                                                                                                                                                                                                                                                                                                             | 158 |
| Table 4-3 Classification accuracy assessment and computation cost with different models and pre-processing methods for quinoa .....                                                                                                                                                                                                                                     | 165 |
| Table 4-4 Classification accuracy assessment and computation cost with different models and pre-processing methods for cowpea.....                                                                                                                                                                                                                                      | 166 |

|                                                                                                                                                                                |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 4-5 Model performance on selected wavebands for classification of cowpea N and P levels at five growth stages.....                                                       | 170 |
| Table 4-6 Model performance on selected wavebands for classification of quinoa N and P levels at five growth stages.....                                                       | 171 |
| Table 4-7 Summary of confusion matrices created for the proposed model for nutrient status identification at five stages of plant growth.....                                  | 172 |
| Table 5-1 List of selected VIs to monitor plant drought stress. ....                                                                                                           | 192 |
| Table 5-2 Model parameters for tuning and training using random search CV .....                                                                                                | 199 |
| Table 5-3 The top 10 wavelengths selected by the Chi-square, SFM-RF and Boruta feature selection methods and the ensemble model for the selection and ranking of features..... | 209 |
| Table 5-4 Performance metrics for RF, SVM and DNN models for identification of drought stress plants (metrics include Average Accuracy (AA), F1-score, Kappa score).....       | 212 |
| Table 5-5 Performance of regression models for $g_s$ and $P_n$ .....                                                                                                           | 216 |

## **List of Abbreviations**

|        |                                                   |
|--------|---------------------------------------------------|
| AD     | Adaptive Boosting                                 |
| AF     | All Features                                      |
| ANN    | Artificial Neural Network                         |
| ASM    | Angular Second Moment                             |
| CC     | Colourchecker                                     |
| CCD    | charge-coupled devices                            |
| CIVE   | colour index of vegetation extraction             |
| CNN    | Convolutional Neural Network                      |
| Cr     | Red Chrominance                                   |
| CV     | Cross Validation                                  |
| CVI    | Colour Vegetation Indices                         |
| DADS   | Day After Drought Stress                          |
| DAS    | Day After Drought Sowing                          |
| DGCI   | Dark Green Colour Index                           |
| DISS   | Dissimilarity                                     |
| DL     | Deep Learning                                     |
| DSHN   | Drought-Stress High Nitrogen                      |
| DSLN   | Drought-Stress Low-Nitrogen                       |
| DT     | Decision Tree                                     |
| DT-RFE | Decision Tree based Recursive Feature Elimination |

|      |                                              |
|------|----------------------------------------------|
| Es   | Error Factor                                 |
| ExG  | Excess Green                                 |
| ExGR | Excess Green Red                             |
| ExR  | Excess Red                                   |
| FS   | F-score                                      |
| GBRT | gradient-boosting decision trees             |
| GLCM | Grey-Level Co-Occurrence Matrix              |
| GPAR | greenness plant area ratio                   |
| GPB  | green blue Colour Index                      |
| gs   | Stomatal Conductance                         |
| HNHP | High Level Nitrogen High Level Phosphorus    |
| HNLP | High Level Nitrogen Low-Level Phosphorus     |
| HOG  | Histogram Of Oriented Gradient               |
| HSI  | Hyperspectral Imaging                        |
| HSV  | Hue Saturation Value                         |
| HTPP | High Throughput Plant Phenotyping            |
| HTPP | Highthroughput Plant Phenotyping             |
| KNN  | K-Nearest Neighbours                         |
| LAI  | Leaf Area Index                              |
| LDA  | Linear Discriminant Analysis                 |
| LNHP | Low-Level Nitrogen And High Level Phosphorus |
| LNLP | Low-Level Nitrogen Low-Level Phosphorus      |

|                |                                        |
|----------------|----------------------------------------|
| LUV            | Luminance, blue axis and red axis      |
| MAE            | Mean Average Error                     |
| ML             | Machine Learning                       |
| MLP            | Multilayer Perceptron Network          |
| N              | Nitrogen                               |
| NDI            | Normalized difference index            |
| NDII           | Normalized difference drought index    |
| NDVI           | Normalized difference vegetation index |
| NIR            | Near Infrared                          |
| NNI            | Nitrogen Nutrition Index               |
| P              | Phosphorus                             |
| PAR            | Perimeter area rati                    |
| PCA            | Principal Component Analysis           |
| Pn             | net CO <sub>2</sub> assimilation       |
| r              | Correlation coefficient                |
| R <sup>2</sup> | Coefficient of determination           |
| ReLU           | Rectified linear unit                  |
| ResNet         | Residual Network                       |
| RF             | Random Forest                          |
| RFE            | Recursive Feature Elimination          |
| RGB            | Red Green Blue                         |
| RGBVI          | Red, green blue vegetation index       |

|                 |                                                    |
|-----------------|----------------------------------------------------|
| RMSE            | Root Mean Square Error                             |
| SF              | Selected Featurs                                   |
| SIFT            | Scale Invariant Feature Transform                  |
| SPAD            | Soil Plant Analysis Development                    |
| SPAD            | Soil plant analysis development                    |
| SVM             | Support Vector Machines                            |
| SWIR            | Short Wave Infrared                                |
| TBR             | Total plant area and bounding rectangle area ratio |
| TCR             | Convex hull area ratio                             |
| UAS             | Unmannered Ariel Systems                           |
| VI <sub>s</sub> | Vegetation Indices                                 |
| VNIR            | Visible and Near Infrared                          |
| WI              | Water Index                                        |
| WWHN            | Well-Watered High Nitrogen                         |
| WWLN            | Well-Watered Low Nitrogen                          |
| XYZ             | Chromatic coordinates                              |
| YCbCr           | Luminance chrominance                              |
| YUV             | Luma, blue projection and red projection           |

## Executive Summary

The quest to achieve the 2050 world food challenge drives researchers to develop disease-resistant and high-yield crops in the face of existing challenges, including climate change and competition for resources. Although high throughput plant phenotyping enables the capacity for rapid and large-scale evaluation of crop performance, which will help accelerate crop breeding processes, there are currently few methods to measure the physical features of plants in response to biotic and abiotic stresses. The ability to phenotype nutrients and drought stress noninvasively, especially identifying key sensitive traits, is limited for crop breeders and researchers. This work focused on developing image processing and predictive modelling tools to phenotype the nutritional and drought status of selected crops. To achieve this, field and glasshouse-based experiments were conducted using digital and hyperspectral sensors to acquire images simultaneously. Experiment design, processing, analysis and results are documented in this thesis, which is made of seven chapters: the introduction, four research chapters, the general discussion and the conclusion. **Chapter 1** introduces the thesis, giving background and literature reviews of image-based plant phenotyping. The significance of this research, as well as the individual and overall objectives, is presented in this chapter. Chapters 2-5 are the main experiment chapters that exploit the literature review of the topic, materials and methods, results, discussion, and conclusions. **Chapter 2** presents a published paper on novel segmentation techniques for field and glasshouse-based images. It comprises developing and testing a method for processing digital images and analyzing nutrient-stress-sensitive phenotypic traits. This involves developing an automated phenotyping pipeline capable of pre-processing images, segmentation and feature extraction for nutrient stress analysis in selected crops. In addition to the published article culminating from this chapter, further details on how the proposed tools and methods were improved are given. **Chapter 3** presents a phenotyping framework for identifying drought stress in wheat using digital RGB images. It involves processing and extracting features using the pipeline in Chapter 2. The extracted features were analyzed using statistical and machine learning techniques to detect and assess drought stress dynamics in plants under variable nitrogen levels. **Chapter 4** presents a published paper that focuses on developing a framework for phenotyping the nutritional status of selected African crops using hyperspectral imaging combined with deep learning techniques. A pre-processing

framework involving radiometric calibration, down-sampling, spectral smoothing and denoising was developed. A hybrid deep-learning model was developed to identify the nutritional status of cowpea and quinoa at different growth stages. **Chapter 5** is a published paper comprising of the integration of full spectra data with machine learning algorithms for early drought stress detection in wheat species. Various waveband selection algorithms were applied to select the most sensitive wavelengths that respond early to wheat drought stress. New crop vegetation indices (VIs) were proposed using the selected wavelengths to assess drought stress in wheat under different nitrogen levels. **Chapter 6** discusses the methods used and further interprets the findings. It also compares the findings from experimental chapters with previous research studies to examine the strengths and weaknesses of the different techniques employed in this research. A link between the proposed methods and African agricultural practices were expatiated in this chapter. Finally, **Chapter 7** concludes the thesis by summarizing the methods and findings of this research. Overall, the methods and tools developed in this PhD thesis demonstrate that image processing combined with appropriate machine learning tools can automatically assess the nutritional and drought status of plants with high accuracy and precision in a high throughput manner. Readers may notice some repetition in some areas of the chapters due to the inclusion of the published articles.



# **1 Introduction**

## **1.1 Background**

According to the Food and Agriculture Organization of the United Nations, a 50% increase in the yield of worldwide consumed crops such as rice, barley and wheat is required to meet the 2050 global food demand agenda (State & Insecurity, 2014). With the population of Africa expected to almost double by 2050, going from 1.3 billion in 2020 to 2.5 billion by 2050, this part of the world is at a high food security risk (Van Ittersum et al., 2016). Hence, the potential to increase its food production and yield must be achieved by accelerating sustainable and multidisciplinary approaches in crop breeding programs (Altieri & Nicholls, 2017). Genotyping techniques such as DNA sequencing have been used in crop breeding programs to develop high-yield crop varieties (Naqvi et al., 2022). Although this provides in-depth crops information, a comprehensive understanding of the interactions between plant genetic traits and their physical characteristics is required (Bian et al., 2022).

In modern agricultural practices, agronomists and plant breeders characterize plant physical traits and their environmental interactions. This process, known as plant phenotyping, combines multiple techniques and protocols to measure complex physical plant traits such as shape, colour and texture at different structural levels, from leaves and shoots to whole plants (Das Choudhury et al., 2020). It provides essential information on how genetics, environment and crop physical characteristics can guide crop selection for yield, nutrient stress and disease resistance (Fiorani & Schurr, 2013). For plant breeders who manage hundreds or thousands of breeding lines, obtaining infield measurements using portable instrumentation or laboratory-based procedures seems infeasible. These methods are time-consuming, labour-intensive and prone to human errors, raising challenges for high throughput, sensitivity and accuracy (Pieruschka & Schurr, 2019). Hence, significant improvements in phenotyping methods are required to harness the wealth of plant phenotypic information.

Recently, novel sensing and image-processing tools have been developed to mitigate these challenges (Fiorani & Schurr, 2013). This development is pivotal to non-invasive phenotyping and has provided new opportunities to phenotype plants at varying growth scales in a high-throughput manner. This phenomenon, known as high throughput plant

phenotyping (HTPP), uses state-of-the-art imaging sensors to acquire still images to study trends and patterns in plants in real-time. This was first conceived in the early 1990s with the use of digital cameras (made of charge-coupled devices (CCD) and complementary metal-oxide-semiconductor (CMOS) sensors) for plant organ phenotyping, e.g., measuring leaves or root growth (Schmundt et al., 1998). Improvement in image-based phenotyping has led to the development of advanced imaging technologies such as 3D scanners, thermal and hyperspectral cameras, among others. Imaging sensors provide fast, highly repeatable and non-invasive data, enabling researchers to observe continuous plant growth in response to changing environmental conditions (Gracia-Romero et al., 2019). They are integrated into aerial and ground-based facilities as high-throughput installations to cost-effectively provide large amounts of unbiased data. Aerial systems, such as satellites, aircraft and unmanned aerial systems (UAS) monitor vegetation (Berni et al., 2009), providing fast coverage over a wide field. However, they are expensive to operate and lack high spatial resolution for specific plant traits. Furthermore, some have limited operating time and are susceptible to harsh weather conditions during operation. Ground-based systems, including handheld devices (e.g., spectrometers or imagers) and fixed or movable platforms (with installed imaging sensors), enable repeated measurements of plant phenotypes at frequent intervals and can also acquire high-resolution data at different view angles (Enciso et al., 2017). Fixed platforms may have a high payload to accommodate multiple sensors, allowing for high-throughput data acquisition, reduced labour and measurements of several phenotypes simultaneously. Such platforms are employed in glasshouse and field experimental stations around the world, including the field scanalyzer (Rothamsted Research, Harpenden, Hertfordshire, UK), the European Plant Phenotyping Network, and the Australian plant phenomics facility.

## **1.2 Image processing for plant phenotyping**

### **1.2.1 Imaging sensors**

A typical HTPP platform houses an array of sensors, such as digital RGB sensors (Ballesteros et al., 2020), hyperspectral cameras (Onoyama et al., 2015), fluorescence (Humplík et al., 2015), thermal (Möller et al., 2007) and 3D imaging laser scanner (Forero et al., 2022) for image data acquisition. Visible sensors output RGB images adopted for colour, texture, morphological and geometric-based operations. They can be

used in a controlled environment or on the field for phenotyping shoot biomass, yield, leaf morphology, canopy-related traits, fruit number and distribution. This imaging approach is commonly used due to the relatively inexpensive sensor cost and proliferated open-source analytical tools.

In hyperspectral imaging, spectroscopy sensors acquire data at narrow wavebands, enabling the assessment of vegetation indices (VIs) and spectral curves for water stress, nutrient stress and disease detection. Fluorescence imaging describes the visualization of the spatial distribution of substances that emit fluorescence; for example, chlorophyll fluorescence (Buschmann et al., 2001). It measures the fluorescence plants emit from chlorophyll, providing a rapid screening technique to identify plant stress response and photosynthesis efficiency. Thermal imaging measures the intensity of infrared radiation (between 300-1400 nm wavelength) emitted from the surface of plants and converted to temperature values (Fiorani & Schurr, 2013). It monitors canopy temperature and plant responses to environmental conditions and stress. 3D imaging sensors capture images in three-dimensional format, providing data on plant canopy architecture, volume and branching patterns. RGB, spectroscopy and fluorescence imaging work on the principle of the dynamics in light absorption and plant interactions. HTPP systems that integrate many sensors allow researchers to gather statistics on different environmental conditions and plant features.

Imaging pipelines are developed from the acquired images to extract relevant phenotypic information from plants. The results from the pipeline are stored in a database, which other researchers can use later. Figure 1.1 is an overview of the data management and processing pipeline for high-throughput phenotyping. As illustrated, the general sequence includes automatic image and environmental data acquisition from the field, glasshouse, or controlled environment. These data are stored in a remote server or cloud, which can be retrieved and used by interested parties. In some cases, the server has algorithms to preprocess and extract important phenotypic data, which are stored back in the server. The extracted data can then be analyzed using statistical and machine-learning tools based on the interest of the researcher.

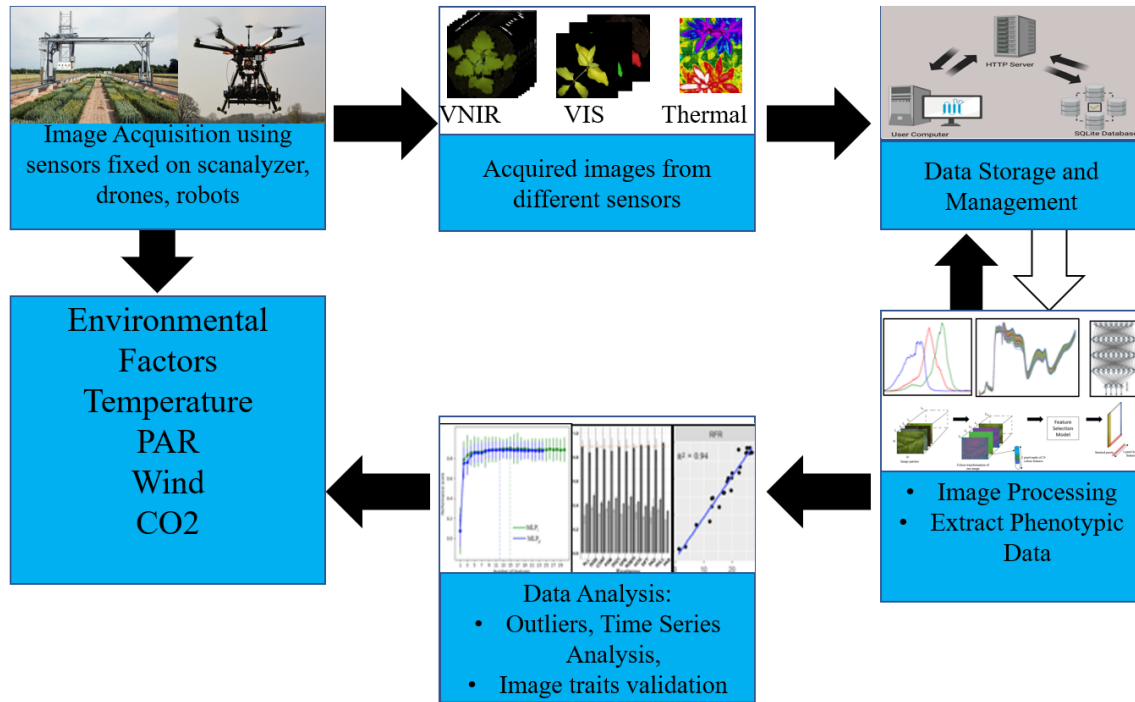


**Figure 1-1** Phenotyping pipeline from image acquisition to data analysis (Virlet et al., 2017). This involves data acquisition using an aerial (eg. UAV) or ground based systems (e.g. field scanalyzers) to acquire images. The sensors may include visible RGB camera, hyperspectral, thermal, 3D laser sensors etc. The data are then stored in a data management or cloud server. The stored data can be pre-processed, post processed and analyzed using statistical or machine learning methods.

### 1.2.2 RGB visible image processing

Developing workflows to process plant images enables the automatic extraction of features from images. Feature analysis is preceded by a sequence of steps to enhance, analyze or manipulate images. The sequence typically involves image acquisition, pre-processing, segmentation and feature extraction as shown in Figure 1.2. Image acquisition is the initial step, where images are obtained using digital sensors that can detect and visualize objects in the visible wavebands of the electromagnetic spectrum. These sensors capture images as unprocessed raw files (with extension raw, .cr2 or .nef file extensions) or processed files (with .jpg and .png file extensions).

After image acquisition, pre-processing follows to improve the quality of images for analysis. Pre-processing steps typically involve image resizing, cropping, contrast enhancement, denoising and colour correction. Transformation techniques such as grayscale conversion, normalization and standardization are also performed as pre-

processing steps to enhance images (Mochida et al., 2018). Data augmentation is often adopted as a preprocessing tool to increase dataset variations for generalized and robust image analysis. This includes noise addition, rotation, flipping, and scaling. Image segmentation is the next step after the basic preprocessing steps and is fundamental but critical to achieving high-throughput phenotyping. It separates image pixels into classes, typically foreground (plant vegetation) and background (non-plant objects).

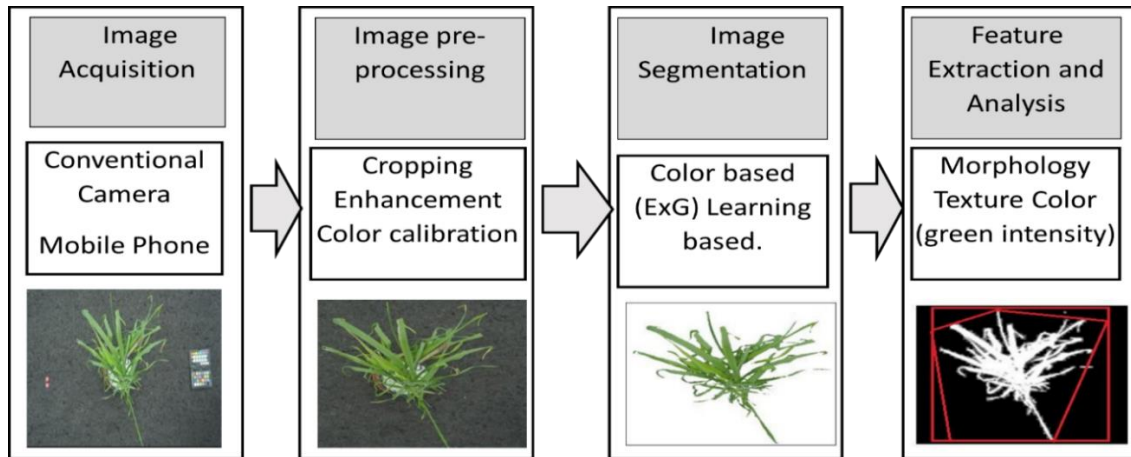


**Figure 1-2** A typical image processing pipeline for plant phenotyping. This mainly involves four steps; (a) image acquisition using visible RGB cameras, (b) image pre-processing which involves cropping, image enhancement and colour calibration. (c) The images are then segmented to remove background from foreground or isolate the region of interest. (d) Feature extraction follows this step where important traits of plants such as colour and texture are extracted for further analysis.

Image segmentation is one of the critical functionalities of an image processing pipeline that influences various stages of analysis and facilitates the extraction of meaningful information from images. Segmentation involves partitioning images into multiple parts, such as leaves, fruits and ears, decreasing the complexities of models to understand trends in images. Several methods have been developed, which can be categorized as threshold-based, colour-index-based (Meyer & Neto, 2008), and learning-based methods (Adams et al., 2020). Segmentation plays a significant role in detecting and quantifying plant nutrients in real-world scenarios under non-limiting imaging conditions. Due to its low computational cost, robustness to differences in resolution and particle occlusion, colour is the most popular feature for plant segmentation

(Sadeghi-Tehran et al., 2017). Transforming colour from the original RGB channel to different spaces, such as hue saturation value (HSV) and luminance chrominance (YCbCr) or using combinations of different colour bands, increases the contrast between plant vegetation and background (soil). They also reduce the impact of illumination caused by differences in colour constancy between images. Several studies have utilized colour information to segment plants using information from various colour spaces. Prasetyo et al. (2017) combined different colour spaces to segment and detect mango trees. They transformed the RGB colour space to HSV and YcbCr, combining individual colour channels with Otsu thresholding for segmentation. Their results showed that Cr (red chrominance) from YcbCr effectively separated mango trees from the background. Similarly, Sadeghi-Tehran et al. (2017) combined multiple colour features from different colour spaces to train a supervised machine-learning classifier for wheat ear segmentation in field images. They found that using multiple colour spaces improved segmentation accuracy and was more robust to background noise and illumination changes than using a single-colour space.

Feature extraction involves generating a significant set of traits that can sufficiently represent images. Features are usually extracted based on patterns and trends that distinguish images and are usually selected based on object characteristics in images. In plant phenotyping, features such as colour (green intensity, HSV), texture (Harr wavelet) and morphology (leaf area, biomass) are extracted for the prediction, classification and identification of plant phenotypes. For example, Zhuang et al. (2017) extracted greenness dominance colour traits and energy features based on wavelet transform (texture features) to identify water stress in maize. They trained a two-stage machine learning model using seven colour and texture feature subsets to evaluate plant water levels, which achieved superior performance. Azimi et al. (2021) also developed an automatic image-based phenotyping system for nitrogen stress detection using a histogram of gradient and SIFT image descriptors.

Although plant scientists are provided with diverse ways to extract and analyze features and functionality of plants, the downstream processing of the images faces challenges hindering its widespread adoption and effectiveness. Several tools have been developed in the past years to facilitate the processing and quantification of plant phenotypic traits. These include PlantCV (Gehan et al., 2017), ImageJ (Schindelin et al., 2015), and

PhenoPhyte (Lobet et al., 2013). PlantCV is an open-source image analysis kit that provides a package of tools designed specifically for plant biologists to statistically analyze and visualise plant features from digital images. It supports a variety of tasks, including segmentation, morphometrics, and colour analysis. It is also designed to handle large datasets, combining them with metadata for comprehensive phenotypic analysis. Like PlantCV, ImageJ is known for its flexibility, extensive plugin ecosystem and user-friendly interface. It has features for filtering, segmentation, trait measurements and 3D analysis. PhenoPhyte is a specialized tool useful for studying plant growth, health and stress responses. It has an embedded automated tool for segmentation, morphological and colour trait analysis that can be combined with experimental metadata for phenotypic and genomic trait analysis

Although these tools are efficient, most are plant-specific, designed for specific platforms (e.g., LemnaTec Scanalyzer, Phenospex PlantEye). The LemnaTec Scanalyzer is a cutting-edge phenotyping platform particularly used for plant research to examine the physical and biochemical properties of plants. The system automates the assessment of multiple plant phenotypes with high-throughput image sensing technology. It has a collection of sensors including visible RGB, hyperspectral, 3D laser scanner and thermal sensors that monitor plant growth, development and plant interactions with their environment. The Phenospex PlantEye is a high-precision scanning and imaging device that tracks and analyses plant traits in real-time. This tool was developed exclusively for plant phenotyping, measuring a variety of plant traits using a plethora of sensors including multispectral photography and 3D laser scanner. Most of these advanced tools require some programming skills. In addition, the different functionalities in the pipelines are limited in terms of automated batch processing, end-to-end data processing and robustness to environmental and plant variabilities. For example, the watershed segmentation approach implemented in PlantCV v2 software may fail in segmenting certain field crops. LeafGP processing tool lacks functionality for colour calibration which may hamper the accurate extraction and processing of colour traits for plant phenotyping. Moreover, non-open-source tools such as ENVI (Environment for Visualizing Images) are costly and time-consuming when processing large-scale data.

### 1.2.3 Hyperspectral image processing

Hyperspectral imaging (HSI) is a technique that analyses the reflection of light on objects over a broad spectrum. It integrates imaging and high-resolution spectroscopy over several spectral bands, providing detailed information about objects in images. Spectroscopy studies the emission or reflectance of light from objects and the variation in energy and wavelength. This outputs a high-dimensional cube image with each pixel having two spatial (x and y dimensions) and one spectral dimension (z-dimension with a continuous reflectance captured over an electromagnetic spectrum) (Figure 1.3). An HSI system records reflectance in a wide range of spectra covering ultraviolet (UV), 315-400 nm, visible (VIS) 400–700 nm, near-infrared (NIR) 700–1100 nm, to the short-wave infrared (1100–2,500 nm) regions (Tardieu et al., 2017).

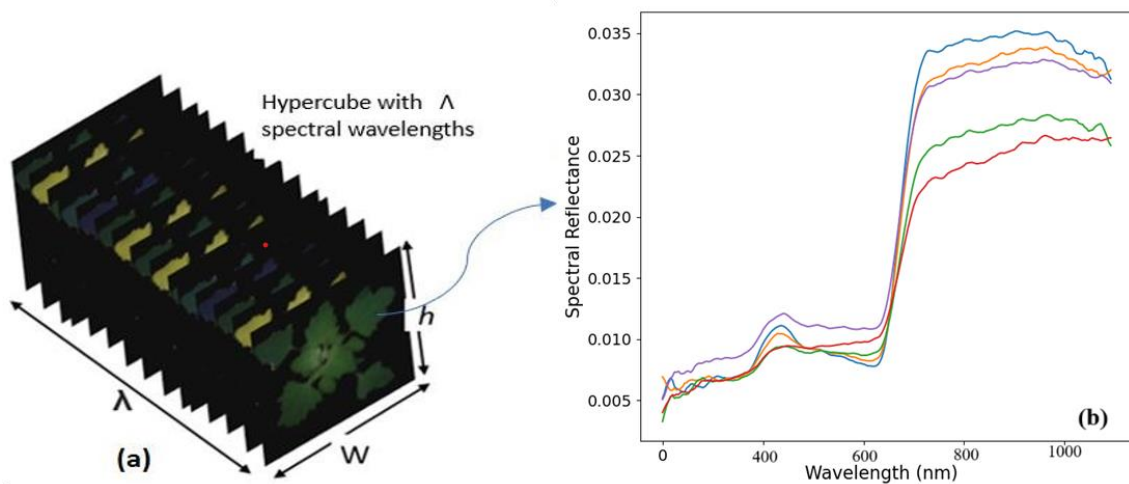


**Figure 1-3** Example of a hyperspectral image of plant leaf canopy. (a) A stack of narrow band sub-images from 394 nm to 2500 nm covering the visible and near infrared regions as well as the short wave infrared regions. This forms a 3-D hypercube made of the two spatial dimensions (height ( $h$ ) x width( $w$ ) and one spectral dimension ( $\lambda$ ); (b) a spectral signature of a plant pixel spanning from 390 nm to 2500 nm .

Until recent years, HSI was mainly limited to remote sensing applications. Lately, it has emerged as a potential tool for non-invasive proximity sensing of plant phenotypes (A. Lowe et al., 2017). Utilizing HSI for plant phenotyping enables the simultaneous extraction of structural and physiological information from plants (Mishra et al., 2020). That is, while plant morphology information (leaves, fruit, height and area) could be



obtained from the spatial imaging modality of HSI, detailed physiology information can be obtained from the temporal information of plants (Mohd Asaari et al., 2018). The 400-700 nm spectral range is associated with plant pigment composition, including chlorophyll, carotenoid and anthocyanin (Figure 1.4) (Siedliska et al., 2021a), while the NIR and SWIR regions contain information for cellulose absorption and compounds containing C-H, N-H and O-H bonds in plants (Mishra et al., 2017). When combined, the structural and physiological data offer a deep understanding of the dynamics involved in plant function. HSI sensing is applied to plant phenotyping in four main ways: for monitoring biotic and abiotic stress, prediction of biochemical parameters, identification and detection of plant diseases and assessment of crop protection compounds potency (Mishra et al., 2020). Although HSI provides rich spatial and spectral information necessary for the detailed characterization of plant phenotypes, it also presents challenges, such as high dimensionality, illumination effect, computational complexity, and data fusion. Subsequent chapters of this thesis have highlighted and tackled some of these challenges.

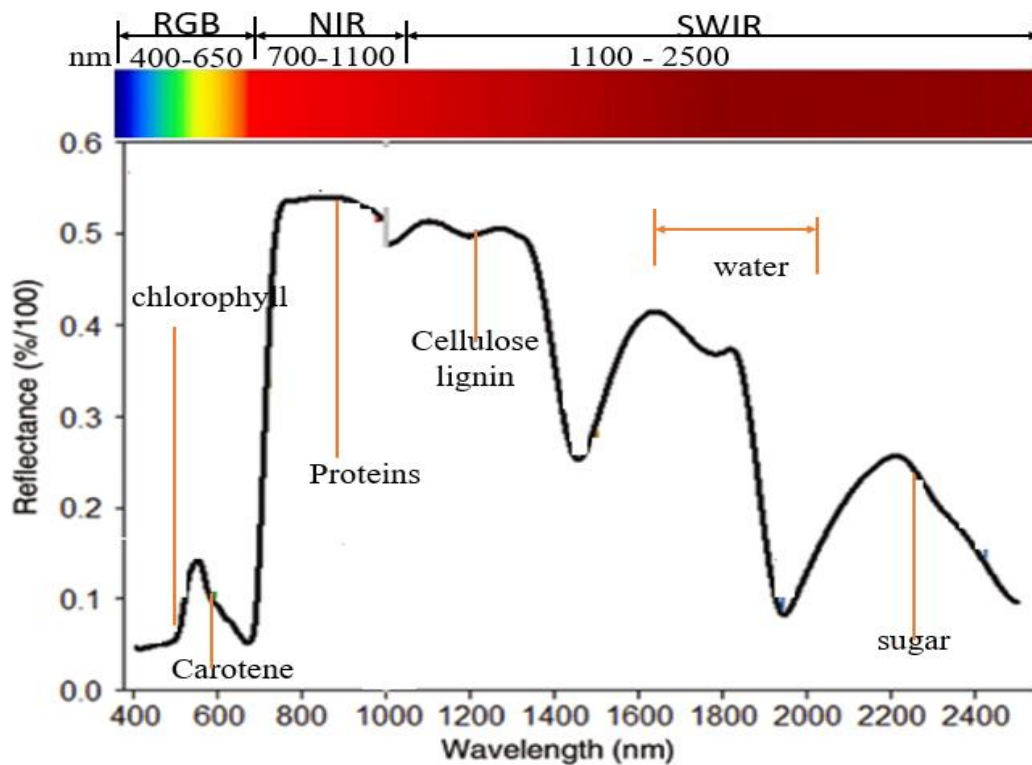


**Figure 1-4** A spectral reflectance curve showing the reflectance properties of a typical green vegetation across different wavelengths of light (Kuska et al., 2018). Machine learning for image-based plant phenotyping. The region around 400-500 nm corresponds

to chlorophyll, indicating strong absorption by chlorophyll pigments while the regions around 700 nm -1000 nm show a high reflectance due to proteins which might be indicative of structural compounds in the plants. The SWIR around 1200 nm -1500 nm is marked for cellulose and lignin while a dip in the region between 1500 nm-1800 nm shows the responsiveness to changes in water components in the plant canopy. A feature near 2300 nm is attributed to sugar, likely a reference to carbohydrate content. Advancements in technology for HTPP have dramatically increased the scale at which plant phenotypes are measured. Extracting data patterns and trends, assimilation and feature identification from large datasets necessitated data mining and machine learning (Singh et al., 2016). Machine learning (ML) is a branch of artificial intelligence and computer science encompassing algorithms that can learn from data without relying on explicit programming. It is often used to analyze results from complex data, exploring the possibility of identifying patterns and trends to interpret measurements. ML can broadly be classified as supervised, unsupervised and reinforcement learning. Supervised learning relies on labelled input and output training data to make future predictions or classifications (Rahaman et al., 2019). Examples include support vector machines (SVM), random forests, k-nearest neighbours (KNN) and linear discriminant analysis (LDA). Unsupervised learning involves training a model with unlabelled data without the help of a known outcome variable. Examples include k-means clustering, hierarchical clustering, and principal components analysis. In plant phenotyping, ML has several applications that allow researchers to speed up breeding efforts and extract valuable insights from big datasets.

Through regression and classification modelling, ML is used for trait extraction and quantification, stress detection and diagnosis, and the development of predictive modelling and decision support systems. Regression models are applied to predict the magnitude, quantity or intensity of plant phenotypic traits. Some specific regression models applied to image-based plant phenotyping include linear, polynomial, ridge, lasso partial least square and support vector regression models. Shi et al. (2021) employed a predictive regression model to phenotype rice nitrogen concentration in the field. Three regression methods, simple nonlinear regression, backpropagation neural network and random forest, were evaluated for dry matter, nitrogen accumulation and leaf index area estimation. The random forest ensemble algorithm was the most effective model, with

strong generalization performance. Hassanijalilian et al. (2020) used a support vector regression model for infield estimation of soybean chlorophyll content. Several colour-related features were extracted. However, the dark green colour index (DGCI), which had the best correlation with SPAD (soil plant analysis development) values ( $r = 90$ ), was used as input to train the SVM model. This model has the potential to predict chlorophyll directly with raw RGB images ( $R^2 = 0.86$  and  $RMSE = 3.20$  SPAD units) without the need to use a standard calibration board and outperformed other models such as linear, polynomial, and multiple linear regression. In plant phenotyping, classification models are frequently employed to categorize, detect or quantify plants into distinct classes or groups according to observed traits. Some classification models commonly used include SVM, RF, LDA and deep neural networks. By leveraging classification techniques, researchers use these approaches to automate plant classification activities, identify traits of interest and assist breeding efforts to produce improved crop varieties with desired attributes. As a representative study, Guo et al., (2022) investigated the potential of using spectral and texture indices to predict SPAD values at different growth stages of maize using two machine learning models (SVM and RF). Their findings show that SVM outperformed the RF model in estimating SPAD values with  $R^2$  (0.81) and RMSE (0.14), respectively. Culman et al. (2017) employed a single multilayer perceptron to classify oil palm into one of four groups- healthy , nitrogen, magnesium or potassium deficient achieving a 0.67 average classification accuracy.

Although ML models achieve satisfactory results in image-based plant phenotyping, they necessitate manual feature engineering, where domain specialists extract and engineer relevant plant traits that are important to the model. This hinders the automation process in HTPP, hence the need for deep learning to automatically learn complex patterns from raw data without the need for feature extraction and engineering. Deep learning (DL) has emerged as a state-of-the-art advanced ML technique capable of detecting, classifying and predicting plant phenotypes. It extends classical ML by adding more complexities to a model to enable higher-level data abstraction using multiple nonlinear transformations (Schmidhuber, 2015). DL models consist of different components (convolutions, pooling layers, gates, memory cells, activation functions, encoding and decoding layers, etc.) and their use depends on the interested network architecture. A convolutional neural network is commonly used in image processing to classify, identify and predict plant

phenotypes. DL models can be trained using either a supervised or unsupervised technique. In unsupervised DL, the goal is to find trends and patterns in the data for various purposes, like hashing and clustering. However, in supervised DL, labelled input data (e.g., image of plant canopy) is mapped to a known output (e.g., nutrient deficiency symptoms on plants) through a weight vector and the errors are backpropagated from the output to the input layer (Singh et al., 2018). DL models, particularly convolutional neural networks, have been used in diverse ways to phenotype plants for disease detection and quantification, segmentation, leaf counting and growth monitoring. Jiajing Zhang et al. (2022) developed a deep-learning disease detection model to automatically assess wheat resistance to *Fusarium* head blight under field conditions. Mask-RCNN model was developed to segment wheat spikes and FHB diseased regions, which resulted in 77.76% and 98.81% detection rates, respectively. One significant difference between traditional ML and DL models, especially using supervised methods, is that while ML has a two-step process of feature extraction and decision-making, DL combines this two-step approach into one model, thereby avoiding manual handcrafting feature extraction.

#### **1.2.4 Phenotyping nutritional status of plants**

Reliably diagnosing plant nutrient stress is integral to farm management since excess or lack of nutrients can be detrimental to crops and the environment. Specifically, excess fertilizer use causes soil and water contamination and a surge in greenhouse gases in the atmosphere (Thorburn et al., 2019). In contrast, the undersupply of plant nutrients causes low yield (Chlingaryan et al., 2018). Since excess or inadequate plant nutritional levels are not always quantifiable by visible symptoms, developing tools to track and estimate plant nutritional status is essential. This will reduce yield losses and be the basis for the rational use of fertilizers as conceptualized by precision agricultural practices. Visual examination of plants using a colour guide is one of the most common methods to track plant nutritional status. However, this technique is error-prone as different abiotic stresses could show similar colour differentials. Additionally, it does not allow the quantification of plant nutrient stress (Graeff et al., 2008). Several works have proven the accuracy of laboratory-based methods to estimate plant nutrient status (Jones et al., 1994). For example, the Kjeldahl digestion method (Kjeldahl, 1883) is widely used as a reference for plant nitrogen estimation. Some indirect alternatives, such as nitrate

concentration in sap, upper layer leaf N concentration and chlorophyll meter (SPAD), have also been developed for plant nitrogen estimation (Balasubramaniam & Ananthi, 2016). However, they are time-consuming, require specific procedures to be followed to interpret data and are not always accurate (Nauš et al., 2010). Hence, considerable efforts have been made to develop non-invasive high-throughput methods to detect and track plant nutritional status (Ali et al., 2015). The review paper by Li et al. (2020) demonstrates how advancements in imaging technology have accelerated non-invasive phenotyping, enabling time-series and high-throughput monitoring of plant nutrients.

Most studies on image-based nutrition analysis available in the literature use images taken by unmanned aerial vehicles (UAVs), aeroplanes or satellites (Barbedo, 2019). While these technologies offer several advantages, such as large-scale coverage and remote sensing capabilities, there are challenges and limitations associated with their application. For example, satellite images are too coarse for individual plant analysis, impeding the detection of the early onset of nutrient deficiency. Moreover, operational skills, cost, and poor weather conditions can degrade image quality or even prevent flights from UAVs. Hence, proximal images are usually good options playing essential roles in plant stress analysis. The detection and prediction of nutrient deficiency depend on visual signs that characterize this disorder. From this perspective, three main problems are tackled: classification (to identify the origin of nutrient deficiency signs), identification (indicates the presence of deficiency disorder by detecting its signs or symptoms), or quantification (aims to estimate the severity of the deficiency) (Barbedo, 2019). All the three problems have been tackled using different sensing systems: digital imaging and hyperspectral or multispectral cameras. From the extensive works done in plant phenotyping, most of the works have explored nitrogen identification and quantification, probably because nitrogen is the most essential nutrient that is economically and socially relevant (Barbedo, 2019). However, only a few studies are dedicated to phosphorus nutritional status analysis.

#### **1.2.4.1 Digital imaging for plant nutrient stress analysis**

For digital imaging, segmentation plays a vital role in plant nutrient analysis by enabling the detection and quantification of nutrient-related traits from still images. After effective segmentation, images undergo further processing to extract colour, texture,

morphology and a host of off-the-shelf features to train classifiers or regression models to identify plant nutrient status accurately. The analysis of nitrogen content is mainly based on the dynamic changes of plant colour features. Studies show that the raw RGB channels (used individually or combined) and colour vegetation indices (CVIs) derived from these channels are associated with nitrogen concentrations. The green channel especially has an indirect relationship with nitrogen. Hence, CVIs such as the green leaf index  $(R-G/R+G)$ , normalized green index  $(G/R+G+B)$ , and modified green-red vegetation index  $(G - R) / (G + R)$  have been widely used to assess nitrogen content. Qui et al. (2021) applied eleven CVIs to determine the nitrogen nutrition index (NNI) of rice at a given growth stage. While the results showed a strong correlation between the NNI and the CVIs, this relationship varied with varying growth stages. In a similar study, Hassanijalilian et al. (2020b) presented an image processing tool to estimate soybean leaves infield chlorophyll content based on SPAD meter measurements. Among numerous CVIs that indirectly evaluate plant chlorophyll concentration, the dark green colour index had the highest association with readings from SPAD metres. Information from different colour spaces such as HSV, Lab, and YCrbr have also been explored. Wang et al.(2014) applied the  $b^*$  component from the Lab\* colour space to identify the nitrogen status of four different cultivars of rice species under six different treatments. Their results show a strong correlation between this feature and nitrogen content ( $R^2 \geq 0.12 \leq 0.8$ ). However, this is dependent on the specific growth stage.

Unlike colour features, which have been widely used for nutrient analysis, a few studies focused on using information from other phenotypic features, such as morphology and texture, to understand plant nutrient stress dynamics. Romualdo et al. (2014) employed Gabor wavelet texture and colour features to diagnose induced nitrogen deficiency in maize leaves. They developed an artificial vision system to detect nitrogen stress in plants at various growth phases, particularly in the early stages, which could help with early diagnosis and growth cycle correction. Their pipeline identified levels of nitrogen deficiency in corn with accuracy rates of 82.5% and 87.5% at the second and third growth stages, respectively.

plants with phosphorus (P) deficiency may exhibit subtle changes that are challenging to distinguish, unlike nitrogen deficiencies, posing difficulties for image processing. Image-based phosphorus deficiency detection is a topic of active research, with several studies

focused on combining different traits of colour, texture and morphology to detect this stress. Ali et al. (2015) applied leaf area, perimeter and RGB colour information to a linear discriminant analysis regression model to estimate leaf phosphorus levels. Three plant species (cotton, tomato and lettuce) were grown in a medium with three different phosphorus levels ( $P_0 = 0 \text{ P}$ ,  $0 \text{ L}$ ;  $P_1 = 2.5 \text{ mL } 10 \text{ L}^{-1}$  and  $P_2 = 5 \text{ mL } 10 \text{ L}^{-1}$  of  $\text{P}$ ) in the glasshouse. The results show a significant correlation between the predicted and actual  $\text{P}$  values for lettuce ( $R^2 = 0.629$ ), tomato ( $R^2 = 0.778$ ), and cotton ( $R^2 = 0.807$ ), indicating the potential use of plant geometric and colour features to identify phosphorus stress in plants. Recent studies are focused on adopting DL architectures like CNN and recurrent neural networks (RNN) to develop comprehensive nutrient stress recognition pipelines, which could be used as practical tools in high-precision management strategies. Azimi et al. (2021b) developed a 23-layer DL model to automatically identify nutrient stress in Sorghum. They compared its performance with traditional ML techniques and other deep learning architectures. From their results, the proposed deep learning model outperformed traditional machine learning techniques, achieving an 8.25% better accuracy. Their findings show that a simple CNN model has equally good performance as state-of-the-art DL models such as ResNet18 and NasNet, which have millions of trainable parameters. As indicated earlier, while nitrogen deficiency assessment has received extensive study, little has been done on phosphorus. Also, few studies have been conducted to determine the sensitivities of extracted features to the stress under study. Moreover, there is little understanding of the effect of different trait combinations on detecting nitrogen or phosphorus deficiencies.

#### **1.2.4.2 Hyperspectral imaging for plant nutrient stress analysis**

Although features from RGB images provide reliable spatial information, they do not provide detailed spectral descriptions of plants. Hyperspectral images convey spatial, spectral, and multi-constituent information, allowing potential detailed, non-invasive plant physicochemical analysis (Zhang et al., 2013). Plants experience a range of physiological and biochemical responses to nutrient stress, which can lead to subtle alterations in their spectral signature.

From previous research, nitrogen stress spectral features are primarily observed in the visible (400nm – 650 nm), and NIR (700 nm -1000 nm) of the electromagnetic spectrum.

For instance, Eshkabilov et al. (2021) used hyperspectral imaging to detect different levels of nutrients in lettuce using a hyperspectral camera in the 400-1000 nm wavelength range. They employed a simple linear regression model to detect the optimal wavebands, 506-601nm and 634-701 nm, which are sensitive to the nutrient dynamics in the plant. Their results showed an  $R^2$  range of 0.784 - 0.987, predicting soluble solid content, pH, nitrate, calcium, potassium and chlorophyll in lettuce plants. Ye et al., (2020) also presented a rapid and non-destructive approach for estimating and mapping nitrogen content in apple trees at the leaf and canopy levels. An ImSpeof plant sector V10 system was employed to acquire images of plants within 400 – 1000 nm. The raw and first derivative spectra were combined with conventional regression models (MLP and PLSR) to predict the nitrogen content of the plants. From the results, the MLR model trained with the raw spectra reflectance achieved the highest score ( $R^2 = 0.7843$ ) in predicting nitrogen content in apples.

Although the SWIR is used less frequently than the visible and NIR regions, it still has some potential for nitrogen stress detection. While the SWIR is sensitive to plant water dynamics, nitrogen stress can impact plant water status and plant biochemical properties, which can cause changes in the spectral signature in this region related to water absorption features and biochemical constituents (Gitelson et al., 2006) . Hence, researchers are also using the SWIR as a tool to measure plant nitrogen deficiency. In most cases, they combine the three main spectral regions (visible, NIR, and SWIR), resulting in a full spectrum (400-2500 nm) for nutrient stress analysis. Yamashita et al., (2020) employed hyperspectral imaging within the 400 nm – 2500 nm spectra to estimate the chlorophyll and nitrogen content of tea leaves. The spectral wavelengths were processed using five pre-processing techniques to correct illumination and noise effects. A regression model was then developed using several machine-learning algorithms and preprocessing techniques. A data-sensitive analysis of the resulting wavelength identified wavelengths between 1325 -1575 nm as sensitive to leaf N content in tea plants, helping to track N status in crops non-destructively.

### **1.2.5 Image-based phenotyping of plant drought stress**

Plant drought stress is a water supply deficiency due to reduced soil water level or changes in plant physiological architecture (Ihuoma & Madramootoo, 2017). It is one of



the leading environmental constraints that compromise and affect sustainable crop production (Zhuang et al., 2017b). It impedes plant growth, causing a decrease in leaf extension rate, leaf area and plant height, subsequently affecting crop yield (Dodig et al., 2021). Tools and strategies to monitor and control plant water loss are essential to increase water savings and agricultural sustainability.

Conventional methods for monitoring plant drought stress are either soil- or plant-based. The soil-based approaches rely on soil moisture measurements, such as soil moisture sensor analysis of water depletion. Although soil moisture sensors are standard tools for monitoring plant water status, these are point-based measuring tools that cannot account for soil spatial variability. Also, to measure the rate of water depletion, the soil is regularly sampled at very few points based on the assumption of uniformity of water holding capacity, plant density and transpiration rate (Clarke, 1997). However, soil heterogeneity makes it difficult to implement this technique. Plant-based methods, including stomatal conductance measurement, leaf water potential, relative water content, fruit and stem diameter and sap-flow measurement, indicate directly or indirectly plant water stress status (Prasad et al., 2006). These methods are, however, invasive, labour-intensive and time-consuming (Gutiérrez et al., 2018).

For this reason, image-based, non-invasive methods have been used in recent times to detect and monitor water stress in plants. From a holistic view, traits such as colour, texture and morphology express enough drought information and can indicate plants response to water stress dynamics (Zhang et al., 2022). Infrared thermometry and thermal imaging are the most widely used non-invasive approaches for drought stress detection (Ihuoma & Madramootoo, 2017). These have been used to obtain average canopy temperature to calculate crop water stress index (CWSI) to assess water deficit for diverse crops such as cotton, olive, and wheat (Elsayed et al., 2017). Although the use of infrared thermometry for drought stress analysis is acknowledged in several studies, digital RGB and hyperspectral imaging are gaining momentum due to their capabilities in assessing complex plant physiological traits.

#### **1.2.5.1 Digital Imaging for drought stress analysis**

Drought stress usually leads to yellowing, browning or wilting of plant leaves, which can be captured by RGB images. This enables researchers to detect and evaluate drought

stress by monitoring changes in colour patterns, especially the green channel or colour vegetation involving the green channel. Also, drought stress can affect plant geometric properties, causing changes in canopy area, plant height, and leaf size. These changes can be monitored to detect and quantify drought stress in plants. Changes in leaf texture, such as curling and wrinkling, can result from prolonged drought stress. These changes are usually more pronounced in texture features such as the grey-level co-occurrence matrix, local binary patterns or the tamaru features. Hence, these features can be used to characterize drought-induced stress in plants. Due to the complexity of plant water stress, combined analysis of multiple phenotypic traits can comprehensively explain plant water physiological conditions (Chen et al., 2014). As a representative study, Duan et al.(2018) proposed a feasible, robust and non-destructive method to quantify the response of rice species to drought under both controlled and field conditions. They experimented with four features, including greenness plant area ratio (GPAR), ratio of total plant area and bounding rectangle area ratio (TBR), perimeter area ratio (PAR) and ratio of total plant area and convex hull area ratio (TCR). The results show that these four features could dynamically measure the drought response under controlled settings over time (on a daily or half-hourly basis) and could reliably distinguish between drought-sensitive and drought-resistant accessions. In addition, their findings revealed that PAR and GPAR features were more robust and highly sensitive to drought stress than the other features in the glasshouse and field environment. Zhuang et al. (2017) also proposed a model for the early detection of drought stress in maize. They generated fourteen-dimensional colour and texture features from plant images to train a two-stage ML model to evaluate water stress. The colour features were indices based on the green dominance channel while the texture features were energy features based on the wavelet transform. Their model had an 80.95% accuracy for recognition of three water treatments, while the accuracy of water stress recognition reached 90.3% accuracy. Their results showed that the proposed model effectively discriminates water suitability and stress conditions in the maize field.

Although digital imaging is a valuable tool for plant drought stress analysis by monitoring the changes in plant morphology, colour, and texture, there is less work on identifying drought stress of plants with multiple stress levels, such as nutrients or diseases. The visual symptoms of drought stress on plants are similar to other stresses like nitrogen deficiency and rust. This makes it difficult to distinguish between these kinds of stresses;

hence, an image-based analysis pipeline is required to mitigate this challenge in plant phenotyping.

#### **1.2.5.2 Hyperspectral imaging for plant drought stress analysis**

Hyperspectral imaging is sensitive to subtle changes in the leaf biochemical and physiological composition and can potentially study drought responses at higher spatial resolution (Proctor et al., 2021). A common approach for HSI-based drought stress detection is using VIs derived from ratios or linear combinations of spectral reflectance at a few single wavelengths. VIs reduce the possible influence of variations in illumination during data processing (Jay et al., 2017). They offer valuable insights on plant drought stress detection by capturing subtle changes in the spectral reflectance associated with water deficiency. VIs can be used for the early detection of drought stress in crops due to their sensitivity to changes in plant physiology and water content. Several VIs have been developed and are used to monitor drought stress. The common ones include water index (WI)  $[(NIR - SWIR) / (NIR + SWIR)]$ , normalized difference water index (NDWI)  $[(NIR - Green) / (NIR + Green)]$ , normalized difference drought index (NDII)  $[(SWIR - MIR) / (SWIR + MIR)]$  and leaf water content index (LWCI)  $[NIR / SWIR]$ . These known indices have been widely used to monitor plant response to water dynamics. While the indices above are particularly for water stress analysis, the normalized difference vegetation index (NDVI) is the best-known VI, which has been extensively used to monitor different phenotypic traits. Although it is a numerical indicator of vegetation greenness, indicating the amount of chlorophyll and fraction of green cover, it is also used in irrigation management and scheduling (Kim & Glenn, 2017). The work of Almouctar et al. (2024) assessed the drought effect on vegetation in the Niamey region caused by low rainfall and high-temperature rise. They monitored two biophysical parameters- NDVI and land surface temperature (LST) derived from the multi-temporal data of Landsat 8 between 2013 and 2019. Their findings revealed a linear relationship between LST and NDVI in 2013 ( $R^2 = 0.34$ ,  $P = 0.58$ ) and an inverse relationship in 2019 ( $R^2 = 0.06$ ,  $P = 0.25$ ) between the two parameters.

Although hyperspectral VIs are widely used for drought stress monitoring, they do not exploit information from the regions highly affected by drought since they only use a few bands. Also, due to species-specific variations in drought response, appropriate VIs

are not always apparent for the early detection of drought stress. Data-driven machine learning-based algorithms that retrieve plant spectral and spatial characteristics hold great potential in offering new insights into drought stress analysis. Asaari et al. (2019) developed a data-driven method that combines clustering, band selection and spectral similarity measures to analyze hyperspectral data for detecting and recovering drought impact maize plants.

### **1.3 Significance of the Research Problem**

Identifying essential plant features sensitive to nutrient and water stress is critical in crop management. Using conventional phenotyping methods for plant stress feature identification is destructive, expensive and laborious when done on a large scale. In contrast, digital RGB and hyperspectral cameras can acquire large phenotypic information to monitor nutrient and water stress in plants over their growth period. However, studies in this area are scarce and have been limited to commonly consumed crops. Few studies have tried establishing a relationship between various plant phenotypic traits and the dynamics of different stresses on plants. Also, there is little knowledge on the response of spectral features to drought-stressed plants in variable nitrogen levels. Therefore, this study proposes methods to detect plant nutrient and water stress using digital RGB and hyperspectral data combined with machine learning approaches.

Identifying changes in plant colour, morphology and physiological traits as indicators of abiotic stress is the foundation for any impactful image-based phenotyping research to monitor plant nutrient and water stress. Drought and nutrient-sensitive in tandem with machine learning approaches to identify the early signs of nutrition and water stress in plants. This process could substitute traditional phenotyping methods where the entire growth cycle of the plant can be monitored.

Hyperspectral imaging shows promising results for image-based phenotyping, providing the capacity to measure plant traits over a wide range of wavebands simultaneously. VIs could be explored for their efficiency in estimating water or drought stress. However, these indices only capture specific aspects of plant vegetation, such as chlorophyll or photosynthetic activities, offering limited information for whole plant phenotyping.

Using machine learning approaches enables the modelling of the spatial and spectral properties to monitor plant nutrient and water stress.

### **1.5 Hypothesis Statement**

1. Compared to traditional methods, implementing an image processing pipeline will improve the automation, accuracy and accessibility of image processing tasks, such as colour calibration, segmentation and feature extraction for plant phenotyping.
2. Statistical methods and machine learning techniques combined with image processing can help identify plant traits that are responsive to water stress beyond the limitations of traditional image processing.
3. A hybrid convolutional neural network architecture can be applied to hyperspectral imaging to detect and monitor the onset of nutrient stress in plants at different growth stages.
4. Combining existing spectral indices with newly developed drought stress indices can be used to detect plant drought stress early and accurately assess the impact of high and low nitrogen levels on drought stress development.

### **1.6 Statement of Objectives**

This research aims to develop an image processing and machine learning framework for phenotyping nutrition and drought stresses in selected crops. The specific objectives are presented in individual chapters, while the main objectives are:

1. To develop an image processing pipeline to analyze plant phenotypes responsive to different nutrient levels.
2. To combine statistical and machine learning methods to identify plant physical traits sensitive to plant water stress under different nitrogen levels.
3. To develop a deep learning discrimination model to monitor the dynamics of plant nitrogen and phosphorus nutrients across selected plant growth stages using hyperspectral imaging data.
4. To evaluate the use of hyperspectral data integrated with machine learning algorithms for the early detection and classification of drought stress in wheat.

## 1.7 References

- Adams, J., Qiu, Y., Xu, Y., & Schnable, J. C. (2020). Plant segmentation by supervised machine learning methods. *Plant Phenome Journal*, 3(1), 1–11. <https://doi.org/10.1002/ppj2.20001>
- Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. (2015). Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 7(12), 16398–16421. <https://doi.org/10.3390/rs71215841>
- Almouctar, M. A. S., Wu, Y., Zhao, F., & Qin, C. (2024). Drought analysis using normalized difference vegetation index and land surface temperature over Niamey region, the southwestern of the Niger between 2013 and 2019. *Journal of Hydrology: Regional Studies*, 52(January), 101689. <https://doi.org/10.1016/j.ejrh.2024.101689>
- Altieri, M. A., & Nicholls, C. I. (2017). The adaptation and mitigation potential of traditional agriculture in a changing climate. *Climatic Change*, 140(1), 33–45. <https://doi.org/10.1007/s10584-013-0909-y>
- Asaari, M. S. M., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., & Scheunders, P. (2019). Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Computers and Electronics in Agriculture*, 162(March), 749–758. <https://doi.org/10.1016/j.compag.2019.05.018>
- Azimi, S., Kaur, T., & Gandhi, T. K. (2021a). A deep learning approach to measure stress level in plants due to Nitrogen deficiency. *Measurement: Journal of the International Measurement Confederation*, 173(September), 108650. <https://doi.org/10.1016/j.measurement.2020.108650>
- Azimi, S., Kaur, T., & Gandhi, T. K. (2021b). CHAPTER 3: A deep learning approach to measure stress level in plants due to Nitrogen deficiency. *Measurement: Journal of the International Measurement Confederation*, 173(June 2020), 108650. <https://doi.org/10.1016/j.measurement.2020.108650>
- Balasubramaniam, P., & Ananthi, V. P. (2016). Segmentation of nutrient deficiency in incomplete crop images using intuitionistic fuzzy C-means clustering algorithm. *Nonlinear Dynamics*, 83(1–2), 849–866. <https://doi.org/10.1007/s11071-015-2372-y>
- Ballesteros, R., Intrigliolo, D. S., Ortega, J. F., Ramírez-Cuesta, J. M., Buesa, I., & Moreno, M. A. (2020). Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. *Precision Agriculture*, 21(6), 1242–1262. <https://doi.org/10.1007/s11119-020-09717-3>

- Barbedo, J. G. A. (2019). Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. *Computers and Electronics in Agriculture*, 162(December 2018), 482–492. <https://doi.org/10.1016/j.compag.2019.04.035>
- Berni, J. A. J., Zarco-Tejada, P. J., Sepulcre-Cantó, G., Fereres, E., & Villalobos, F. (2009). Mapping canopy conductance and CWSI in olive orchards using high resolution thermal remote sensing imagery. *Remote Sensing of Environment*, 113(11), 2380–2388. <https://doi.org/10.1016/j.rse.2009.06.018>
- Bian, L., Zhang, H., Ge, Y., Čepl, J., Stejskal, J., & EL-Kassaby, Y. A. (2022). Closing the gap between phenotyping and genotyping: review of advanced, image-based phenotyping technologies in forestry. *Annals of Forest Science*, 79(1), 1–21. <https://doi.org/10.1186/s13595-022-01143-x>
- Buschmann, C., Langsdorf, G., & Lichtenthaler, H. K. (2001). Imaging of the blue, green, and red fluorescence emission of plants: An overview. In *Photosynthetica* (Vol. 38, Issue 4, pp. 483–491). <https://doi.org/10.1023/A:1012440903014>
- Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., & Klukas, C. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis with open. *Plant Cell*, 26(12), 4636–4655. <https://doi.org/10.1105/tpc.114.129601>
- Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151(November 2017), 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>
- Clarke, T. R. (1997). An empirical approach for detecting crop water stress using multispectral airborne sensors. In *HortTechnology* (Vol. 7, Issue 1, pp. 9–16). <https://doi.org/10.21273/horttech.7.1.9>
- Culman, M. A., Gomez, J. A., Talavera, J., Quiroz, L. A., Tobon, L. E., Aranda, J. M., Garreta, L. E., & Bayona, C. J. (2017). A Novel Application for Identification of Nutrient Deficiencies in Oil Palm Using the Internet of Things. *Proceedings - 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, MobileCloud 2017*, 169–172. <https://doi.org/10.1109/MobileCloud.2017.32>
- Das Choudhury, S., Maturu, S., Samal, A., Stoerger, V., & Awada, T. (2020). Leveraging Image Analysis to Compute 3D Plant Phenotypes Based on Voxel-Grid Plant Reconstruction. *Frontiers in Plant Science*, 11(December), 1–18. <https://doi.org/10.3389/fpls.2020.521431>
- Dodig, D., Božinović, S., Nikolić, A., Zorić, M., Vančetović, J., Ignjatović-Micić, D., Deliće, N., Weigelt-Fischer, K., Altmann, T., & Junker, A. (2021). Dynamics of

- Maize Vegetative Growth and Drought Adaptability Using Image-Based Phenotyping Under Controlled Conditions. *Frontiers in Plant Science*, 12(May), 1–18. <https://doi.org/10.3389/fpls.2021.652116>
- Duan, L., Han, J., Guo, Z., Tu, H., Yang, P., Zhang, D., Fan, Y., Chen, G., Xiong, L., Dai, M., Williams, K., Corke, F., Doonan, J. H., & Yang, W. (2018). Chapter 3. Novel digital features discriminate between drought resistant and drought sensitive rice under controlled and field conditions. *Frontiers in Plant Science*, 9(April). <https://doi.org/10.3389/fpls.2018.00492>
- Elsayed, S., Elhoweity, M., Ibrahim, H. H., Dewir, Y. H., Migdadi, H. M., & Schmidhalter, U. (2017). Thermal imaging and passive reflectance sensing to estimate the water status and grain yield of wheat under different irrigation regimes. *Agricultural Water Management*, 189, 98–110. <https://doi.org/10.1016/j.agwat.2017.05.001>
- Enciso, J., Maeda, M., Landivar, J., Jung, J., & Chang, A. (2017). A ground based platform for high throughput phenotyping. *Computers and Electronics in Agriculture*, 141, 286–291. <https://doi.org/10.1016/j.compag.2017.08.006>
- Eshkabilov, S., Lee, A., Sun, X., Lee, C. W., & Simsek, H. (2021). Hyperspectral imaging techniques for rapid detection of nutrient content of hydroponically grown lettuce cultivars. *Computers and Electronics in Agriculture*, 181(December 2020), 105968. <https://doi.org/10.1016/j.compag.2020.105968>
- Fiorani, F., & Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual Review of Plant Biology*, 64, 267–291. <https://doi.org/10.1146/annurev-arplant-050312-120137>
- Forero, M. G., Murcia, H. F., Méndez, D., & Betancourt-Lozano, J. (2022). LiDAR Platform for Acquisition of 3D Plant Phenotyping Database. *Plants*, 11(17). <https://doi.org/10.3390/plants11172199>
- Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., Hoyer, J. S., Lin, A., Liu, S., Lizárraga, C., Lorence, A., Miller, M., Platon, E., Tessman, M., & Sax, T. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ*, 2017(12). <https://doi.org/10.7717/peerj.4088>
- Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, 33(11). <https://doi.org/10.1029/2006GL026457>
- Gracia-Romero, A., Kefauver, S. C., Fernandez-Gallego, J. A., Vergara-Díaz, O., Nieto-Taladriz, M. T., & Araus, J. L. (2019). UAV and ground image-based phenotyping:



- A proof of concept with durum wheat. *Remote Sensing*, 11(10). <https://doi.org/10.3390/rs11101244>
- Graeff, S., Pfenning, J., Claupein, W., & Liebig, H. P. (2008). Evaluation of image analysis to determine the N-fertilizer demand of broccoli plants (*Brassica oleracea* convar. botrytis var. Italica). *Advances in Optical Technologies*, 2008. <https://doi.org/10.1155/2008/359760>
- Guo, Y., Chen, S., Li, X., Cunha, M., Jayavelu, S., Cammarano, D., & Fu, Y. H. (2022). Machine Learning-Based Approaches for Predicting SPAD Values of Maize Using Multi-Spectral Images. *Remote Sensing*, 14(6). <https://doi.org/10.3390/rs14061337>
- Gutiérrez, S., Fernández-Navales, J., Diago, M. P., & Tardaguila, J. (2018). On-the-go hyperspectral imaging under field conditions and machine learning for the classification of grapevine varieties. *Frontiers in Plant Science*, 9(July), 1–11. <https://doi.org/10.3389/fpls.2018.01102>
- Hassanijalilian, O., Igathinathane, C., Doetkott, C., Bajwa, S., Nowatzki, J., & Haji Esmaeili, S. A. (2020a). CHAPTER 2: Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Computers and Electronics in Agriculture*, 174(August 2019), 105433. <https://doi.org/10.1016/j.compag.2020.105433>
- Hassanijalilian, O., Igathinathane, C., Doetkott, C., Bajwa, S., Nowatzki, J., & Haji Esmaeili, S. A. (2020b). Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Computers and Electronics in Agriculture*, 174(April), 105433. <https://doi.org/10.1016/j.compag.2020.105433>
- Humplík, J. F., Lazár, D., Husíčková, A., & Spíchal, L. (2015). Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses - A review. *Plant Methods*, 11(1), 1–10. <https://doi.org/10.1186/s13007-015-0072-8>
- Ihuoma, S. O., & Madramootoo, C. A. (2017). Recent advances in crop water stress detection. *Computers and Electronics in Agriculture*, 141, 267–275. <https://doi.org/10.1016/j.compag.2017.07.026>
- Jay, S., Gorretta, N., Morel, J., Maupas, F., Bendoula, R., Rabatel, G., Dutartre, D., Comar, A., & Baret, F. (2017). Estimating leaf chlorophyll content in sugar beet canopies using millimeter- to centimeter-scale reflectance imagery. *Remote Sensing of Environment*, 198, 173–186. <https://doi.org/10.1016/j.rse.2017.06.008>
- Jones, J. B., Mills, H. A., Jones, J. B., Smittle, D. A., & Hussey, C. G. (1994). A Comparison of analytical methods for nitrogen analysis in plant tissues. *Communications in Soil Science and Plant Analysis*, 25(7–8), 943–954. <https://doi.org/10.1080/00103629409369090>

- Kim, J. Y., & Glenn, D. M. (2017). Multi-modal sensor system for plant water stress assessment. *Computers and Electronics in Agriculture*, 141, 27–34. <https://doi.org/10.1016/j.compag.2017.07.009>
- Kjeldahl, J. (1883). Neue Methode zur Bestimmung des Stickstoffs in organischen Körpern. *Zeitschrift Für Analytische Chemie*, 22(1), 366–382. <https://doi.org/10.1007/BF01338151>
- Kuska, M. T., Behmann, J., & Mahlein, A. K. (2018). Potential of hyperspectral imaging to detect and identify the impact of chemical warfare compounds on plant tissue. *Pure and Applied Chemistry*, 90(10), 1615–1624. <https://doi.org/10.1515/pac-2018-0102>
- Li, D., Li, C., Yao, Y., Li, M., & Liu, L. (2020). Modern imaging techniques in plant nutrition analysis: A review. *Computers and Electronics in Agriculture*, 174(May), 105459. <https://doi.org/10.1016/j.compag.2020.105459>
- Lobet, G., Draye, X., & Périlleux, C. (2013). An online database for plant image analysis software tools. *Plant Methods*, 9(1). <https://doi.org/10.1186/1746-4811-9-38>
- Lowe, A., Harrison, N., & French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods*, 13(1), 1–12. <https://doi.org/10.1186/s13007-017-0233-z>
- Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, 63(2), 282–293. <https://doi.org/10.1016/j.compag.2008.03.009>
- Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., & Scheunders, P. (2017). Close range hyperspectral imaging of plants: A review. *Biosystems Engineering*, 164, 49–67. <https://doi.org/10.1016/j.biosystemseng.2017.09.009>
- Mishra, P., Lohumi, S., Ahmad Khan, H., & Nordon, A. (2020). Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. *Computers and Electronics in Agriculture*, 178(July), 105780. <https://doi.org/10.1016/j.compag.2020.105780>
- Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., & Melgani, F. (2018). Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective. *GigaScience*, 8(1), 1–12. <https://doi.org/10.1093/gigascience/giy153>
- Mohd Asaari, M. S., Mishra, P., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., & Scheunders, P. (2018). Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping

- platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138, 121–138. <https://doi.org/10.1016/j.isprsjprs.2018.02.003>
- Möller, M., Alchanatis, V., Cohen, Y., Meron, M., Tsipris, J., Naor, A., Ostrovsky, V., Sprintsin, M., & Cohen, S. (2007). Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. *Journal of Experimental Botany*, 58(4), 827–838. <https://doi.org/10.1093/jxb/erl115>
- Naqvi, R. Z., Siddiqui, H. A., Mahmood, M. A., Najeebullah, S., Ehsan, A., Azhar, M., Farooq, M., Amin, I., Asad, S., Mukhtar, Z., Mansoor, S., & Asif, M. (2022). Smart breeding approaches in post-genomics era for developing climate-resilient food crops. *Frontiers in Plant Science*, 13(September), 1–19. <https://doi.org/10.3389/fpls.2022.972164>
- Nauš, J., Prokopová, J., Řebíček, J., & Špundová, M. (2010). SPAD chlorophyll meter reading can be pronouncedly affected by chloroplast movement. *Photosynthesis Research*, 105(3), 265–271. <https://doi.org/10.1007/s11120-010-9587-z>
- Nicolas Virlet, Sabermanesh, K., Sadeghi-tehran, P., & Hawkesford, M. J. (2017). Field Scanalyzer : An automated robotic field phenotyping platform for detailed crop monitoring. *Functional Plant Biology*, 44, 143–153.
- Onoyama, H., Ryu, C., Suguri, M., & Iida, M. (2015). Nitrogen prediction model of rice plant at panicle initiation stage using ground-based hyperspectral imaging: growing degree-days integrated model. *Precision Agriculture*, 16(5), 558–570. <https://doi.org/10.1007/s11119-015-9394-9>
- Pieruschka, R., & Schurr, U. (2019). Plant phenotyping: Past, present, and future. *Plant Phenomics*, 2019. <https://doi.org/10.34133/2019/7507131>
- Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1), 26–33. <https://doi.org/10.1016/j.jag.2005.06.002>
- Prasetyo, E., Adityo, R. D., Suciati, N., & Fatichah, C. (2017). CHAPTER 2: Mango leaf image segmentation on HSV and YCbCr color spaces using Otsu thresholding. *Proceeding - 2017 3rd International Conference on Science and Technology-Computer, ICST 2017*, 99–103. <https://doi.org/10.1109/ICSTC.2017.8011860>
- Proctor, C., Dao, P. D., & He, Y. (2021). Close-range, heavy-duty hyperspectral imaging for tracking drought impacts using the PROCOSINE model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 263, 107528. <https://doi.org/10.1016/j.jqsrt.2021.107528>
- Rahaman, M. M., Ahsan, M. A., & Chen, M. (2019). Data-mining Techniques for Image-

- based Plant Phenotypic Traits Identification and Classification. *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-55609-6>
- Romualdo, L. M., Luz, P. H. C., Devechio, F. F. S., Marin, M. A., Zúñiga, A. M. G., Bruno, O. M., & Herling, V. R. (2014). Use of artificial vision techniques for diagnostic of nitrogen nutritional status in maize plants. *Computers and Electronics in Agriculture*, 104, 63–70. <https://doi.org/10.1016/j.compag.2014.03.009>
- Sadeghi-Tehran, P., Virlet, N., Sabermanesh, K., & Hawkesford, M. J. (2017). Chapter 2. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. *Plant Methods*, 13(1), 1–17. <https://doi.org/10.1186/s13007-017-0253-8>
- Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, 82(7–8), 518–529. <https://doi.org/10.1002/mrd.22489>
- Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Schmundt, D., Stitt, M., Jähne, B., & Schurr, U. (1998). Quantitative analysis of the local rates of growth of dicot leaves at a high temporal and spatial resolution, using image sequence analysis. *Plant Journal*, 16(4), 505–514. <https://doi.org/10.1046/j.1365-313X.1998.00314.x>
- Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., & Sun, Q. (2021). Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Computers and Electronics in Agriculture*, 180(October 2020), 105860. <https://doi.org/10.1016/j.compag.2020.105860>
- Siedliska, A., Baranowski, P., Pastuszka-Woźniak, J., Zubik, M., & Krzyszcak, J. (2021). CHAPTER 4: Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. *BMC Plant Biology*, 21(1), 1–17. <https://doi.org/10.1186/s12870-020-02807-4>
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, 21(2), 110–124. <https://doi.org/10.1016/j.tplants.2015.10.015>
- Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in Plant Science*, 23(10), 883–898. <https://doi.org/10.1016/j.tplants.2018.07.004>
- State, T., & Insecurity, F. (2014). *The State of Food Insecurity in the World*.
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., & Bennett, M. (2017). Plant Phenomics,

- From Sensors to Knowledge. *Current Biology*, 27(15), R770–R783. <https://doi.org/10.1016/j.cub.2017.05.055>
- Thorburn, P., Fitch, P., Zhang, Y., Shendryk, Y., Webster, T., Biggs, J., Mooij, M., Ticehurst, C., Vilas, M., & Fielke, S. (2019). *Helping Farmers Mitigate Nutrient Losses “Digital Agriculture.”* 32, 1–6. <https://www.reefcredit.org/>
- Van Ittersum, M. K., Van Bussel, L. G. J., Wolf, J., Grassini, P., Van Wart, J., Guilpart, N., Claessens, L., De Groot, H., Wiebe, K., Mason-D’Croz, D., Yang, H., Boogaard, H., Van Oort, P. A. J., Van Loon, M. P., Saito, K., Adimo, O., Adjei-Nsiah, S., Agali, A., Bala, A., ... Cassman, K. G. (2016). Can sub-Saharan Africa feed itself? *Proceedings of the National Academy of Sciences of the United States of America*, 113(52), 14964–14969. <https://doi.org/10.1073/pnas.1610359113>
- Yamashita, H., Sonobe, R., Hirono, Y., Morita, A., & Ikka, T. (2020). CHAPTER 4: Dissection of hyperspectral reflectance to estimate nitrogen and chlorophyll contents in tea leaves based on machine learning algorithms. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-73745-2>
- Ye, X., Abe, S., & Zhang, S. (2020). Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging. *Precision Agriculture*, 21(1), 198–225. <https://doi.org/10.1007/s11119-019-09661-x>
- Zhang, J., Min, A., Steffenson, B. J., Su, W. H., Hirsch, C. D., Anderson, J., Wei, J., Ma, Q., & Yang, C. (2022). Wheat-Net: An Automatic Dense Wheat Spike Segmentation Method Based on an Optimized Hybrid Task Cascade Model. *Frontiers in Plant Science*, 13(February), 1–13. <https://doi.org/10.3389/fpls.2022.834938>
- Zhang, X., Liu, F., He, Y., & Gong, X. (2013). Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. *Biosystems Engineering*, 115(1), 56–65. <https://doi.org/10.1016/j.biosystemseng.2013.02.007>
- Zhang, Y., Zha, Y., Jin, X., Wang, Y., & Qiao, H. (2022). Changes in Vertical Phenotypic Traits of Rice (*Oryza sativa* L.) Response to Water Stress. *Frontiers in Plant Science*, 13(July), 1–19. <https://doi.org/10.3389/fpls.2022.942110>
- Zhuang, S., Wang, P., Jiang, B., Li, M., & Gong, Z. (2017a). CHAPTER 3: Early detection of water stress in maize based on digital images. *Computers and Electronics in Agriculture*, 140, 461–468. <https://doi.org/10.1016/j.compag.2017.06.022>
- Zhuang, S., Wang, P., Jiang, B., Li, M., & Gong, Z. (2017b). Chapter 3. Early detection of water stress in maize based on digital images. *Computers and Electronics in Agriculture*, 140, 461–468. <https://doi.org/10.1016/j.compag.2017.06.022>

## **2 Machine Learning Methods for Automatic Segmentation of Images of Field- and Glasshouse-Based Plants for High-Throughput Phenotyping**

Frank Gyan Okyere <sup>1,2</sup>, Daniel Cudjoe <sup>1,2</sup>, Pouria Sadeghi-Tehran<sup>1</sup>, Nicolas Virlet<sup>1</sup>, March Castle <sup>1</sup>, Andrew B. Riche<sup>1</sup>, Latifa Greche<sup>1</sup>, Fady Mohareb <sup>2</sup>, Daniel Simms <sup>2</sup>, Manal Mhada <sup>3</sup> and Malcolm John Hawkesford <sup>1, \*</sup>

<sup>1</sup>Sustainable Soils and Crops, Rothamsted Research, Harpenden, United Kingdom, 2 School of Water.

<sup>2</sup>Energy and Environment, Cranfield University, Cranfield, United Kingdom, 3 AgroBioSciences.

<sup>3</sup>Department, University of Mohammed VI Polytechnic, Ben Guerir, Morocco

Based on: Okyere, F.G.; Cudjoe, D.; Sadeghi-Tehran, P.; Virlet, N.; Riche, A.B.; Castle, M.; Greche, L.; Mohareb, F.; Simms, D.; Mhada, M.; et al. Machine Learning Methods for Automatic Segmentation of Images of Field- and Glasshouse-Based Plants for High-Throughput Phenotyping. *Plants* 2023, 12, 2035. <https://doi.org/10.3390/plants12102035>

### **2.1 Abstract**

Image segmentation is a fundamental but critical step for achieving automated high-throughput phenotyping. While conventional segmentation methods perform well in homogenous environments, the performance decreases when used in more complex environments. This study aimed to develop a fast and robust neural-network-based segmentation tool to phenotype plants in both field and glasshouse environments in a high-throughput manner. Digital images of cowpea (from glasshouse) and wheat (from field) with different nutrient supplies across their full growth cycle were acquired. Image patches from 20 randomly selected images from the acquired dataset were transformed from their original RGB format to multiple colour spaces. The pixels in the patches were annotated as foreground and background, with a pixel having a feature vector of 24 colour properties. A feature selection technique was applied to choose the sensitive features, which were used to train a multilayer perceptron network (MLP) and two other traditional machine learning models: support vector machines (SVMs) and random forest (RF). The

performance of these models, together with two standard colour-index segmentation techniques (excess green (ExG) and excess green–red (ExGR)), was compared. The proposed method outperformed the other methods in producing quality segmented images with over 98%-pixel classification accuracy. Regression models developed from the different segmentation methods to predict soil plant analysis development (SPAD) values of cowpea and wheat showed that images from the proposed MLP method produced models with high predictive power and comparably high accuracies. This method will be an essential tool for developing a data analysis pipeline for high-throughput plant phenotyping. The proposed technique can learn from different environmental conditions with a high level of robustness.

**Keywords:** feature extraction; imaging; machine learning; phenotyping; segmentation

## 2.2 Introduction

There is an increasing need to improve agriculture to meet the 2050 global food demand agenda. One aspect of this improvement is the implementation of high-throughput plant phenotyping (HTPP) to enable the large-scale evaluation of plant performances in breeding programs. HTPP involves automatically sensing plant physical features to understand the dynamic relationship between plants and their environment (Lee et al., 2018). This method could involve the use of non-invasive methods to collect plant data for the analysis of plant physiology, phenology and morphological traits by using sequentially acquired images from multiple sensors, including visible RGB, thermal, hyperspectral cameras, 3D laser scanners and fluorescence imagers, taken from multiple view angles (Sugiura et al., 2018). Digital RGB cameras are the most common and widely used image-sensing tools due to their relatively low cost, high resolution and portability (Sadeghi-Tehran et al., 2017). RGB cameras measure the light reflected from objects by using three broad bands centered at 650 nm (red), 550 nm (green) and 450 nm (blue) (Li et al., 2014). Segmentation is a critical step in processing digital images that separates and identifies regions of interest in images.

At the basic level, segmentation separates image pixels into two classes: foreground (i.e., plant vegetation) and background (i.e., non-plant objects). Over the years, several segmentation techniques have been developed, and broadly categorized as threshold, colour-index and learning-based methods (Hamuda et al., 2016). Threshold-

based methods separate image pixels into two groups based on their intensity values above or below a given threshold. This technique transforms an image from RGB to grayscale and binarizes the gray level based on the threshold value. The threshold is either global (using a single threshold value) or variable (threshold value varies). Binary, adaptive and Otsu thresholding are commonly used threshold-based segmentation techniques. Colour-index-based techniques use combinations of colour indices to differentiate plant vegetation from other non-plant objects. This method works on the assumption that the histogram distribution of an input image is bimodal and can be distinctively separated by a colour threshold index. The colour index of vegetation extraction (CIVE), excess green (ExG), excess green–red index (ExGR) and normalized difference index (NDI) are examples of commonly used colour-based segmentation techniques (Zheng et al., 2017). (Meyer & Neto, 2008) developed an improved colour index method (ExG-ExR) to segment greenhouse and field images of young soybean plants. The ExG–ExR index has a fixed, built-in zero threshold, so it does not need Otsu or any user-selected threshold value. Their results show the improved colour index method achieved the highest factor of  $0.88 \pm 0.12$  for greenhouse plants with soil residue. This proposed index performed remarkably well in segmenting plants with different soil-residue backgrounds showing superior vegetation separation accuracy over other methods like ExG\_Otsu and NDI+Otsu indices. Kirk et al(2009) utilized a combination of colour and threshold methods in estimating the leaf area index (LAI) of cereal crops. Their method used a combination of greenness and intensity-derived indices from the red and green spectrum. The intensity feature gives a better separation of the plants and nonplant matter, especially under variable sunny conditions. Their method is capable of estimating LAI with a similar accuracy to that of the LAI-2000 leaf area meter. Learning-based methods utilize supervised and unsupervised machine learning approaches to perform semantic and instance segmentation of plant images for disease detection (Iqbal & Talukder, 2020), plant growth analysis (Lee et al., 2018), leaf counting (Aich & Stavness, 2017), wheat spike counting (Zhang et al., 2022), etc. Examples include k-means clustering, supervised mean-shift algorithm, fuzzy clustering, etc. Studies have shown that machine learning models outperform the segmentation methods described above when trained with carefully selected features. Adams et al. (2020) utilized a supervised learning method to segment plants by classifying pixels-based datasets whose labels were



produced using k-means clustering algorithm. Their results showed a better performance of their approach than traditional thresholding methods. Deep learning (DL), a branch of the learning-based method, has been extensively applied to segment plants with different backgrounds. A commonly used deep learning architecture called the convolutional neural network has gained much attention in computer-vision-based plant phenotyping. DL approaches for the semantic segmentation of plants, including Region-based CNN (R-CNN), pyramid networks and encoder–decoder-based models, rely on CNN modelling (Chen et al., 2018). These methods have been extensively used in plant organ segmentation for root and shoot phenotyping. These include the works of Aich & Stavness(2017), who used SegNet architecture, a deconvolutional network without a fully connected layer to segment plant leaves from its background. Morris (2018) also performed instance segmentation to detect and discriminate leaf boundaries using a pyramid convolutional neural network. Recently, Narisetti et al.,(2022) developed a fully automated segmentation pipeline based on pre-trained U-Net models to phenotype shoots of different crops. Although DL methods have been extensively used in plant segmentation and phenotyping, they require a large-volume labelled dataset, which is not readily available (Jiang & Li, 2020). Furthermore, DL models have many trainable parameters (e.g., ResNet18 has around 11 million, while DenseNet has around 8 million trainable parameters), resulting in slow data processing, particularly when dealing with high-throughput data. This forms a bottleneck around plant phenotyping, especially when dealing with a small number of plant datasets. Due to these limitations, deep learning approaches were not considered in this study.

Many segmentation methods rely on colour information (Guijarro et al., 2015), which is a strong descriptor used in object detection and recognition. However, colour-based segmentation may not perform well in certain scenarios, such as cluttered backgrounds, variable illumination conditions, etc.). Plant colour can vary depending on the position of the sun during image acquisition (Williamson et al., 1984), which can lead to poor segmentation results due to the effect of variable illumination conditions on colour dynamics. Different colour spaces can minimize the effects of illumination and colour variability on segmentation. Colour spaces provide additional information about pixel intensity and colour distribution and can be used individually or in combination to segment plant images (Ganesan et al., 2017). Prasetyo et al., (2017) combined different

colour spaces to segment and detect mango trees. They transformed the RGB colour space to HSV (Hue Saturation Value) and YCbCr (luminance chrominance) and combined the individual colour channels with Otsu thresholding for segmentation. Their results showed that Cr from YCbCr effectively separated mango trees from the background. Similarly, Sadeghi-Tehran et al., (2017) used multiple colour features obtained from different colour spaces to train a supervised machine learning classifier for wheat ear segmentation on the field. They found that using multiple colour spaces improved segmentation accuracy and was more robust to background noise and illumination changes than using a single-colour space.

Although using different colour spaces for segmentation can be effective, it can also be computationally expensive, particularly when computing millions of pixels which requires high computer speed and large memory capacity for processing. In addition, some of the individual colour channels may have redundant effects on the segmentation process, while others may have little or no effect. Furthermore, this method may fail to properly segment plants with different shades of colour and texture across different parts of the plant image if the appropriate colour features are not utilized. This study sought to (1) develop an automated phenotyping pipeline based on an artificial neural network to segment field- and glasshouse-based plant images and (2) predict plant soil plant analysis development (SPAD) values based on the proposed and other existing segmentation methods.

## **2.3 Materials and Methods**

Experiments were set up in a glasshouse (cowpea) and the field (wheat). The segmentation pipeline, as shown in Figure 2.1, involves (1) image acquisition from two different environments, (2) image annotation and colour transformation, (3) feature extraction and selection and (4) modelling of the machine learning classifier.

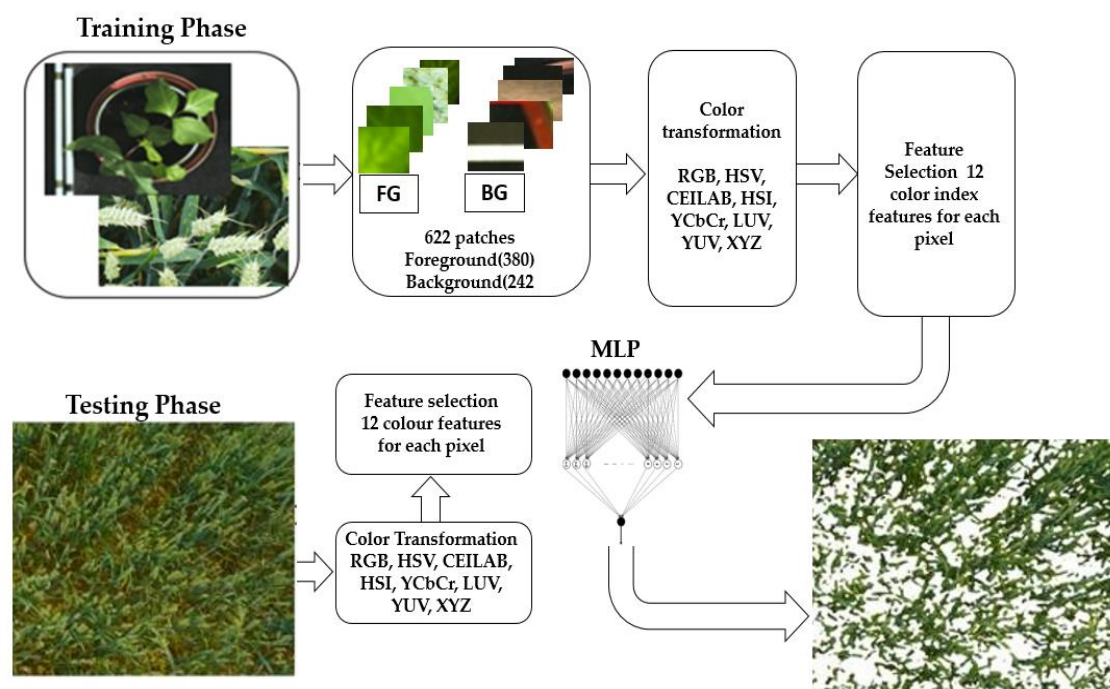


Figure 2-1 Diagrammatic representation of the proposed method for glasshouse and field image, using MLP, multilayer perceptron. The steps involved in this method include: image acquisition and preprocessing, patch extraction (for background (BG) and foreground (FG)), colour transformation, feature selection and training a MLP classifier using the annotated pixel features as input.

## 2.3.1 Experimental Setup and Imaging

### 2.3.1.1 Cowpea grown in the glasshouse.

*Vigna unguiculata* (cowpea) plants were grown at the Plant Growth Facility (<https://www.cranfield.ac.uk/facilities/plant-growth-facility>, assessed on 10<sup>th</sup> March 2023) located at Cranfield University (Cranfield, Bedfordshire UK). The plants were grown in pots filled with compost that was washed and depleted of macro- and micronutrients. The compost was then reconstituted with prepared nutrient solutions of defined composition. The compost-washing method followed the method by Masters-Clark et al (2020), with some modifications. This was achieved by flooding one-part compost to five-part deionized water, mixing and breaking aggregates. The soluble solution was drained through a double-headed 0.5  $\mu\text{m}$  sieve. The steps were repeated five times and the depleted-nutrient compost was dried in an oven at 70  $^{\circ}\text{C}$  for 72 h. After drying, the compost was mixed with a nutrient solution prepared according to the

treatment required. Four nutrient solutions were prepared based on the Letcombe nutrient solution (Masters-Clark et al., 2020), comprising a high level of nitrogen high-level phosphorus (HNHP), high-level nitrogen low-level phosphorus (HNLP), low-level nitrogen high-level phosphorus (LNHP) and low-level nitrogen low-level of phosphorus (LNLP). The concentrations were 49.1 mM and 14.6 mM N for HN and LN and 13.4 mM and 3.3 mM P for HP and LP, respectively. Each treatment was replicated five times. The different treatments induced changes in leaves' pigmentation, resulting in variations in spectral characteristics (colour) during plant development. During their development stages, plants with adequate nitrogen levels exhibited green leaves, while deficiency was characterized by chlorosis progressing from light green to yellow to brown and phosphorus deficiency inhibited shoot growth and decolourized leaves from a blue-green colour to pale green/yellow in severely affected regions (McCauley et al., 2011). Digital images were taken twice weekly, using a high-resolution ( $3296 \times 2472$  pixels) RGB camera (colour 12-bit Prosilica GT3300) installed on the Lemnatec phenotyping platform in the glasshouse. The camera was set 3.5 m above the top plant canopy, covering an area of  $0.48 \text{ m}^2$  during imaging. Imaging commenced at the vegetative growth stage, 28 DAS (days after sowing), when the first trifoliate leaf had unfolded at node three and ended at the physiological maturity stage (90 DAS), when the seed pods were matured and ready for harvesting. The format of all the images used in this study was JPEG. Supplementary information (Figure B-1) shows samples of cowpea plants at various growth stages in the glasshouse

### **2.3.1.2 Wheat grown on the field**

Wheat (*Triticum aestivum*), cv Chilham, was cultivated in the field at Rothamsted Research, Harpenden, UK, from October 2020 to August 2021. Rows were planted 15 cm apart, with 350 seeds/ $\text{m}^2$  sowing density, on a  $9 \text{ m} \times 1.8 \text{ m}$  plot. The soil phosphorus (P) treatments were established by successive cropping and differential P application over several years at 12 Olsen phosphorus rates (approximately 3, 6, 9, 12, 15, 18, 21, 25, 30, 40, 50 and 60 ppm). The phosphorus-deficient treatments induced colour variations, especially on the leaves, veins and stems, resulting in purple (Barrett-lennard et al., 1982). Digital images were taken weekly, from the jointing to the booting stage, producing over 200 images. A digital camera-Sony Alpha a6300 (Sony Electronics, Sony UK, Weybridge, (Surrey), UK) was used to capture the plant images. The camera was mounted

on a 1.5 m monopod (Joalland et al., 2017) and held perpendicular to the ground, approximately 2 m above the canopy. A cable release trigger (UR-232-Tasman) was attached to the monopod to control the camera trigger. The setup was centered in the middle of the plots. The camera was set to automatic exposure mode to compensate for the changes in natural light. Supplementary information (Figure B-2) presents samples of wheat plants under two phosphorus levels (3 and 60 ppm) at various growth stages obtained from the field.

### **2.3.2 White balance and colour correction of plant images**

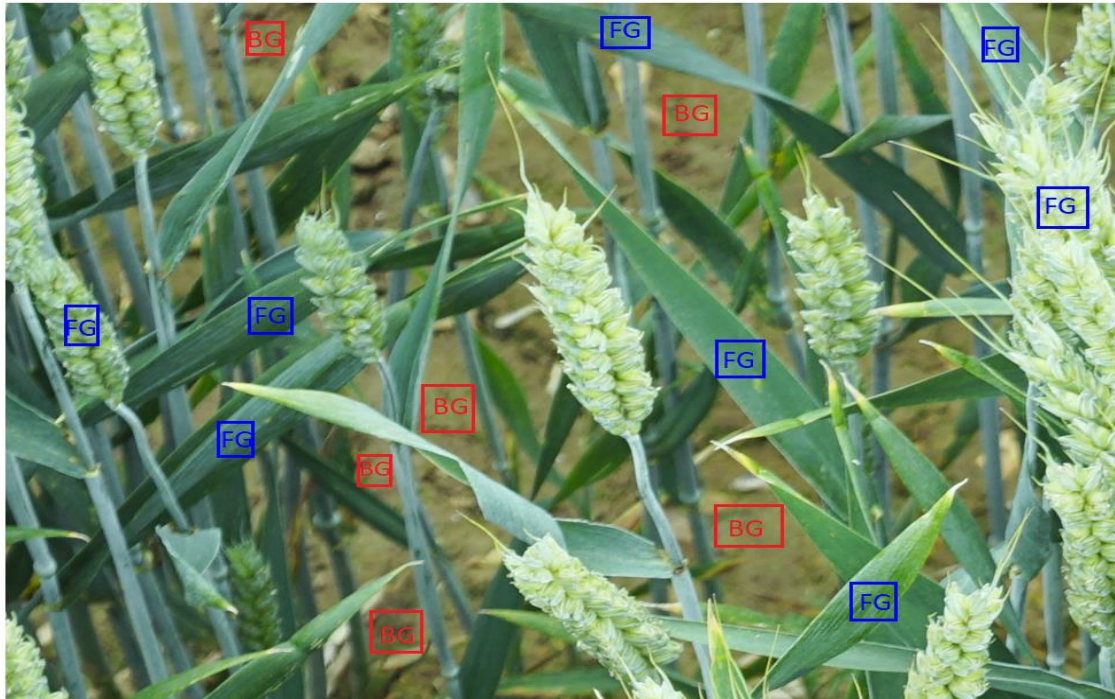
Changes in light intensity and brightness throughout the day can impact the appearance of plant colour in images, potentially leading to inaccurate interpretations (Sunoj et al., 2021). To address this issue, we utilized colour normalization/correction techniques. The goal of colour correction is to transform image pixels so that the pixel values of a reference (in this case, colour checker tiles) match the transformed images (Chopin et al., 2018). To accomplish this, we compared the pixel values of the captured images to the true colorimetric values of a colour checker passport (X-Rite Pantone, Manchester, UK), which contains 24 squares of colours ranging from gray to brown (H. S. Naik et al., 2017). The gray squares were used to establish the white balance, while the green, brown and yellow squares were used to calibrate the corresponding hue values. The calibration process identified how much the hue value of the green, brown and yellow squares on the colour checker deviated from the standard hue values. The correction was then applied to the plant images.

### **2.3.3 Machine learning based segmentation process**

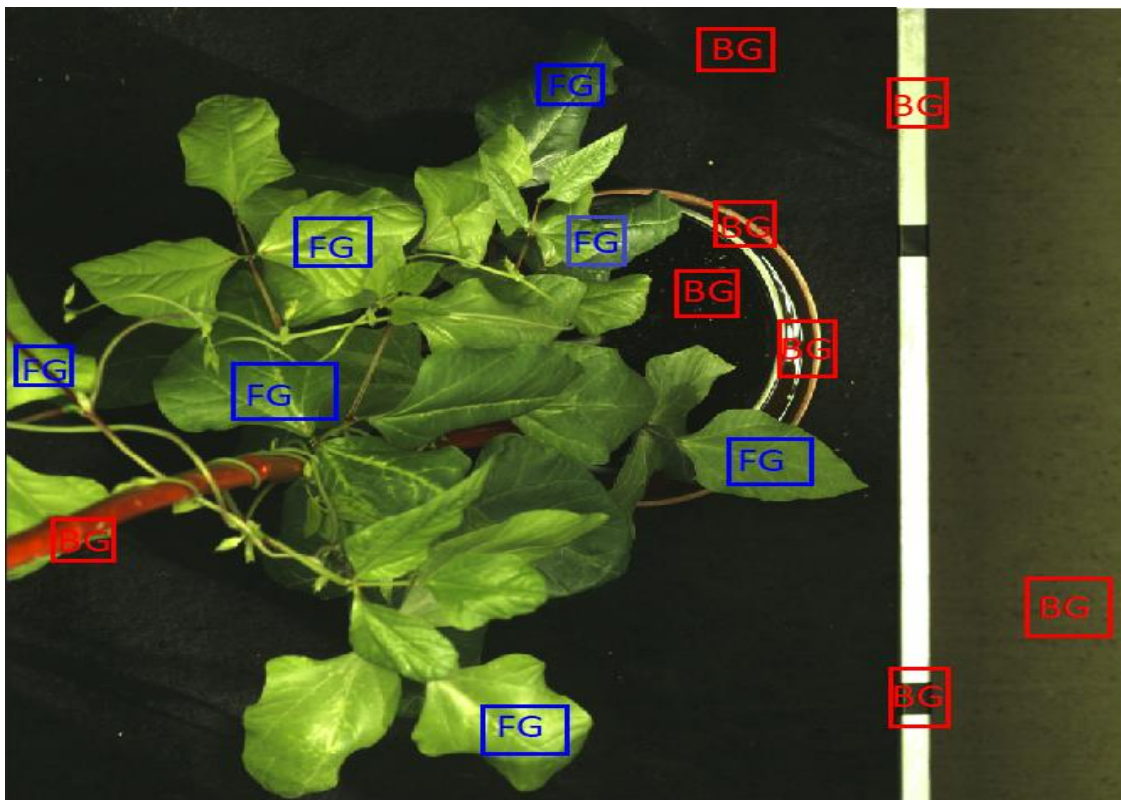
#### **2.3.3.1 Image annotation and colour transformation**

During the image-processing, illumination differences can create difficulties in achieving accurate segmentation. Various colour spaces possess unique characteristics that can minimize the impact of illumination on images and can resist noise variations during processing. Based on this concept, colour features were extracted from different colour channels to build a segmentation model. To accomplish this, image patches were created for plants and non-plant materials at various growth stages, which were manually annotated (Dutta & Zisserman, 2019) as the foreground and background (Figure 2.2). The image patches refer to smaller localized sections of the image for annotation and

classification (for segmentation). The patches were extracted and annotated by drawing rectangular regions of interest (i.e. bounding boxes) for foreground (plant images) and background (for non-plants objects). This was achieved using the VGG Image Annotator software (Dutta & Zisserman, 2019). The patches are made of pixels with  $n$  number of features for the classification model. The patches were manually chosen by picking randomly any space within the plants (as foreground) or non-plant (as background). The foreground annotation consisted of 1400 patches (800 from field and 600 from glasshouse images), while the background included 1100 patches (600 from field images and 500 from glasshouse images). Each patch was then resized to a  $20 \times 20$  pixels patch and transformed into eight different sets of colour channels, including RGB, HSV, YCbCr, Lab, YUV (Y = luma or brightness, U = blue projection and V = red projection), Luv (L = luminance, u = blue axis and v = red axis), hls (h = hue, l = lightness and s = saturation) and XYZ (chromatic coordinates). Since the patches are made of pixels, resizing each patch to  $20 \times 20$  size gives a patch made of 400 pixels. That is, each patch represents a  $m \times n$  matrix, where  $m$  is the number of pixels and  $n$  is the number of colour channels. Hence, for each image patch,  $(20 \times 20) \times 24$ -pixel features were obtained. In this case, each pixel with eight transformed colour channels gives a feature vector of size 24. With 1400 foreground and 1100 background patches, 24,000,000-pixel features (foreground = 13,440,000 and background = 10,560,000) were used to train a binary classifier for segmentation.



(a) Wheat image annotation (field).



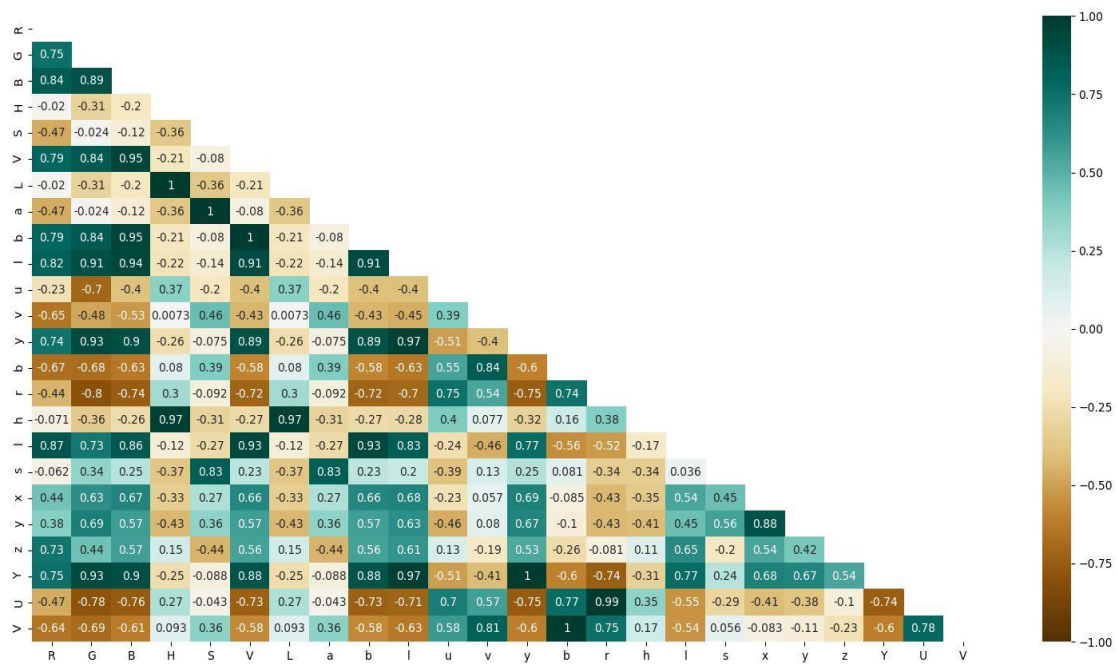
(b) Cowpea image annotation (glasshouse).



**Figure 2-2** Annotation of images into foreground and background patches for feature extraction. (a) Wheat image annotation and (b) Cowpea image annotation. The FG represents the foreground annotation and the BG is the background annotation. Patches were extracted from two crops (wheat from the field and cowpea from the glasshouse)

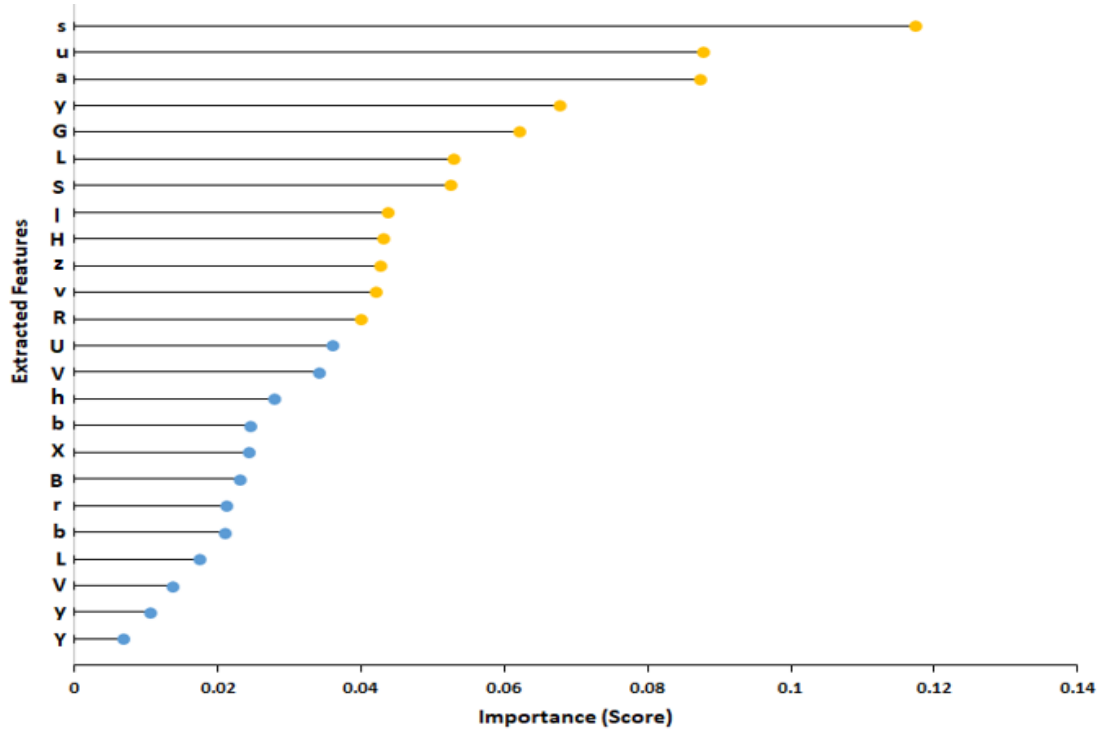
### 2.3.3.2 Feature extraction and selection

Training a binary classifier with many pixel features can be challenging due to high computational memory and speed requirements. Additionally, some features were observed to be highly correlated (Figure 2.3) and redundant, leading to similar effects on the model. To address these issues, a feature-selection method called decision-tree-based recursive feature elimination (DT-RFE) was applied. DT-RFE is a wrapper-type feature selection method that uses a backward feature illumination technique to remove features in order of importance. It works on the principle of decreasing the size of a feature set by recursively selecting the most important ones. The method first trains the decision tree model with a set of features. A significance mark for each feature is obtained, and weights are assigned to each feature. Then, a ranking criterion of all features is computed by using the sum of the weights of each feature in the model (Zhang et al., 2021). The features are then sorted according to the final score (weighted sum) and the smallest ranking features are removed. The process repeats until the desired number of highly sensitive features is achieved (Zulfikar et al., 2019; Darst et al., 2018).





(a) Heat map of all the extracted features for feature selection using the Pearson correlation.



(b) A plot of feature importance for feature selection

**Figure 2-3** Feature-selection process involving correlation analysis and feature ranking based on importance score. (a) is the heatmap of all extracted features and (b) is a plot of feature importance for feature selection. Abbreviations: R, G,B (R = red, G = green and B = blue channels); H, S,V (H = hue, S = saturation and V = value); ybr (y = luma, b = blue component and r = red component); Lab (L = lightness and a and b = chromaticity); YUV (Y = luma or brightness, U = blue projection and V = red projection); Luv (L = luminance, u = blue axis and v = red axis); hls (h = hue, l = lightness and s = saturation); and XYZ (X and Z = spectral weighting curves and Y = luminance).

This study selected 12 features with an importance score above a 0.04 threshold as input features for training the classifier (Figure 2.3b). This threshold was selected based on a trial experiment with different feature combinations. It was observed that a combination of the 12 selected channels (with a high importance score) resulted in high classification accuracy, whilst additional channels had no impact on the overall accuracy. Moreover, the classification accuracy dropped when fewer than 12 channels were used. The important features selected for training the binary classifier are shown in Figure 2.3b.

This includes the  $R$  and  $G$  (from RGB),  $H$  and  $S$  (from HSV),  $L$  and  $a$  (from Lab),  $v$  (from Luv),  $y$  (from YCbCr),  $u$  (from YUV),  $l$  and  $s$  (from hls) and  $Z$  (from XYZ). The selected channels have properties such as high contrast, illumination invariant and colour differentiation that help in accurately separating plant pixels from non-plant pixels. For instance, the  $H$  and  $S$  channels separate the colour information from intensity, reducing the effect of illumination and shadows on the robustness of the segmentation method. Additionally, the  $L$  and  $a$  components differentiate shades of colour, with  $L$  indicating lightness and  $a$  showing the red/green coordinate, which is dominant in the colour makeup of plant pixels. Colour transformation channels and their impact on image processing are highlighted in the work of Kumar & Miklavcic (2018)

### 2.3.3.3 Training and testing the model

An artificial neural network (ANN) was utilized to perform the plant image segmentation using a typical MLP architecture. MLP is a supervised neural network that maps a set of input features to a set of output in a feedforward propagation manner (Atkinson & Tatnall, 1997). It comprises a set of processing elements (neurons) that form different layers (Desai & Shah, 2021). It has three main layers: an input layer, one or more hidden layers and an output layer. The input layer takes in the values of the input data, while the output layer has one unit for each value of the MLP's output. (It has a single unit for regression or binary classification and  $M$  units for  $M$ -class classification.) The hidden layers are fixed between these two layers and are fully connected, transmitting information to the output layer. The neurons in the different layers are interconnected to each other through weights whose values are updated after a learning process (Barbedo, 2018). The weights define the importance of the interactions between the neurons in order to control the errors in the output layer (Neto et al., 2021). Inputs from the last hidden layer converge to a single value in the output layer (Hesami et al., 2019).

The MLP is trained on a dataset with  $X$  number of inputs:  $X = (x_1, x_2, x_3, x_4, \dots, x_n)$  and gives  $y$  number of outputs  $Y = (y_1, y_2, y_3, \dots, y_n)$ , using a linear or nonlinear function  $f(\cdot): R^x \rightarrow R^y$  to solve classification and regression problems (Lo & Marculescu, 2019). The success of an MLP network is dependent on parameters such as the activation function and loss error, which can be expressed mathematically as a linear combination, as shown in Equation 2.1 (Thomas & Suhner, 2015):

$$P = \left[ \sum_{j=1}^n \left( W_{kj} \alpha_h \sum_{j=1}^m (W_{ji} X_i + W_{jb}) \right) + W_{kb} \right] \quad \text{Equation 2.1}$$

where P is the predicted output; n is the number of output neurons;  $W_{kj}$  is the weight for the connecting neuron of hidden and output layers;  $\alpha_h$  is the hidden neuron's activation function; and m is the number of hidden neurons;  $W_{ji}$  is the weight for the connecting neuron of input and hidden layers;  $X_i$  is the input variable;  $W_{jb}$  is the bias for the hidden neuron;  $W_{kb}$  is the bias for the output neuron (Arulmozhi, 2021); and  $\alpha_h$  and  $\alpha_0$  are the activation functions for the hidden neuron and the output neuron, respectively. The activation function is a key part of a neural network. It is a mathematical expression that defines how the weighted sum is transformed from one layer to another. Basically, it tells whether a node should be activated or not. It is calculated from the product of the weight and input and adding a bias. Commonly used activation functions include linear, sigmoid, rectified linear unit (ReLU), etc. The activation function is usually linear for regression problems and sigmoid or rectified linear unit for classification problems (Thomas & Suhner, 2015). The default activation function for the hidden layers is ReLU.

In this study, we employed a modified version of the MLP network used by Sari et al.(2012). This network was trained on the principle of backpropagation algorithm (Hornik et al., 1989). This algorithm involves the forward propagation of data to calculate the loss between the target and predicted value, followed by the backward propagation through weight adjustment to reduce the loss (Aslam et al., 2020). The weights are updated after each iteration based on the expression in Equation 2.2:

$$W_u^* = W_o - a \left( \frac{\partial Error}{\partial W_o} \right) \quad \text{Equation 2.2}$$

where  $W_u^*$  is the updated weight,  $W_o$  is the previous weight, a is learning rate and  $\partial Error$  is the derivative of error with respect to the weight.

A preliminary study was conducted to select the number of hidden layers with the 12 input features per pixel. We tested 1 hidden layer (with 3 neurons), 2 hidden layers (3 and 10 neurons) and 4 hidden layers (3, 5, 10 and 20 neurons) and the results showed that there was no significant difference in the model performance when the number of hidden

layers was increased. Hence, a grid search cross-validation ( $k = 10$ ) was conducted to select the number of nodes in the hidden layer. MLP was experimentally set to 12, 20, 35, 60 and 100 neurons. ReLU was used as the activation function for processing the nodes of the MLP model (Oh et al., 2021). MLP models were selected as the prime model for this study because, although conventionally, MLP models require a large number of datasets, they can be effective with small number of datasets as compared to CNN models (Niza El Aisnadaa et al., 2024). MLP models can also perform well when trained with carefully selected features which captures the relevant information about the underlying data. MLP models with selected features can lead to better performance, reduced overfitting and faster training times (Naik and kiran, 2021). Furthermore, the MLP models trained with extracted features offer high interpretability, is effective in cases of limited computational resources.

#### **2.3.4 Regression for SPAD values predictions**

SPAD readings are used to determine the greenness of plants and this is highly correlated with their nitrogen content (Nguyen et al., 2016). In this study, the plant chlorophyll index was measured using a SPAD meter (SPAD-502, Konica Minolta Sensing, Osaka, Japan) at different growth stages of the plants. For the field experiment, three readings were taken from five randomly selected wheat leaves within each experimental plot. SPAD readings were taken on the first, second and third sectors of each fully expanded leaf. The final reading for each plot was the average of the fifteen measured SPAD values. In the glasshouse experiments, SPAD readings were taken concurrently with the image acquisition from the cowpea plant canopy. Three readings were taken from fully expanded leaves located at the top, middle and bottom part of each plant in each pot. The final plant SPAD values were calculated as the average of these readings.

Quality plants segmentation is crucial for the precise and non-destructive measurement of plant traits. Poorly segmented images can result in higher levels of pixel misclassifications during segmentation, which may ultimately affect the analysis of plant traits. To evaluate the performance of the segmentation methods used in this study, we implemented a gradient-boosting decision trees (GBRTs) algorithm called eXtreme Gradient Boosting (XGBoost) as a regression model to predict SPAD readings of plants using the segmented image data. XGBoost is an ensemble machine learning algorithm for classification or regression predictive modeling problems (Chen & Guestrin, 2016). It is computationally efficient and highly effective for structured datasets on classification and regression predictive modeling problems (Chen & Guestrin, 2016). XGBoost builds regression trees one by one such that the subsequent tree is trained from the

residuals of the previous tree. That is, the new model (tree) improves on its performance by correcting errors made by the previously trained model (tree) (Pesantez-Narvaez et al., 2019). To build an XGBoost model, an initial tree (model)  $X_0$  defined to predict a target variable  $y$  is associated with a residual  $(y - X_0)$ . A new model,  $X_1$ , is fit to the residuals from the previous model. In the next step, the previous model is combined with the new model  $X_1$  to produce a boosted version of  $X_0$ . In this case, the mean squared error from  $X_1$  will be lower than that from  $X_0$ . To improve the performance of  $X_1$ , the model could be boosted by modeling after the residuals of  $X_1$  to create a new model  $X_2$ . This process can be repeated until the residuals have been reduced as much as possible. In this work, the XGBoost implementation uses a histogram-based split finding method, which groups features into discrete bins to improve the training speed and reduce memory usage. The hyperparameters of the XGBoost, including  $n\_estimators$  (the number of trees used in the model), subsample size used to train each tree, maximum depth (max-depth) of trees and learning rate( $\alpha$ )- (indicating how fast the model learns), were fine-tuned to optimize the performance of the models. XGBoost was selected for the regression problem because of its strong predictive performances. This model is based on the gradient boosting algorithm which combines weak learners to form a stronger learner making it highly accurate and precise. Moreover, XGBoost can handle complex data and is robust to overfitting (due to the built-in regularization (both lasso and ridge) which prevents overfitting. Additionally, XGBoost is flexible in handling different loss function making it more efficient and is designed to be highly efficient with optimizations such as: parallelized tree building, out-of-core computation and distributed computing for large-scale problems. Due to its high efficiency in terms of both speed and memory, it is suitable for both small and large datasets often encountered in regression problems. The segmentation algorithms and extraction of plant traits were performed in the Python 3.7 environment (Python Software Foundation, [http: www.python.org/](http://www.python.org/), assessed on 1 August 2022) with the scikit-learn, opencv2, and SciPy packages. The entire machine learning process was carried out on an Intel Core computer with a 2 GHz processor, 8.00 GB of RAM, and the MS Windows 10 operating system.

### 2.3.5 Comparing segmentation performance

The performance of the MLP segmentation method was compared to the two ML-based segmentation methods and two well-known colour-threshold methods: ExG (excess green) and ExGR (excess green–red). ExG is a simple colour-index-based method that provides a clear distinction between soil and plant vegetation by increasing the contrast of the green part of the spectrum against the red and blue parts (Hindman & Meyer, 1998). It operates on the principle of bimodal distribution of pixels and is given by the expression  $ExG = 2 \times G - R - B$ , where R, G and B are the pixel values of the red, green and blue colour channels, respectively. ExGR was introduced by Meyer et al (2004), where two colour indices, ExG and excess red (ExR), were combined to reduce the residual background

pixels, achieving a high segmentation accuracy. ExGR is expressed as  $ExG - (1.4 \times R - G)$ , where  $R$  and  $G$  are the mean values of the red and green channels, respectively and  $ExG$  is the excess green index. Both  $ExG$  and  $ExGR$  indices were combined with automatic Otsu thresholding to segment the images.

To evaluate the performance of the different methods, ten images each for glasshouse and field plants were randomly selected and segmented using all five segmentation methods. The performance of all the segmentation methods was evaluated with reference images in which plant vegetations were manually segmented from the ten plants, using Adobe Photoshop Software (Adobe Inc., Maidenhead, UK). Figure 2.4 shows two examples of the test images and the manually segmented reference images for glasshouse and field conditions. Three quality factors, namely  $Q_{seg}$  (Meyer & Neto, 2008),  $S_r$  (W. Guo et al., 2013) and an error factor ( $Es$ ) (Yu et al., 2017), expressed in Equations 2.3 – 2.5 were used to evaluate the quality of the segmentation methods.

$$Q_{seg} = \frac{\sum_{x,y=0}^{x,y=h,w} (S(p)_{x,y} \cap R(p)_{x,y})}{\sum_{x,y=0}^{x,y=h,w} (S(p)_{x,y} \cup R(p)_{x,y})} \quad \text{Equation 2.3}$$

$$S_r = \frac{\sum_{x,y=0}^{x,y=h,w} (S(p)_{x,y} \cap R(p)_{x,y})}{\sum_{x,y=0}^{x,y=h,w} R(p)_{x,y}} \quad \text{Equation 2.4}$$

$$Es = \frac{\sum_{x,y=0}^{x,y=h,w} (S(p)_{x,y} \cap \neg R(p)_{x,y})}{\sum_{x,y=0}^{x,y=h,w} R(p)_{x,y}} \quad \text{Equation 2.5}$$

where  $S$  is the segmented plant,  $S(p) = 255$ , or background pixels  $S(p) = 0$ ;  $R$  is the reference image; indices  $h$  and  $w$  are the height and width of the image, respectively; and  $x$  and  $y$  are the pixels' coordinates. The segmentation accuracy is based on the logical operations logical or ( $\cup$ ), logical  $\cap$  ( $\cap$ ) and logical not ( $\neg$ ). It is compared on a pixel-by-pixel basis of the segmented image ( $S$ ) to the reference image ( $R$ ).

$Q_{seg}$  measures the consistency on a pixel-by-pixel basis between the  $S$  and  $R$  images and ranges between 0 and 1, where 1 is perfect segmentation and vice versa. Similarly,  $S_r$  measures the consistency of plant pixels between image regions (Yu et al., 2017). The  $Es$  measures the section of misclassified plant pixels in relation to the true total plant pixels selected from dataset.

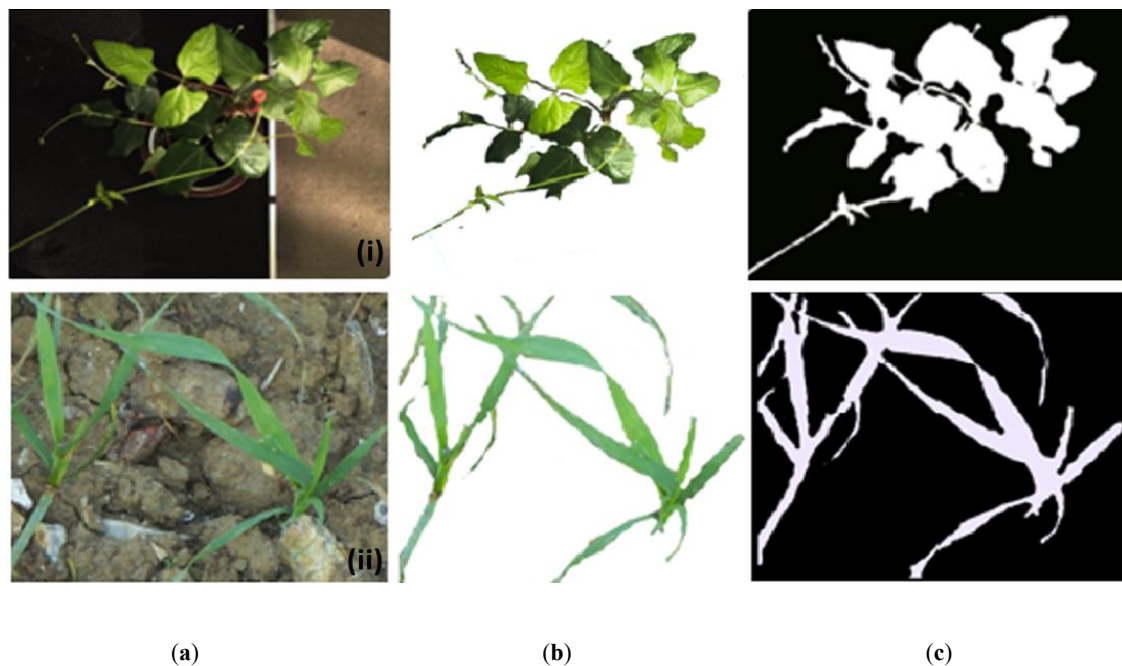


**Figure 2-4** Examples of images obtained from the glasshouse and field, segmented as reference images. The manually segmented images were obtained using Adobe Photoshop software (a) Original image, (b) manually segmented image and (c) binary image. Reference images were randomly selected from the dataset with different nutrient content and illumination and at variable growth stages.

### 2.3.6 Evaluating classification model performance

To compare the performance of the MLP classifier with other methods, classical machine learning algorithms—random forest (RF) and support vector machine (SVM)—were employed to segment the plant images. RF is an ensemble learning method composed of a collection of multiple decision trees capable of modelling nonlinear variable interactions while avoiding overfitting. It evaluates an optimal set of decision trees that can analyze high-dimensional data. Each decision tree is developed using a set of data that does not depend on earlier decision trees. In this study, the probability distribution of each decision tree was aggregated and averaged to provide the final classification by labelling each pixel as 0 (background) or 1 (foreground). SVM is a supervised learning method that is useful for the classification and regression of complex variables. In classification, an SVM model fits a hyperplane to a set of input data that separates one class from another. The best hyperplane is the one with the maximum margin between a set of classes (Akbarzadeh et al., 2018). Generally, SVM model performs better and is more tolerant to irrelevant and interdependent features.

To assess the impact of feature selection on the performance of the models, the ML models (MLP, SVM and RF) were trained with two features sets: all features (AF) and selected features (SF). The performance of the models was evaluated using the average accuracy score, F-score and recall score metrics, as given in Equations 2.6– 2.8:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad \text{Equation 2.6}$$

$$\text{R} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{Equation 2.7}$$

$$\text{F-Score} = 2 \frac{\text{P} \times \text{R}}{\text{P} + \text{R}} \quad \text{Equation 2.8}$$

where  $\text{P} = \frac{\text{TP}}{\text{TP} + \text{FP}}$ ; P is precision, R is recall, TP is true positive, TN is true negative, FN is false negative and FP is false negative. Note that a true positive occurs when the model correctly predicts the positive class (plant pixels) while a true negative occurs when the model correctly predicts the negative class (a non-plant pixels) Each model was trained on the whole dataset, which was divided into a 3:1:1 ratio for training, validation and testing subsets, respectively. The validation subset was used to test the model during training, while the testing subset (independent data) was used to test the model after training. Supplementary information (Table B-1) shows the amount of data used for each feature dataset for training, validating and testing the models. To ensure a robust and stable model, the training and testing procedures were repeated 100 times (with a random seed set to 0–99) and the results represent the average classification metrics for each set of data. Table 2-1 outlines the optimal hyperparameters used to train each model.

**Table 2-1** Optimal hyperparameters for the different classification models



| Hyper Parameters    | Model |       |     |
|---------------------|-------|-------|-----|
|                     | MLP   | SVM   | RF  |
| Hidden layers       | 1     |       | -   |
| Neurons             | 20    |       | -   |
| Activation function | ReLU  |       | -   |
| Kernel              | -     | RBF   | -   |
| C                   | -     | 0.001 | -   |
| $\gamma$            | -     | 1     | -   |
| n_estimator         | -     | -     | 200 |
| max_depth           | -     | -     | 20  |
| min_samples_leaf    | -     | -     | 4   |

### 2.3.7 Assessment of performance of SPAD prediction models

Three metric scores, namely the root mean square error (RMSE), adjusted determination coefficient ( $R^2$ ) and the mean absolute error (MAE), were used to comparatively assess the performance of the regression models. RMSE and MAE were used to evaluate the accuracy of the models, using the ground-truth SPAD data, while  $R^2$  evaluated the performance of the models based on k-fold cross-validation technique (Angel & McCabe, 2022). The segmentation methods were labelled as MLP-FR, SVM-FR, RF-FR, ExG-FR and ExGR-FR for field images and MLP-GR, SVM-GR, RF-GR, ExG-GR and ExGR-GR for glasshouse data. Normalized colour vegetation indices (CVIs), which are commonly used in estimating plant chlorophyll and nitrogen, were extracted from the segmented images and correlated (Pearson correlation) with the SPAD values. The dataset included 180 segmented cowpea images and 240 segmented wheat images, which were analyzed separately. Each dataset was divided into three portions in a 70:15:15 ratio for model training, testing and validation. The training data were used to fit the model, while the validation dataset reduced overfitting during the fine-tuning of the hyperparameters (Shi et al., 2021). Finally, the testing dataset provided an unbiased evaluation of the final model. The top five indices of each segmentation method were used as input features to train the respective regression models.

## 2.4 Results

For pixel-wise plant segmentation, plant colour features were extracted and used as inputs for training a multilayer perceptron neural network. The features belonged either to the AF (all features) dataset, which contains all the transformed colour features, or the SF (selected features) dataset, which contains only the colour features after feature selection.

### 2.4.1 Performance of the ML Models Based on Different Number of Training Features

Figure 2.5 depicts a whisker plot summarizing the accuracy scores of the models. Table 2-2 also summarizes the classification scores of all three models. With the MLP model, the SF datasets outperformed the AF dataset (99.321% vs. 98.464% accuracy). A similar trend was observed in the RF model, with a 0.343% accuracy difference between the SF and AF datasets. However, in the SVM model, the SF and AF datasets were statistically non-significant. As anticipated, all the models trained with the SF dataset had faster processing speed than those trained with the AF dataset. During training, the processing speed for the SVM model significantly decreased from 560.560 s in the AF dataset to 321.502 s in the SF dataset. Similarly, during testing (Supplementary information Table B-2), the SF models had higher processing speed (2.450 s, 44.81 s and 86.163 s for the MLP, RF and SVM models, respectively) than the AF-dataset-trained models (5.150 s, 81.225 s and 132.158 s for the MLP, RF and SVM models, respectively). This shows that the selected features not only improved the performance of the classification models but also reduced the processing time by using important and informative features for training the models. For this reason, all three models were trained with the SF datasets for segmentation.

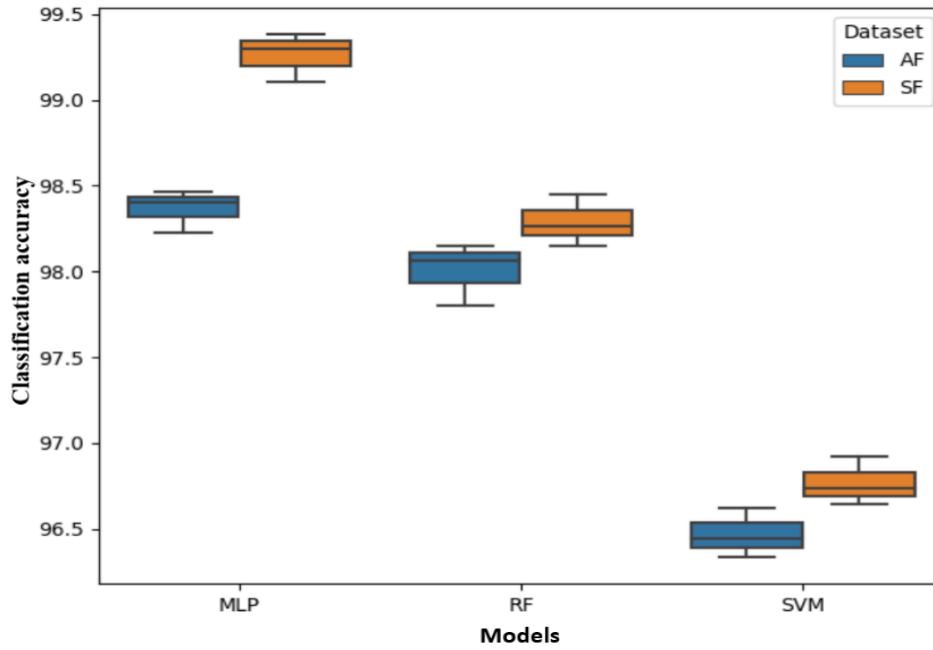


**Figure 2-5** Classification accuracy scores for the three models (multilayer perceptron (MLP), random forest (RF) and support vector machines (SVMs). This accuracy score analysis was based on the type of dataset used for the training the model. The datasets include; Selected Features dataset (SF) and All Features (AF) dataset. The SF dataset was obtained after feature selection using feature importance based selection technique while the AF dataset comprise all the extracted features obtained earlier.

**Table 2-2** Classification model performance assessment. MLP, RF and SVM are the multilayer perceptron, random forest and support vector machine segmentation models, respectively.

| Model | All Features Dataset (AL) |                     |                     |                        | Selected Features Dataset (SF) |                     |                     |                        |
|-------|---------------------------|---------------------|---------------------|------------------------|--------------------------------|---------------------|---------------------|------------------------|
|       | Accuracy                  | F-Score             | Recall              | Computational Time (s) | Accuracy                       | F-Score             | Recall              | Computational Time (s) |
| MLP   | 98.204 <sup>a</sup>       | 96.142 <sup>a</sup> | 98.144 <sup>a</sup> | 175.098 <sup>a</sup>   | 99.382 <sup>a</sup>            | 97.423 <sup>a</sup> | 98.321 <sup>a</sup> | 113.521 <sup>a</sup>   |
| RF    | 97.811 <sup>a</sup>       | 95.331 <sup>b</sup> | 97.421 <sup>a</sup> | 191.450 <sup>b</sup>   | 98.354 <sup>ab</sup>           | 96.632 <sup>a</sup> | 97.931 <sup>a</sup> | 128.455 <sup>b</sup>   |
| SVM   | 96.134 <sup>b</sup>       | 92.402 <sup>c</sup> | 95.651 <sup>b</sup> | 560.560 <sup>c</sup>   | 96.624 <sup>b</sup>            | 92.822 <sup>b</sup> | 96.514 <sup>b</sup> | 321.502 <sup>c</sup>   |

Superscripts <sup>a</sup>, <sup>b</sup> and <sup>c</sup> represent significant differences with  $p < 0.05$  within the columns.

The same letters mean no significant difference, and different letters mean a significant difference exists.

### 2.4.2 Segmentation Accuracy Analysis

Segmentation accuracy was evaluated using three quality factors, namely Qseg, Sr and an error factor (Es), as explained in Section 2.3.5. Figure 2.6 shows examples of segmented wheat and cowpea, using the MLP and ExG methods. By visual inspection, the proposed method produced high-quality images as compared to the traditional ExG segmentation method. Moreover, as shown in Figure 2.7, the ML-based methods (MLP, SVM and RF) outperformed the colour-index methods (ExG and ExGR). The MLP had the highest average Qseg (0.916) and Sr (0.956), with the lowest standard deviations (0.012 and 0.040, respectively), indicating less variability in the Qseg and Sr scores. The RF model followed closely, achieving 0.874 (Qseg) and 0.932 (Sr), with standard deviations of 0.046 and 0.044, respectively. ExG had the lowest Qseg and Sr values (0.785 and 0.816), along with the highest Es (0.220) indicating more pixel miss-classification, respectively. Again, MLP had the lowest classification error (0.058), followed by RF and SVM (0.074 and 0.170, respectively). ExG performed poorly, with the highest error rate of 0.455.

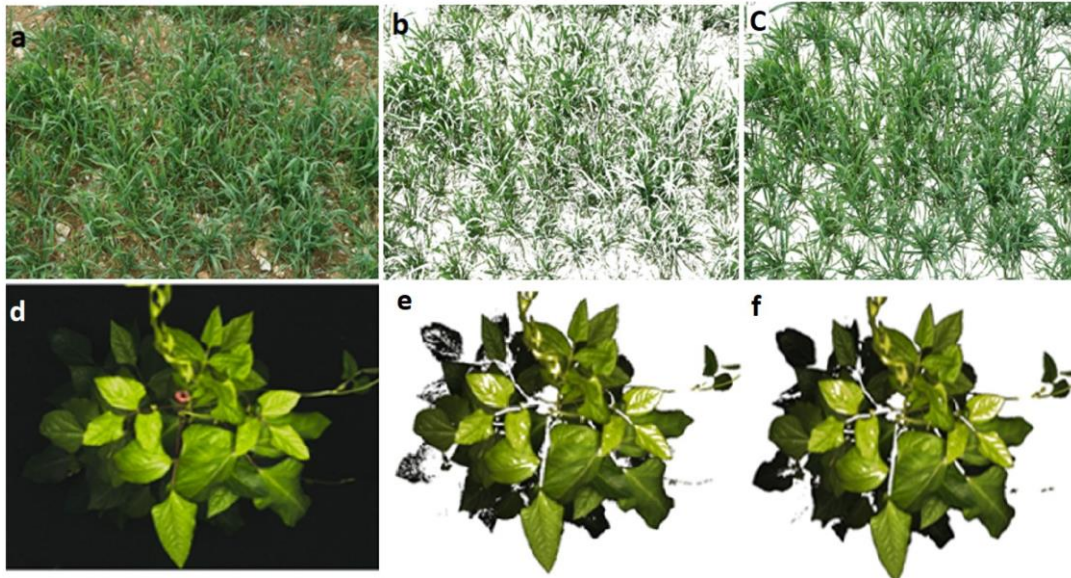


**Figure 2-6** Examples of glasshouse and field-segmented plants using the proposed method and selected segmentation methods. (a) Original wheat image, (b) ExG wheat segmented image, (c) proposed method (MLP) segmented image, (d) original cowpea image, (e) ExG cowpea and (f) MLP segmented image. The machine learning based methods including the proposed MLP were trained with the SF datasets whiles for the ExG and ExGR techniques, established colour indices together with Otsu thresholding were used.

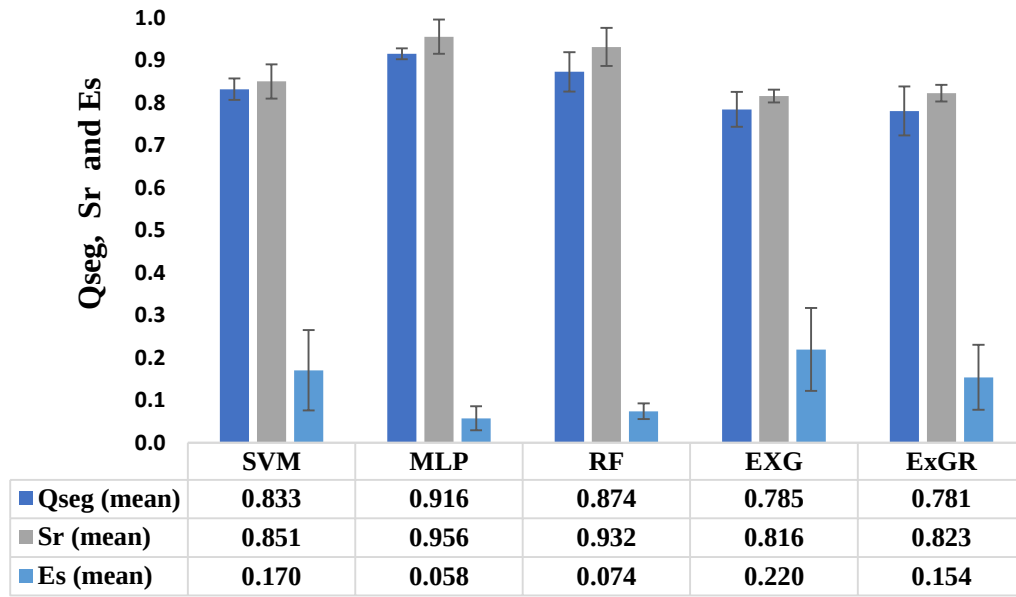


**Figure 2-7** Mean segmentation accuracy rate comparison for quality assessment of five different segmentation methods for glasshouse-based images; Qseg measures the segmentation consistency on a pixel-by-pixel basis, Sr measures the consistency of plant pixels between image regions and Es measures the rate of pixel misclassification. These were applied on the five segmentation methods: multilayer perceptron (MLP), support vector machines (SVMs), random forest (RF), excess green (ExG) and excess green–red (ExGR).

The results presented in Figure 2.8 show that the colour-index-based methods (ExG and ExGR) had the lowest Qseg values (0.725 and 0.693, respectively), with high standard deviations (0.082 and 0.112, respectively). On the other hand, the MLP method had the best performance, earning the highest Qseg rating (0.880), with a low standard deviation (0.055). This indicates that the high performance of the MLP model, producing better pixel classification compared to the other techniques, with an error rate of 0.127. In both environments, i.e., the glasshouse and the field, the ML-based methods outperformed the colour-based methods, and this is consistent with the finding of Adams et al (2020), who suggested that ML-based techniques are an improvement over traditional segmentation methods.

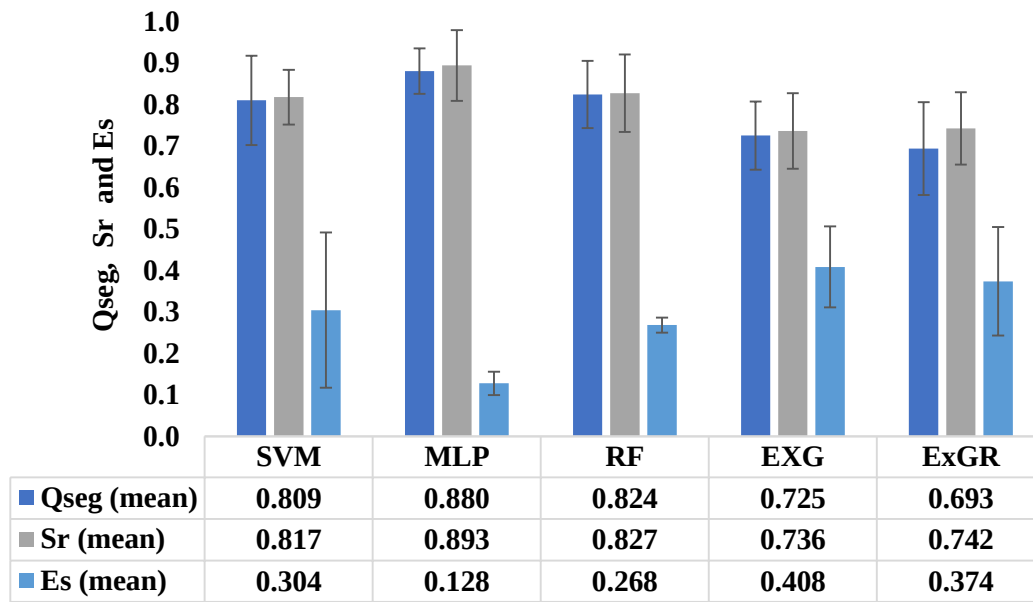


**Figure 2-8** Comparison of segmentation accuracy rate (Qseg-, Sr, Es) for quality assessment of five different segmentation methods; multilayer perceptron (MLP), support vector machines (SVMs), random forest (RF), excess green (ExG) and excess green–red (ExGR) for field-based images. Qseg measures the segmentation consistency on a pixel-by-pixel basis, *Sr* measures the consistency of plant pixels between image regions and *Es* measures the rate of pixel misclassification. These were applied to the five segmentation methods.

## 2.4.3 Performance of SPAD prediction regression models

The root mean square error (RMSE), the adjusted determination coefficient ( $R^2$ ) and the mean absolute error (MAE) were the metrics used to comparatively assess the performance of the SPAD regression models.

### 2.4.3.1 Cowpea SPAD Values Prediction

Table 2-3 shows the list of extracted CVIs with their corresponding correlation coefficients ( $r$ ) for cowpea. The quality of the segmented images influenced the correlations of the measured SPAD values and the extracted CVIs; for example, poorly segmented images with high pixel misclassification and noise led to lower correlations with the measured SPAD values and vice versa. The CVIs extracted using the MLP segmentation method had high correlation coefficients, with DGCI having the highest (0.948) and EGI having the lowest (0.464) coefficient. This was followed by the RF segmentation method, with the highest correlation (0.850) recorded with the DGCI index

and lowest (0.295) recorded with the EGI color vegetation index. The ExG segmentation method had the lowest correlation coefficient (0.228) recorded for the blue CVI, while its highest correlation coefficient (0.742) was recorded with ExG CVI.

Table 2-4 shows the results of the regression models for predicting SPAD values for the glasshouse-based images. The MLP-GR model had the highest  $R^2$  values in both the training and testing phases, while the ExG-GR model had the lowest. The  $R^2$  values for both training and testing datasets were close, indicating that the models did not underfit or overfit. The RMSE has an inverse trend with the  $R^2$  metric (the lower the RMSE, the better). Figure 2.9 is a scatterplot of observed and predicted SPAD values from test data which shows that the MLP-GR model with a high  $R^2$  (0.904) had the best prediction performance, followed by the RF-GR model ( $R^2 = 0.837$ ), while the ExG-GR model had the lowest prediction power ( $R^2 = 0.701$ ). Overall, all the models performed well, with high prediction power ( $R^2 > 0.7$ ) in both the training and testing phases.

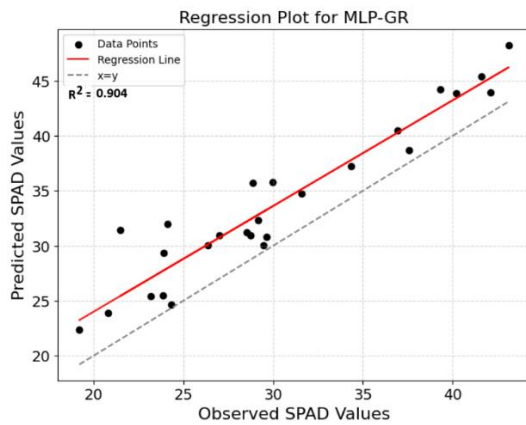
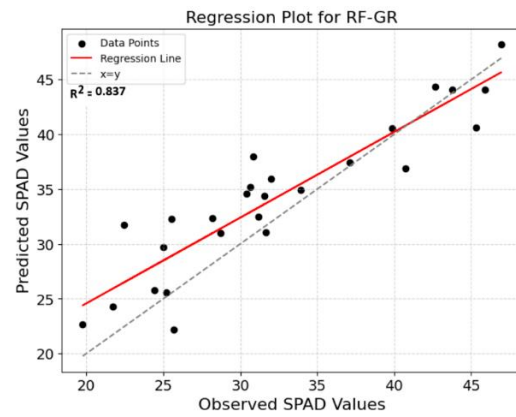
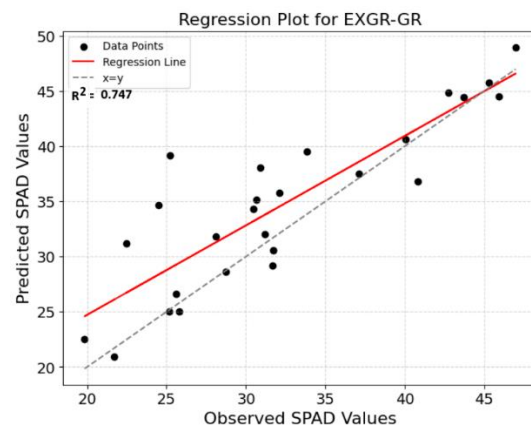
**Table 2-3** Correlation analysis of colour vegetation indices (CVIs) and SPAD readings for cowpea. R, G and B are the average red, green and blue values. The ExG, ExGR, MLP, SVM and RF are the excess green, excess green–red, multilayer perceptron, support vector machines and random forest segmentation methods, respectively.

| CVI                               | Mathematical Expression            | Correlation Coefficient (r) |        |        |        |        |
|-----------------------------------|------------------------------------|-----------------------------|--------|--------|--------|--------|
|                                   |                                    | Segmentation Method         |        |        |        |        |
|                                   |                                    | ExG                         | ExGR   | MLP    | SVM    | RF     |
| Green                             | -                                  | 0.420                       | 0.291  | 0.554  | 0.418  | 0.459  |
| Blue                              | -                                  | −0.228                      | −0.424 | −0.682 | −0.562 | −0.605 |
| ExG (Woebbecke et al., 1995)      | $2 \times G - R - B$               | 0.742                       | 0.763  | 0.893  | 0.824  | 0.838  |
| ExR (Meyer & Neto, 2008)          | $1.4 \times R - G$                 | 0.692                       | 0.822  | 0.885  | 0.745  | 0.820  |
| CIVE (Kataoka et al., 2003)       | $0.441R - 0.811G + 0.385B + 18.78$ | 0.670                       | 0.745  | 0.924  | 0.787  | 0.822  |
| ERI (Golzarian & Frick, 2011)     | $(R-G) \times (R-B)$               | −0.551                      | −0.568 | −0.661 | −0.517 | −0.642 |
| DGCI (Karcher & Richardson, 2003) | $(H-60)/60 + (1-S) + (1-B)]/3$     | 0.672                       | 0.759  | 0.948  | 0.821  | 0.850  |
| GR                                | G-R                                | −0.585                      | −0.721 | −0.842 | −0.766 | −0.671 |
| COM1 (Guijarro et al., 2015)      | ExG + CIVE + ExGR                  | 0.623                       | 0.748  | 0.901  | 0.792  | 0.782  |
| GBRG (Woebbecke et al., 1995)     | $(G-R)/(R-G)$                      | 0.689                       | 0.824  | 0.870  | 0.844  | 0.802  |
| EGI (Golzarian & Frick, 2011)     | $(G-R)/(G-B)$                      | 0.231                       | 0.425  | 0.464  | 0.382  | 0.295  |



**Table 2-4** Performance of the regression models during training and testing for glasshouse data. MLP-GR, RF-GR, SVM-FR, ExGR-GR and ExG-GR are the glasshouse-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green-red segmentation methods, respectively.

| Model   | Training Set   |       |       | Testing Set    |       |       |
|---------|----------------|-------|-------|----------------|-------|-------|
|         | R <sup>2</sup> | RMSE  | MAE   | R <sup>2</sup> | RMSE  | MAE   |
| MLP-GR  | 0.942          | 2.152 | 2.120 | 0.904          | 2.682 | 2.931 |
| RF-GR   | 0.881          | 3.885 | 2.581 | 0.837          | 3.958 | 3.780 |
| SVM-GR  | 0.864          | 4.255 | 3.601 | 0.826          | 3.620 | 3.881 |
| ExGR-GR | 0.790          | 5.102 | 4.512 | 0.747          | 5.790 | 4.910 |
| ExG-GR  | 0.772          | 6.881 | 5.155 | 0.701          | 7.151 | 6.851 |



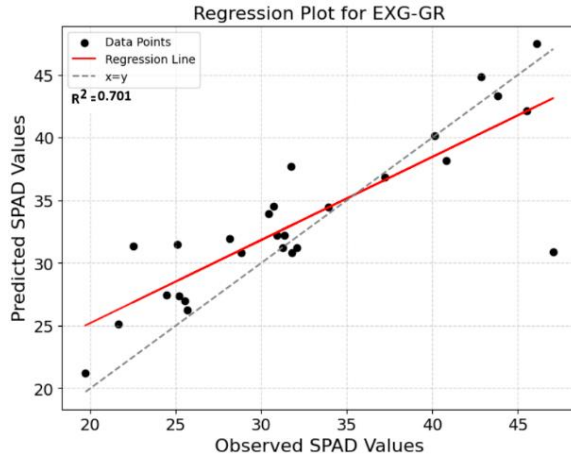


**Figure 2-9** Scatterplots of predicted and observed SPAD values for the glasshouse plants. MLP-GR, RF-GR, SVM-GR, ExGR-GR and ExG-GR represent the glasshouse-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green–red segmentation methods, respectively. All the regressions methods were performed using the XGBoost regression technique and evaluated using the  $R^2$ , MAE and RMSE metrics.

#### 2.4.3.2 Wheat SPAD Value Prediction

Table 2-5 shows the Pearson correlation coefficient between the selected CVIs and SPAD readings. The MLP segmented method shows a high correlation with the SPAD values for all the extracted CVIs, with the highest correlation observed for GBRG (0.904) and the lowest for EGI (0.412). The ExG segmentation method yielded relatively low correlations, with the highest correlation for ExG CVI (0.704) and the lowest for Green CVI (0.242), indicating poorly segmented images. Among the extracted CVIs, DGCi demonstrated a high correlation with all the segmentation methods, which is consistent with the finding of (Hassanijalilian et al., 2020b). Table 2-6 shows the performance of the regression models for the training and testing datasets. The MLP-FR model exhibited the best performance, achieving the highest  $R^2$  score for both training (0.893) and testing (0.821) datasets, with the lowest MAE scores (2.520 and 3.233, respectively). This indicates that the MLP method segmented the images well, with minimal errors. RF-FR and SVM-FR followed with  $R^2 = 0.815$  and  $0.791$ , respectively, for training and  $0.770$  and  $0.747$ , respectively, for testing. The ExG-FR model had relatively the lowest performance, with  $R^2 = 0.642$  and  $0.521$  for training and testing, respectively. This could be attributed to the poorly segmented images obtained using the ExG method, resulting in a large number of misclassified pixel features for training the regression model. Figure 2.10 shows the linear regression between the predicted and observed SPAD values for the

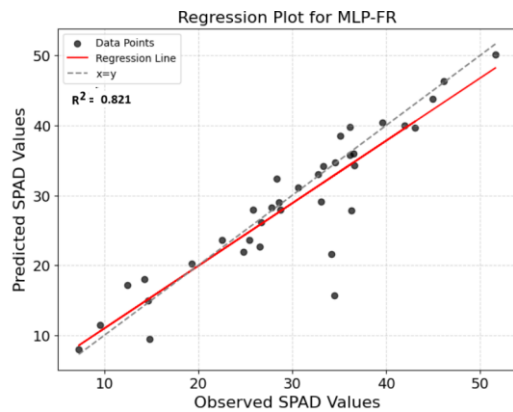
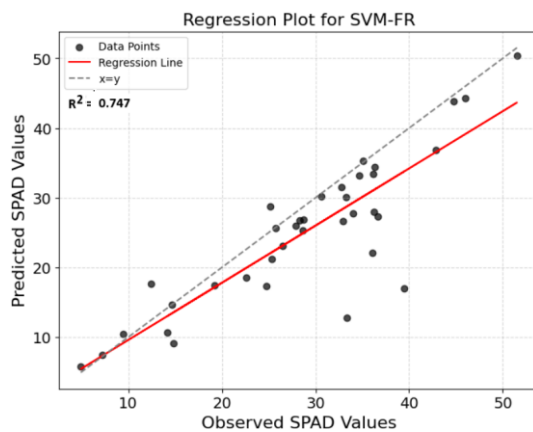
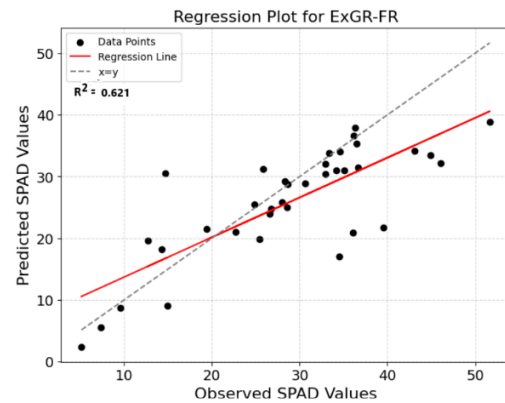
different segmentation methods using the wheat testing dataset. The MLP-FR model demonstrated better prediction accuracy ( $R^2 = 0.821$ ), outperforming all the other methods, while the ExG-FR model had the lowest prediction accuracy ( $R^2 = 0.546$ ).

**Table 2-5** Correlation analysis of colour vegetation indices (CVIs) and measured SPAD readings for wheat. R, G and B are the average red, green and blue values. MLP, RF, SVM, ExGR and ExG are the segmentation methods using multilayer perceptron, random forest, support vector machines, excess green and excess green–red segmentation methods, respectively.

| CVI                               | Mathematical Expression                 | Correlation Coefficient(r) |        |        |        |        |
|-----------------------------------|-----------------------------------------|----------------------------|--------|--------|--------|--------|
|                                   |                                         | Segmentation Method        |        |        |        |        |
|                                   |                                         | ExG                        | ExGR   | MLP    | SVM    | RF     |
| Green                             | -                                       | 0.242                      | 0.212  | 0.443  | 0.38   | 0.291  |
| Blue                              | -                                       | -0.377                     | -0.344 | -0.618 | -0.522 | -0.552 |
| ExG (Woebbecke et al., 1995)      | $2 * G - R - B$                         | 0.704                      | 0.726  | 0.927  | 0.841  | 0.877  |
| ExR (Meyer & Neto, 2008)          | $1.4 * R - G$                           | 0.675                      | 0.722  | 0.845  | 0.751  | 0.803  |
| CIVE (Kataoka et al., 2003)       | $0.441R - 0.811G + 0.385B + 18.78$      | 0.601                      | -0.652 | -0.842 | -0.773 | -0.782 |
| ERI (Golzarian & Frick, 2011)     | $(R - G) * (R - B)$                     | -0.407                     | -0.476 | -0.681 | -0.572 | -0.644 |
| DGCI (Karcher & Richardson, 2003) | $(H - 60) / 60 + (1 - S) + (1 - B) / 3$ | 0.623                      | 0.690  | 0.881  | 0.791  | 0.747  |
| GR                                | $G - R$                                 | -0.551                     | -0.606 | -0.740 | -0.657 | -0.671 |
| COM1 (Guijarro et al., 2015)      | $ExG + CIVE + ExGR$                     | 0.532                      | 0.682  | 0.813  | 0.722  | 0.78   |
| GBRG (Woebbecke et al., 1995)     | $(G - R) / (R - G)$                     | 0.690                      | 0.743  | 0.904  | 0.781  | 0.811  |
| EGI (Golzarian & Frick, 2011)     | $(G - R) / (G - B)$                     | 0.111                      | 0.251  | 0.412  | 0.322  | 0.352  |

**Table 2-6** Performance evaluation of the regression models. MLP-FR, RF-FR, SVM-FR, ExGR-FR and ExG-FR represent the field-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green–red methods, respectively.

| Model   | Training Set   |      |      | Testing Set    |        |        |
|---------|----------------|------|------|----------------|--------|--------|
|         | R <sup>2</sup> | RMSE | MAE  | R <sup>2</sup> | RMSE   | MAE    |
| MLP-FR  | 0.893          | 3.25 | 2.52 | 0.821          | 3.680  | 3.233  |
| RF-FR   | 0.815          | 4.05 | 2.78 | 0.775          | 3.785  | 3.781  |
| SVM-FR  | 0.791          | 5.10 | 3.90 | 0.747          | 7.903  | 4.252  |
| ExGR-FR | 0.670          | 7.88 | 5.12 | 0.621          | 8.202  | 6.102  |
| ExG-FR  | 0.642          | 9.85 | 7.55 | 0.546          | 13.155 | 11.851 |



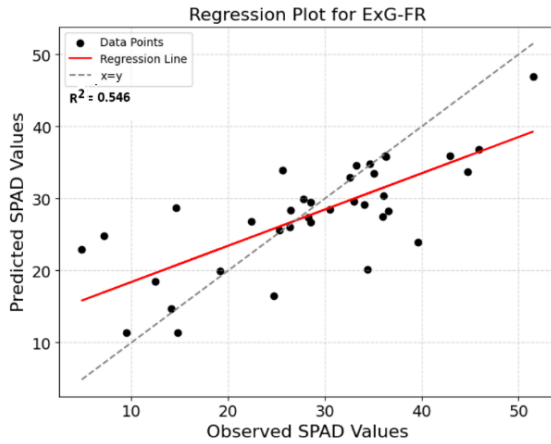


**Figure 2-10** Scatterplots of predicted and observed SPAD values for the field plants. MLP-FR, RF-FR, SVM-FR, ExGR-FR and ExG-FR are the field-based regression models for the multilayer perceptron, random forest, support vector machine, excess green and excess green and excess green and excess green–red segmentation methods, respectively. All the regressions methods were performed using the XGBoost regression technique and evaluated using the  $R^2$ , MAE and RMSE metrics.

## 2.5 Discussion

Images of plants acquired from environments with varying conditions present daunting tasks for accurate segmentation of the foreground and background objects. This is due to conditions such as illumination variability and different colouration of plant leaves resulting from factors such as nutrient deficiencies, diseases and plant senescence. These challenges are particularly predominant in field-based images, where there are other complex situations, such as cluttered backgrounds and leaf occlusion, resulting in the underestimation of vegetation pixels within the images. While colour-index-based segmentation is a widely used low-level pixel feature for images with uniform background and contrast, it is not effective for images with complex backgrounds or canopies with non-uniform colour distribution. In this study, the colour-index methods failed to adequately segment vegetation, resulting in multiple fragments of the plants parts in the image due to colour disparities. However, ExGR demonstrated good segmentation quality for images taken from the glasshouse and field compared to ExG. This is in agreement with Guijarro et al. (2011) and Meyer & Neto(2008), who found that ExGR and combinations of other colour indices produce improved segmentation quality compared to the ExG method.

A single-layer MLP network that takes in annotated and colour-transformed pixels and outputs binary classified pixels was proposed. The annotation covered all plant growth stages to ensure the model effectively learns different scenarios of plant colour characteristics, such as yellow-to-brown colour due to nitrogen deficiency and deep green colour from healthy plant leaves. The MLP model with a high training and testing accuracy produced quality segmented images with reduced noise levels. In a similar study, Bakhshipour et al.,(2017) used an artificial neural network (ANN) trained with wavelet texture features to achieve a 96% accuracy in discriminating weeds from sugar beets in the field. This supports the idea that neural networks such as MLP and ANN can successfully segment plant images when trained with the right phenotypic traits. The results show that the proposed MLP segmentation model is robust, achieving a high segmentation accuracy when used on plants from different environments and with complex leaf architecture and occlusion.

Furthermore, the feature selection technique employed improved the robustness of the model by reducing the processing time and ensuring that the model was trained on only sensitive and non-redundant features. Using the DT-RFE method was a helpful part of the feature selection process, which potentially reduced multicollinearity by eliminating redundant features, potentially reducing multicollinearity. Rahaman et al., (2019) suggests combining data mining and machine learning techniques improved their feature ranking which is generally useful to make plant phenotype data analysis more effective and robust throughout the classification. While DT-RFE can help in identifying and removing some correlated features, it doesn't explicitly measure or correct for multicollinearity. Techniques such as Variance Inflation Factor (VIF) would be able to address the multicollinearity issue if this is a primary concern in further analysis.

With a reduced number of training features, the MLP processing time was reduced by 62.8%, whilst the SVM and RF processing time were reduced by 49.1% and 57.4%, respectively. This is important during real-time processing, where high computational speed is required for the batch processing of images. Although the proposed method by Sadeghi-Tehran et al( 2017) successfully segmented plant vegetation from cluttered backgrounds, the large-features dataset used for model training presented a computational challenge. Unlike the proposed method, a high-powered computer memory is required to implement such segmentation techniques for the automated processing of images.

Furthermore, although deep learning networks such as convolutional neural networks (CNNs) have been successful in plant segmentation (Sulistyo et al., 2017; Xiong et al., 2017; Bakhshipour et al., 2017), they require large datasets to train and validate the model to ensure better performance. However, in this study, only a small dataset (less than 10% of the total dataset) was used to train the neural network for segmentation. That is a small size datasets (smaller proportion of the whole datasets) was used. Moreover, due to the fast nature and less training data requirement of the proposed method, it can easily be deployed on edge devices such as mobile phones, Raspberry Pi and other devices just like other complex deep learning models. In reality, MLPs are frequently easier to deploy on resource-constrained edge devices due to their simpler design, which generally requires less computing resources and memory than more complex models such as convolutional neural networks (CNNs) or transformers.

Generally, ML-based methods have been successful in segmenting plant vegetation from complex environments under varying illumination conditions compared to colour-index-based methods. The additional pixel information resulting from the colour space transformations enhances the discriminatory power of the segmentation model. Supervised ML models demonstrated improved segmentation quality, which is consistent with the finding of McCauley et al (2011), who suggested that ML techniques that learn colour features from multiple colour orientations improve the robustness of the segmentation model. The RF model exhibited good segmentation quality slightly lower than the MLP but better than the SVM model for both glasshouse and field based plants. In effect, the foreground consistency of the RF model ( $S_r = 0.827$ ) is similar to SVM model but lower than MLP model. This shows that with good optimization (using fine tuning of parameters and hyperparameters), the RF model can exhibit good performance and can be used in place of the MLP model. While the SVM model performed well, its high computational speed and memory requirements make it less feasible for high-throughput plant phenotyping.

After segmenting the images, machine learning regression models were developed based on the different segmentation methods to predict plant SPAD values. Generally, the method with good segmentation quality produced high prediction accuracies. In this case, the MLP segmentation method outperformed the others, achieving a high  $R^2$  (0.904, 0.821) and low RMSE (2.862, 3.680) for both greenhouse and field plants. This is in line



with the Qui et al. (2021) study, which asserts that neural networks have high potential in estimating the nitrogen nutrition index (NNI) in rice compared to other statistical and machine learning algorithms. They developed six machine learning models, namely SVM, ANN, RF, k-nearest neighbour (KNN), partial least square regression (PLSR) and adaptive boosting, based on rice RGB images from an unmanned aerial vehicle (UAV). Their results show that ANN can estimate the NNI at the early jointing and filling stages of rice with  $R^2 = 0.72$  and  $0.80$ , respectively. While Hassanijalilian et al(2020a) achieved a similar performance in predicting soybean chlorophyll by using an SVM model ( $R^2 = 0.86$  and  $RMSE = 3.20$ ), their method was based on a few randomly selected leaves, which may not represent the actual scenario of chlorophyll dynamics in the field. Moreover, their object-recognition method (segmentation) is only feasible for small plant datasets.

Overall, the MLP method achieved the best performance in all the conditions tested in this study. It produced the highest segmentation quality with lower variations. The MLP method is versatile and can be used to segment not just green vegetation in the glasshouse but also field-based plant images with variable colour information resulting from, for example, a nutrient deficiency or senescence.

## 2.6 Conclusions

This study presented a novel method for automated non-invasive segmentation of greenhouse and field plant images. Relevant information was extracted from sensitive features and used to train a one-layer MLP model that accurately distinguishes plant vegetation pixels from non-plant pixels. The model is capable of segmenting plants from diverse environments, ranging from glasshouses with simple backgrounds to field plants with cluttered backgrounds and varying illumination conditions. The selected features were also used to train classical ML models, SVM and RF and their performances were compared to the proposed method. Although the classical ML models performed well, the MLP outperformed them in segmenting both glasshouse- and field-based images with high precision. The robustness of all segmentation methods was tested on plants from the glasshouse (cowpea) and the field (wheat) under different illumination conditions and the proposed method produced the best performance in both cases. This method is not limited to segmenting green vegetation, but also plants with diseases and nutrient deficiencies,

making it suitable for high throughput segmentation of plant images in both glasshouse and field environments.

## Acknowledgements

This research was supported by the OCP S.A. under the University of Mohammed VI Polytechnic, Rothamsted Research and Cranfield University project (FPO4). Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) through the Delivering Sustainable Wheat program (BB/X011003/1).

## 2.7 References

- Adams, J., Qiu, Y., Xu, Y., & Schnable, J. C. (2020). Plant segmentation by supervised machine learning methods. *Plant Phenome Journal*, 3(1), 1–11. <https://doi.org/10.1002/ppj2.20001>
- Aich, S., & Stavness, I. (2017). Leaf counting with deep convolutional and deconvolutional networks. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, 2018-Janua*, 2080–2089. <https://doi.org/10.1109/ICCVW.2017.244>
- Akbarzadeh, S., Paap, A., Ahderom, S., Apopei, B., & Alameh, K. (2018). Plant discrimination by Support Vector Machine classifier based on spectral reflectance. *Computers and Electronics in Agriculture*, 148(August 2017), 250–258. <https://doi.org/10.1016/j.compag.2018.03.026>
- Angel, Y., & McCabe, M. F. (2022). Machine Learning Strategies for the Retrieval of Leaf-Chlorophyll Dynamics: Model Choice, Sequential Versus Retraining Learning, and Hyperspectral Predictors. *Frontiers in Plant Science*, 13(March). <https://doi.org/10.3389/fpls.2022.722442>
- Arulmozhi, E. (2021). *Machine Learning-Based Microclimate Model for Indoor Air*.
- Aslam, M., Lee, J. M., & Hong, S. (2020). A multi-layer perceptron based deep learning model to quantify the energy potentials of a thin film a-Si PV system. *Energy Reports*, 6, 1331–1336. <https://doi.org/10.1016/j.egyr.2020.11.025>
- Atkinson, P. M., & Tatnall, A. R. L. (1997). Introduction neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), 699–709. <https://doi.org/10.1080/014311697218700>
- Bakhshipour, A., Jafari, A., Nassiri, S. M., & Zare, D. (2017). Weed segmentation using

- texture features extracted from wavelet sub-images. *Biosystems Engineering*, 157, 1–12. <https://doi.org/10.1016/j.biosystemseng.2017.02.002>
- Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture*, 153(July), 46–53. <https://doi.org/10.1016/j.compag.2018.08.013>
- Barrett-lennard, E. G., Robson, A. D., & Greenway, H. (1982). Effect of phosphorus deficiency and water deficit on phosphatase activities from wheat leaves. *Journal of Experimental Botany*, 33(4), 682–693. <https://doi.org/10.1093/jxb/33.4.682>
- Chen, H., Dou, Q., Yu, L., Qin, J., & Heng, P. A. (2018). VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, 170(April 2017), 446–455. <https://doi.org/10.1016/j.neuroimage.2017.04.041>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-Aug*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chopin, J., Kumar, P., & Miklavcic, S. J. (2018). Land-based crop phenotyping by image analysis: Consistent canopy characterization from inconsistent field illumination. *Plant Methods*, 14(1), 1–15. <https://doi.org/10.1186/s13007-018-0308-5>
- Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics*, 19(Suppl 1), 1–6. <https://doi.org/10.1186/s12863-018-0633-8>
- Desai, M., & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical EHealth*, 4(2021), 1–11. <https://doi.org/10.1016/j.ceh.2020.11.002>
- Dutta, A., & Zisserman, A. (2019). The VIA annotation software for images, audio and video. *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2276–2279. <https://doi.org/10.1145/3343031.3350535>
- Ganesan, P., Sajiv, G., & Megalan Leo, L. (2017). CIELuv color space for identification and segmentation of disease affected plant leaves using fuzzy based approach. *ICONSTEM 2017 - Proceedings: 3rd IEEE International Conference on Science Technology, Engineering and Management, 2018-Janua*, 889–894. <https://doi.org/10.1109/ICONSTEM.2017.8261330>
- Golzarian, M. R., & Frick, R. A. (2011). Classification of images of wheat, ryegrass and brome grass species at early growth stages using principal component analysis. *Plant Methods*, 7(1), 1–11. <https://doi.org/10.1186/1746-4811-7-28>

- Guijarro, M., Pajares, G., Riomoros, I., Herrera, P. J., Burgos-Artizzu, X. P., & Ribeiro, A. (2011). Automatic segmentation of relevant textures in agricultural images. *Computers and Electronics in Agriculture*, 75(1), 75–83. <https://doi.org/10.1016/j.compag.2010.09.013>
- Guijarro, M., Riomoros, I., Pajares, G., & Zitinski, P. (2015). Discrete wavelets transform for improving greenness image segmentation in agricultural images. *Computers and Electronics in Agriculture*, 118, 396–407. <https://doi.org/10.1016/j.compag.2015.09.011>
- Guo, W., Rage, U. K., & Ninomiya, S. (2013). Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model. *Computers and Electronics in Agriculture*, 96, 58–66. <https://doi.org/10.1016/j.compag.2013.04.010>
- Hamuda, E., Glavin, M., & Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, 125, 184–199. <https://doi.org/10.1016/j.compag.2016.04.024>
- Hassanijalilian, O., Igathinathane, C., Doetkott, C., Bajwa, S., Nowatzki, J., & Haji Esmaeili, S. A. (2020). Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Computers and Electronics in Agriculture*, 174(April), 105433. <https://doi.org/10.1016/j.compag.2020.105433>
- Hesami, M., Naderi, R., & Tohidfar, M. (2019). Modeling and optimizing In vitro sterilization of chrysanthemum via multilayer perceptron-non-dominated sorting genetic algorithm-II (MLP-NSGAI). In *Frontiers in Plant Science* (Vol. 10). <https://doi.org/10.3389/fpls.2019.00282>
- Hindman, T., & Meyer, G. E. (1998). Fuzzy Logic Detection Parameters for Plant Species Identification. *Systems Engineering*, 3543(1986), 327–335.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Iqbal, M. A., & Talukder, K. H. (2020). Detection of Potato Disease Using Image Segmentation and Machine Learning. *2020 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2020*, 43–47. <https://doi.org/10.1109/WiSPNET48689.2020.9198563>
- Jiang, Y., & Li, C. (2020). CHAPTE- ALL- Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics*, 2020. <https://doi.org/10.34133/2020/4152816>
- Joalland, S., Screpanti, C., Liebisch, F., Varella, H. V., Gaume, A., & Walter, A. (2017).

- Comparison of visible imaging, thermography and spectrometry methods to evaluate the effect of *Heterodera schachtii* inoculation on sugar beets. *Plant Methods*, 13(1), 1–14. <https://doi.org/10.1186/s13007-017-0223-1>
- Karcher, D. E., & Richardson, M. D. (2003). Quantifying turfgrass color using digital image analysis. In *Crop Science* (Vol. 43, Issue 3, pp. 943–951). <https://doi.org/10.2135/cropsci2003.9430>
- Kataoka, T., Kaneko, T., Okamoto, H., & Hata, S. (2003). Crop growth estimation system using machine vision. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, 2(Aim), 1079–1083. <https://doi.org/10.1109/AIM.2003.1225492>
- Kirk, K., Andersen, H. J., Thomsen, A. G., Jørgensen, J. R., & Jørgensen, R. N. (2009). Estimation of leaf area index in cereal crops using red-green images. *Biosystems Engineering*, 104(3), 308–317. <https://doi.org/10.1016/j.biosystemseng.2009.07.001>
- Kumar, P., & Miklavcic, S. J. (2018). Analytical study of colour spaces for plant pixel detection. *Journal of Imaging*, 4(2), 1–12. <https://doi.org/10.3390/jimaging4020042>
- Lee, U., Chang, S., Putra, G. A., Kim, H., & Kim, D. H. (2018). An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image analysis. *PLoS ONE*, 13(4), 1–17. <https://doi.org/10.1371/journal.pone.0196615>
- Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors (Switzerland)*, 14(11), 20078–20111. <https://doi.org/10.3390/s141120078>
- Lo, C., & Marculescu, R. (2019). MetaNN: Accurate classification of host phenotypes from metagenomic data using neural networks. *BMC Bioinformatics*, 20(Suppl 12), 1–14. <https://doi.org/10.1186/s12859-019-2833-2>
- Masters-Clark, E., Shone, E., Paradelo, M., Hirsch, P. R., Clark, I. M., Otten, W., Brennan, F., & Mauchline, T. H. (2020). Development of a defined compost system for the study of plant-microbe interactions. *Scientific Reports*, 10(1), 1–9. <https://doi.org/10.1038/s41598-020-64249-0>
- McCauley, A., Jones, C., & Jacobsen, J. (2011). Plant Nutrient Functions and Deficiency and Toxicity Symptoms. *Nutrient Management Module*, 9(9), 1–16.
- Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, 63(2), 282–293. <https://doi.org/10.1016/j.compag.2008.03.009>
- Meyer, G. E., Neto, J. C., Jones, D. D., & Hindman, T. W. (2004). Intensified fuzzy

- clusters for classifying plant, soil, and residue regions of interest from color images. *Computers and Electronics in Agriculture*, 42(3), 161–180. <https://doi.org/10.1016/j.compag.2003.08.002>
- Morris, D. (2018). A pyramid CNN for dense-leaves segmentation. *Proceedings - 2018 15th Conference on Computer and Robot Vision, CRV 2018*, 238–245. <https://doi.org/10.1109/CRV.2018.00041>
- Naik, D. L., & kiran, R. (2021). A novel sensitivity-based method for feature selection. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00515-w>
- Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., Singh, A., Singh, A. K., & Ganapathysubramanian, B. (2017). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods*, 13(1), 1–12. <https://doi.org/10.1186/s13007-017-0173-7>
- Narisetti, N., Henke, M., Neumann, K., Stolzenburg, F., Altmann, T., & Gladilin, E. (2022). Deep Learning Based Greenhouse Image Segmentation and Shoot Phenotyping (DeepShoot). *Frontiers in Plant Science*, 13(July), 1–16. <https://doi.org/10.3389/fpls.2022.906410>
- Neto, N. F. S., Stefenon, S. F., Meyer, L. H., Bruns, R., Nied, A., Seman, L. O., Gonzalez, G. V., Leithardt, V. R. Q., & Yow, K. C. (2021). A study of multilayer perceptron networks applied to classification of ceramic insulators using ultrasound. *Applied Sciences (Switzerland)*, 11(4), 1–19. <https://doi.org/10.3390/app11041592>
- Nguyen, G. N., Panozzo, J., Spangenberg, G., & Kant, S. (2016). Phenotyping approaches to evaluate nitrogen-use efficiency related traits of diverse wheat varieties under field conditions. *Crop and Pasture Science*, 67(11), 1139–1148. <https://doi.org/10.1071/CP16173>
- Niza El Aisnadaa, A., van der Kruit, K. B. R., Draijer, K. M., Lopez-Zorrilla, J., Miyauchi, M., Yamaguchi, A., & Artrith, N. (2024). A cost-effective strategy of enhancing machine learning potentials by transfer learning from a multicomponent dataset on ænet-PyTorch. *ArXiv E-Prints*, arXiv:2408.12939. <https://doi.org/10.48550/arXiv.2408.12939>
- Oh, S., Lee, D. Y., Gongora-Canul, C., Ashapure, A., Carpenter, J., Cruz, A. P., Fernandez-Campos, M., Lane, B. Z., Telenko, D. E. P., Jung, J., & Cruz, C. D. (2021). Tar spot disease quantification using unmanned aircraft systems (Uas) data. In *Remote Sensing* (Vol. 13, Issue 13). <https://doi.org/10.3390/rs13132567>
- Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claims using telematics data—XGboost versus logistic regression. *Risks*, 7(2). <https://doi.org/10.3390/risks7020070>

- Prasetyo, E., Adityo, R. D., Suciati, N., & Fatichah, C. (2017). CHAPTER 2: Mango leaf image segmentation on HSV and YCbCr color spaces using Otsu thresholding. *Proceeding - 2017 3rd International Conference on Science and Technology-Computer, ICST 2017*, 99–103. <https://doi.org/10.1109/ICSTC.2017.8011860>
- Rahaman, M. M., Ahsan, M. A., & Chen, M. (2019). Data-mining Techniques for Image-based Plant Phenotypic Traits Identification and Classification. *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-55609-6>
- Sadeghi-Tehran, P., Virlet, N., Sabermanesh, K., & Hawkesford, M. J. (2017). Chapter 2. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. *Plant Methods*, 13(1), 1–17. <https://doi.org/10.1186/s13007-017-0253-8>
- Sari, T., Kefali, A., & Bahi, H. (2012). An MLP for binarizing images of old manuscripts. *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, 247–251. <https://doi.org/10.1109/ICFHR.2012.176>
- Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., & Sun, Q. (2021). Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Computers and Electronics in Agriculture*, 180(October 2020), 105860. <https://doi.org/10.1016/j.compag.2020.105860>
- Sugiura, R., Tsuda, S., Tsuji, H., & Murakami, N. (2018). *by UAV Imagery Written for presentation at the 2018 ASABE Annual International Meeting Sponsored by ASABE*. 2–6.
- Sulistyo, S. B., Woo, W. L., & Dlay, S. S. (2017). Regularized Neural Networks Fusion and Genetic Algorithm Based On-Field Nitrogen Status Estimation of Wheat Plants. *IEEE Transactions on Industrial Informatics*, 13(1), 103–114. <https://doi.org/10.1109/TII.2016.2628439>
- Sunoj, S., McRoberts, K. C., Benson, M., & Ketterings, Q. M. (2021). Digital image analysis estimates of biomass, carbon, and nitrogen uptake of winter cereal cover crops. *Computers and Electronics in Agriculture*, 184(February), 106093. <https://doi.org/10.1016/j.compag.2021.106093>
- Thomas, P., & Suhner, M. C. (2015). A New Multilayer Perceptron Pruning Algorithm for Classification and Regression Applications. *Neural Processing Letters*, 42(2), 437–458. <https://doi.org/10.1007/s11063-014-9366-5>
- Williamson, S. J., Cummins, H. Z., & Kidder, J. N. (1984). Light and Color in Nature and Art. *American Journal of Physics*, 52(2), 186–186. <https://doi.org/10.1119/1.13928>
- Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1995). Color

- indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the American Society of Agricultural Engineers*, 38(1), 259–269. <https://doi.org/10.13031/2013.27838>
- Xiong, X., Yu, L., Yang, W., Liu, M., Jiang, N., Wu, D., Chen, G., Xiong, L., Liu, K., & Liu, Q. (2017). A high-throughput stereo-imaging system for quantifying rape leaf traits during the seedling stage. *Plant Methods*, 13(1), 1–17. <https://doi.org/10.1186/s13007-017-0157-7>
- Yu, K., Kirchgessner, N., Grieder, C., Walter, A., & Hund, A. (2017). An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping. *Plant Methods*, 13(1), 1–13. <https://doi.org/10.1186/s13007-017-0168-4>
- Zhang, J., Min, A., Steffenson, B. J., Su, W. H., Hirsch, C. D., Anderson, J., Wei, J., Ma, Q., & Yang, C. (2022). Wheat-Net: An Automatic Dense Wheat Spike Segmentation Method Based on an Optimized Hybrid Task Cascade Model. *Frontiers in Plant Science*, 13(February), 1–13. <https://doi.org/10.3389/fpls.2022.834938>
- Zhang, Q., Liu, P., Wang, X., Zhang, Y., Han, Y., & Yu, B. (2021). StackPDB: Predicting DNA-binding proteins based on XGB-RFE feature optimization and stacked ensemble classifier. *Applied Soft Computing*, 99, 106921. <https://doi.org/10.1016/j.asoc.2020.106921>
- Zheng, Y., Zhu, Q., Huang, M., Guo, Y., & Qin, J. (2017). Maize and weed classification using color indices with support vector data description in outdoor fields. *Computers and Electronics in Agriculture*, 141, 215–222. <https://doi.org/10.1016/j.compag.2017.07.028>
- Zulfikar, A. F., Supriyadi, D., Heryadi, Y., & Lukas. (2019). Comparison performance of decision tree classification model for spam filtering with or without the recursive feature elimination (RFE) approach. *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2019*, 6, 311–316. <https://doi.org/10.1109/ICITISEE48480.2019.9004001>



## **2.8 Additions to Published Article**

The following sections provide additional information on improvements made in the proposed segmentation algorithm using various machine-learning methods.

### **2.8.1 Challenges with the Proposed Segmentation Algorithm**

While the proposed segmentation method demonstrated commendable performance in various plant architecture and environmental conditions, it also revealed significant challenges. The model's reliance on colour information posed difficulties in segmenting certain scenarios. For instance, the model mistakenly identified some of the affected portions as background in plants with brown or black blotches (Figure 2.11) caused by prolonged stresses like nitrogen or drought. This was because the blotches share similar colour characteristics with the soil or other non-plant matter, leading the model to eliminate them. Additionally, extreme sunlight reflections on images adversely affected the model's performance. Despite the training features encompassing image patches from diverse light conditions, strong heterogeneous flare artefacts caused by extreme light reflections compromised the overall quality of the segmented images.

An improvement was made to address the challenges affecting the performance of the proposed segmentation algorithm. The improvement was made by incorporating additional plant-specific features and colour features. This modification was instrumental in training the model to better identify and segment plant images, thereby mitigating the issues encountered with the initial algorithm.

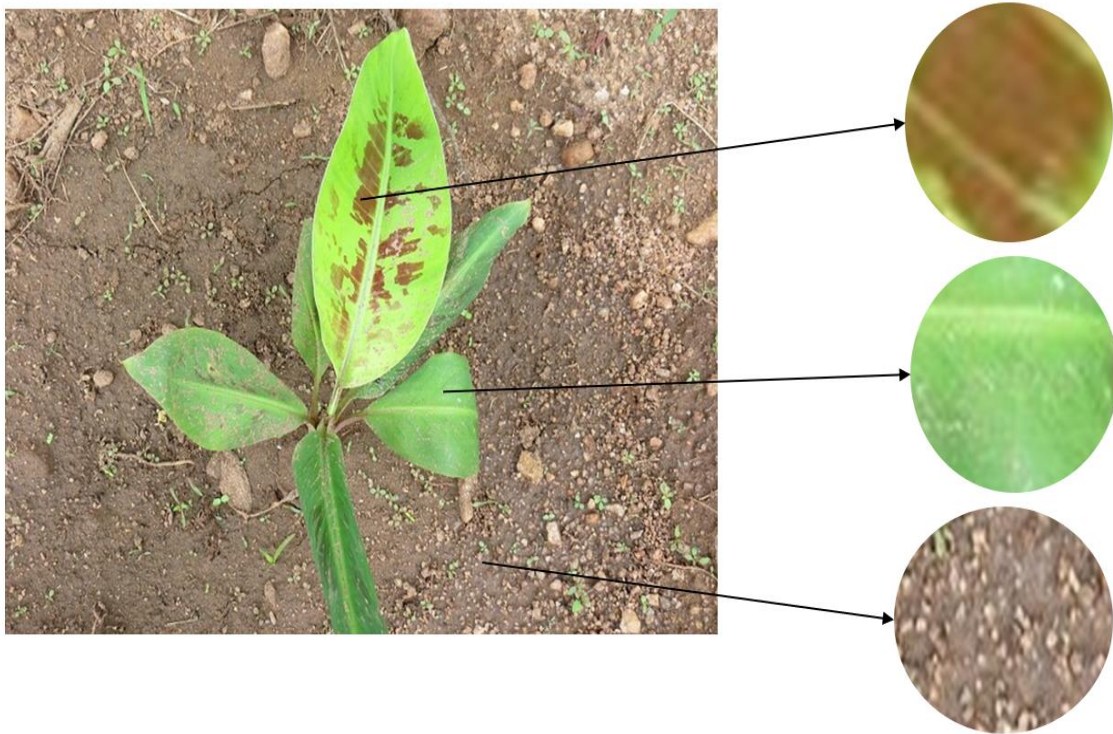


**Figure 2-11** Plant image with stress visual colour similar to the background colour

### **2.8.2 Developing the improved segmentation tool**

Plants exhibit several physical features to characterize their responses to environmental conditions. These features provide valuable information that can help distinguish plants from their surroundings. They are categorized as colour, texture and morphology. The proposed segmentation method used colour features from different spaces and dimensions, resulting in a 24-vector (features) for training the MLP model. The improved segmentation method used a combination of colour and selected texture features. Two Gabor features with different parameter settings, three local binary patterns (LBP) features with different parameter settings and entropy features (6 texture features in all) were extracted on each image patch and added to the 24-colour features to give a 30-dimensional feature vector. Figure 2.12 shows the workflow of the improved algorithm. Feature selection methods were applied to select the most informative features, which were used to retrain the MLP model.

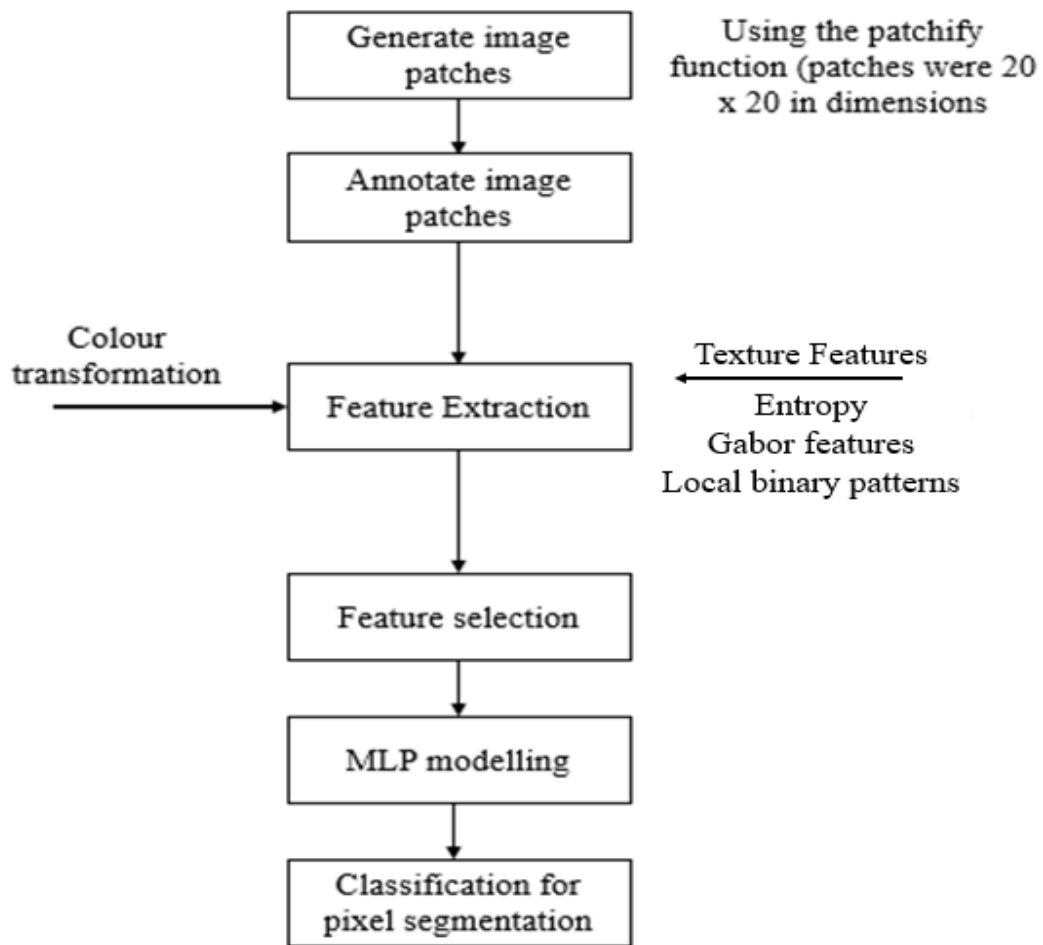


**Figure 2-12** Workflow of the improved segmentation algorithm

## 2.8.3 Machine learning-based segmentation process

### 2.8.3.1 Image patch extraction

Pre-processing steps, including colour corrections, were similar to the one used in the published paper. Unlike the previous method, where VGG software was used to extract and annotate patches, a custom-made algorithm using the patchify function in Python 3 was used to extract patches of size 20 x 20 pixels. The patches with plant and non-plant pixels were selected randomly as foreground and background, respectively and annotated as zero for non-plant and one for plant pixels.

### 2.8.3.2 Feature Extraction

*Colour feature extraction:* The extraction and transformation of colour features followed the same process explained in the published article.

*Entropy* is a statistical measure used to characterize an image's spatial complexity or roughness. It provides information about the texture of different areas within an image, enabling the classification of plant pixel areas from non-plant areas. The entropy feature has been used in diverse ways in image processing, including medical imaging (Pałczyński et al., 2022) and agriculture (Hall-Beyer, 2017) to segment, classify or detect traits of interest. Entropy calculation is based on the distribution of pixel intensities in an image. To compute the entropy of the image, the image is first transformed into a grayscale image, after which the histogram, which represents the frequency of occurrence of each pixel intensity level, is calculated. Next, the histogram is normalized to obtain a probability distribution  $p(i)$ . The entropy is then calculated for each pixel using the equation 2.9.

$$H = - \sum_{i=0}^{M-1} p(i) \cdot \log_2(p(i))$$

**Equation 2.9**

Where  $H$  is the entropy,  $p(i)$  is the probability of pixel intensity occurrence, and  $M$  is the total number of possible intensity levels. In entropy analysis for computational imaging, the higher the entropy value, the greater the randomness or complexity of the image, and vice versa.

*Gabor features:* These are designed to mimic human visualization systems and effectively capture local spatial frequency information and texture characteristics of images. It is a linear filter mainly employed in image processing for edge detection, feature extraction, and object detection (Jolly & Raman, 2016). It possesses optimum locality in the frequency and spatial domain, effectively capturing an image's local spatial frequency information and texture characteristics. When Gabor filters are applied to an image, the image is convolved using the filter function (given mathematically in equation 2.10) at multiple scales and orientations.

$$F(m,n;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{m^2 + \gamma^2 n^2}{2\sigma^2}\right) \cos\left(2\pi \frac{m'}{\lambda} + \psi\right)$$

**Equation 2.10**

Where  $m' = m\cos(\theta) + n\sin(\theta)$  and  $n' = -x\sin(\theta) + y\cos(\theta)$  are the coordinates rotated by an angle  $\theta$ .  $\lambda$  is the wavelength of the sinusoidal wave,  $\theta$  is the orientation of the filter,  $\psi$

is the phase offset,  $\sigma$  controls the standard deviation of the Gaussian envelope, and  $\gamma$  controls the aspect ratio of the Gaussian envelope.

*Local binary pattern (LBP):* LBP is a distinct and effective descriptor widely used in computer vision to describe local texture patterns in images. The fundamental idea behind LBP is to describe the local texture patterns of an image by comparing the grey level of each pixel with its neighbouring pixels (Turkoglu & Hanbay, 2019). LBP first defines a neighbourhood pixel around a centralised pixel in a circular or rectangular region. The central pixel's intensity value (threshold) is obtained and compared to the neighbouring pixels. A value of 1 is assigned to the neighbouring pixel if its intensity is greater or equal to the intensity of the central pixel, otherwise, 0 is assigned (Figure 2.13). This process is repeated for all the remaining pixels, which results in a binary pattern. This is used to construct a histogram representing the distribution of local texture patterns in the image. Mathematically, the LBP is given by equation 2.11.

$$LBP = \sum_{i=0}^{P-1} s(n_i - G_c) 2^i$$

**Equation 2.11**

$$S_x = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

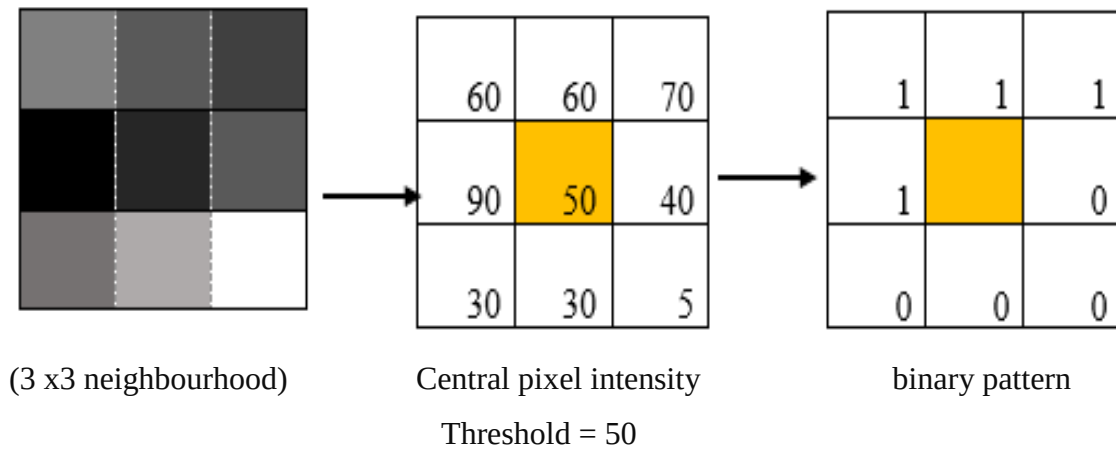


**Figure 2-13** Diagrammatic explanation of the LBP feature extraction

### 2.8.3.3 Feature Selection

In the improved tool (MLP<sub>i</sub>), 30 features were presented: 24 colour features and 6 texture features. The high-dimensional feature vector poses computational challenges. Moreover,

some features were redundant, as observed in the published paper. Hence, a recursive feature elimination method with a decision tree (DT-RFE) as the base model was employed to select the important features. This technique involved (a) training the decision tree model on all feature sets. (b) ranking of features by the RFE based on their importance scores provided by the decision tree algorithm (c) identification and removal of the least important feature based on the ranking (d) retraining of the decision tree model with the reduced feature set (e) repetition of steps b and c iteratively until a stopping criterion is met and the selected features are then used to train the final decision tree model. From Figure 2.14, the DT-RFE selected 15 features as the most important to the model. The features selected are presented in Figure 2.15. To compare the performance of the previous method ( $MLP_p$ ) to the improved one ( $MLP_i$ ), 24 colour features were extracted, and sensitive features were selected using the DT-RFE technique, as explained earlier for the  $MLP_p$ . In this case, 13 features were selected as shown on Figure 2.14. From Figure 2.15a, the R and G (from RGB), H and S (from HSV), y (from ybr), L and  $a^*$  (From Lab), u (from YUV),  $v^*$  (form Luv), l and s (from hls) and Z (form XYZ) were selected to train the proposed MLP model ( $MLP_p$ ). On the other hand, R and G (from RGB), H and S (from HSV), y (from ybr), L and  $a^*$  (From Lab), u (from YUV),  $v^*$  (form Luv), l and s (from hls), Z (form XYZ), GB1 (from Gabor), LB1 (from LBP ) and E (from Entropy) were selected to train the  $MLP_i$  model (Figure 2.15b)

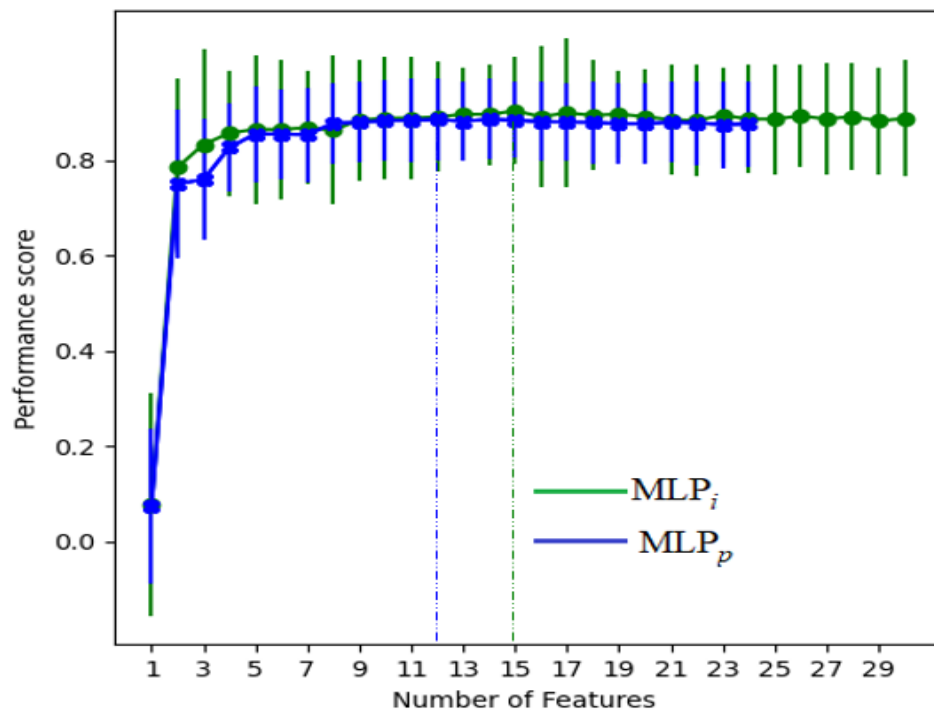
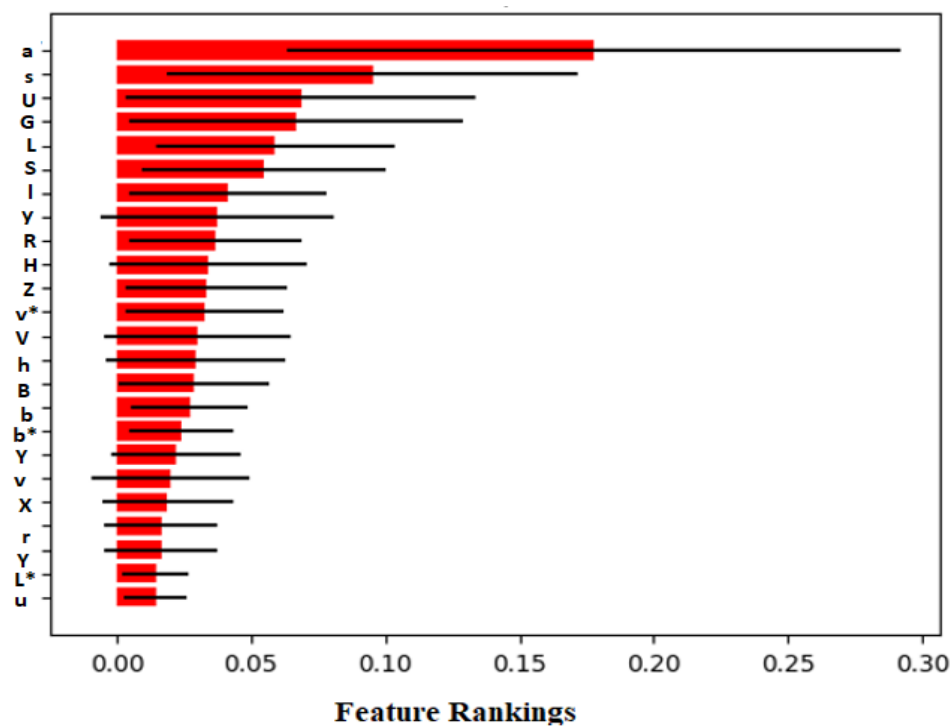
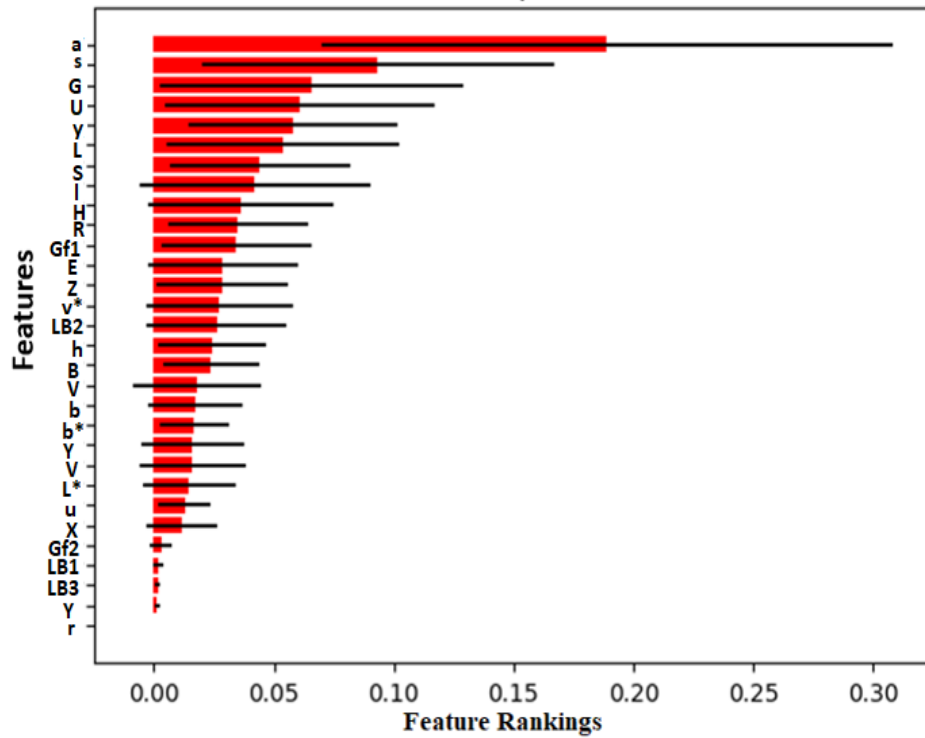


**Figure 2-14** Plot showing feature combinations and resulting output (the short-dashed lines show the optimum number of features relevant for the binary classification of pixels)



(a) Feature importance ranking of  $MLP_p$  model



(b) Feature importance ranking of  $MLP_i$  model

**Figure 2-15** A plot of feature importance for feature selection using the decision-tree-based recursive feature.

## 2.8.4 Image segmentation using state-of-the-arts CNN models

Two state-of-the-art models, U-Net and VGGNet, were trained using the extracted patches. Their performances were compared to the  $MLP_i$  segmentation model.

### 2.8.4.1 U-Net Model for Segmentation

U-Net is a 23-layer network commonly used in image-based plant phenotyping, particularly for segmentation and object detection. As shown in Supplementary information (Figure B-3), U-Net comprises an encoder and decoder network structured in a U-shape (Ronneberger et al., 2015). The encoder network, also called the contracting network, follows the typical CNN architecture. It consists of four blocks of convolutional layers with kernel size  $3 \times 3$  followed by a ReLU activation function. After each convolution, a  $2 \times 2$  max-pooling layer with a stride of 2 is applied to down-sample the feature maps, reducing the computation cost of the training model. The decoder network is symmetrical to the encoder block. It comprises  $3 \times 3$  convolutional layers that perform up-sampling operations with a ReLU activation function, halving the number of feature



channels. This improves the spatial dimensions of the feature maps, allowing the network to generate segmentation masks that are the same size as the input images. The final layer of the U-Net consists of a 1 x 1 convolutional layer with a sigmoid function that maps each 64-component feature vector to the desired number of classes.

#### **2.8.4.2 Pretrained VGG-16 and Random Forest Models for Segmentation**

The VGG-16 proposed by Simonyan & Zisserman (2015) is commonly used in semantic segmentation tasks. It is characterized by its small convolutional filters and its deep architecture. It has thirteen convolutional layers and three fully connected layers with the convolutional layers followed by max-pooling layers with fully connected layers at the end, as shown in Supplementary information (Figure B-4). While the convolutional layer extracts features from the input images, the fully connected layers are responsible for classifying the final results based on the characteristics of the extracted features. Using the VGG-16 as a pre-trained tool, generic information was extracted from the plant images for segmentation (Nguyen et al., 2022). The VGG-16 is used as a feature extractor, trained on our custom datasets, and used as input on a random forest classifier to predict classes of image pixels as either background or foreground.

Random forest is an ensemble learning method that builds multiple decision trees and combines their output to improve the overall accuracy and control overfitting. Each tree is trained on a set of randomly selected features, making the forest robust to noise. The model's overall performance was determined by majority voting on the individual trees' performance for classification purposes. The RF model receives the extracted features from the VGG-16 network as input, trains them, and outputs a binary classification.

#### **2.8.5 Dataset for training the model**

The datasets for the three models are top-view RGB images of quinoa, wheat, and sugar beet species with samples shown in Figure 2.16. The quinoa and wheat datasets are experimental data collected as part of our ongoing experiments at Cranfield University and Rothamsted Research, respectively. The sugar beet images are publicly available datasets downloaded from <https://www.v7labs.com/open-datasets/sugar-beets-2016>. Each image is of size 2448 x 2448 pixels resolution with different background. The quinoa images are of size 2548 x 2551 with slightly complex backgrounds. The wheat images are field-based data with 2400 x 2280 pixels with cluttered backgrounds, and

variable illumination. The datasets were 4500 in total: 1970 sugar beets, 1200 quinoa datasets and 630 wheat datasets. The quinoa and wheat images were captured at different growth stages, which ensures variations in the datasets. These datasets present challenges, including illumination variations, overlapping leaves, and complex backgrounds with different soil types, rocks, plant pots and trays. The leaves in the plants had different shapes and orientations. Moreover, for the quinoa and wheat plants, the leaves coil up under low nitrogen or phosphorus, decreasing the overall vegetation cover, while the leaves are open and broad under adequate N and P.

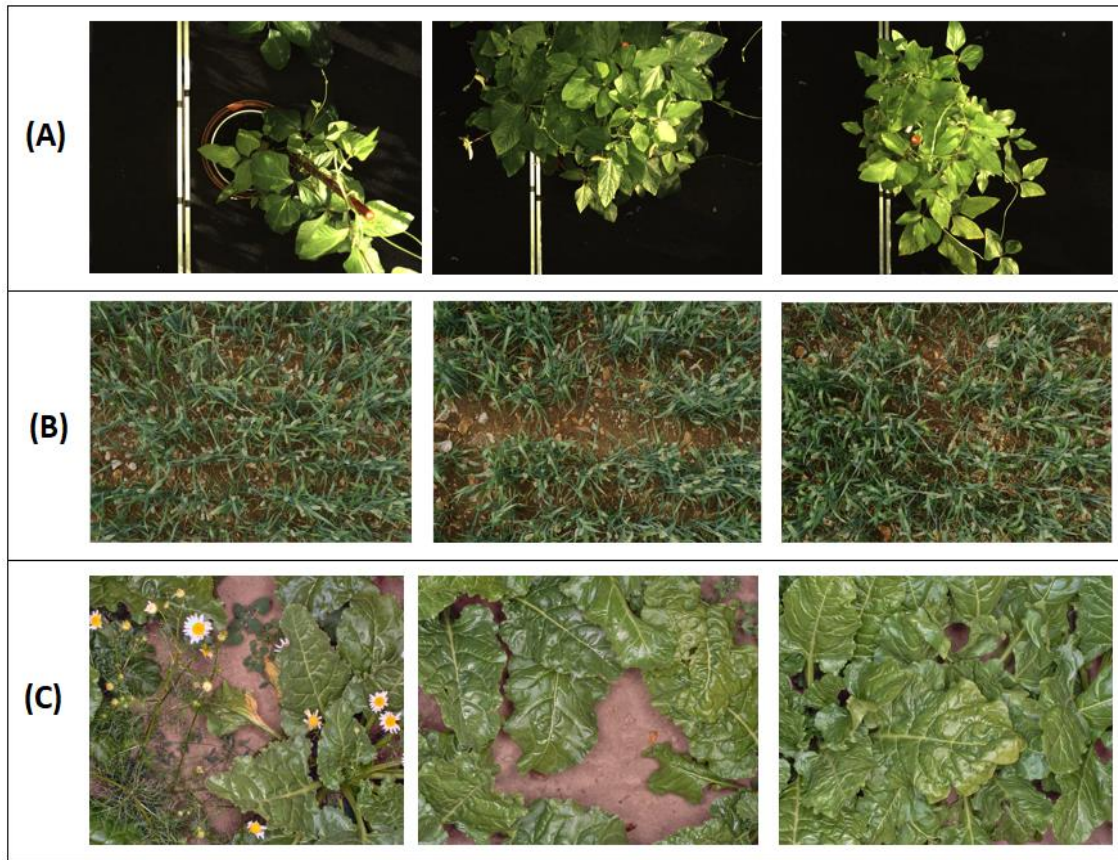


**Figure 2-16** Sample dataset images (A) Glasshouse quinoa images with different lightening conditions and superimposed leaves (b) Field images (wheat) with different non-plant matter (sand, rocks, etc.) (C) Sugar beet datasets with different background matter and leaves colouration

## 2.8.6 Training the networks

### 2.8.6.1 MLP model training

To train the MLP network, 10% of the images were randomly selected from the three datasets (quinoa, sugar beet, and wheat) and used for training and validation. Patches were extracted from the selected images and annotated, which were fed to the MLP

classifier for training. Patches were extracted and annotated, resulting in 4500 patches from the foreground (2300) and background (2200). This gives a total of  $4 \times 10^6$  pixels, split into 70%:30% for training and validation. To compare the performance of the improved model to the state-of-the-art models, 18,000 patches were extracted from 4500 plant images. The MLP model was retrained using the selected features. Due to changes in the number and types of training features used, the hyperparameters were fine-tuned to ensure the robustness and generalization of the model. The MLP hyperparameters include hidden layer [(50,50,50), (50,100,50), (100,)], activation function: ReLU, solver: adam, learning rate: adaptive, alpha: 0.001,0.05. In this case, the hidden layer and alpha parameters were systematically tested for possible parameter values using the grid search cross-validation technique (K=5) while the rest remained constant. The trained model was randomly tested on 10 % of each dataset that was not part of the training processes.

#### **2.8.6.2 Training the U-Net and VGG-16 models**

To train the U-Net model, batches of images with their ground truth masks were used as input to the network. Different types of augmentation methods, including shear, rotation ( $50^\circ$ ), zoom range (0.2), width shift (0.2), and height shift range (0.2), were applied to increase the diversity of the dataset, resulting in a total of 20,000 images. 70% of the datasets were used for training and 30% for testing. The images were resized to  $224 \times 224 \times 3$  pixels for both networks. The parameters of the network was trained using Adam optimizer with a fixed learning rate of 0.00001, exponential decay rate of 0.9 for the first-order moments and 0.0999 for the second-order-moments and epsilon with a small value of  $1e^{-08}$  for numerical stability and batch size of 10.

To train the VGG-16 network, the images were resized into a  $224 \times 224 \times 3$  pixel dataset and fed into the model. The model was modified by freezing or removing the fully connected layers at the end while retaining only the convolutional layers. The images were passed into the modified VGG-16 to extract high-dimensional feature maps encapsulating complex patterns and trends useful for segmentation. A stochastic gradient descent with a fixed learning rate of  $1 \times 10^{-3}$  and a momentum of 0.9 was used. The optimization process was carried out over a maximum of 100 epochs, where each epoch passes through the entire dataset. A corresponding feature vector from the feature maps is produced for each pixel of the input image, resulting in a set of feature vectors for training the random forest classifier.

It should be noted that the U-Net was trained from scratch, while the VGG-16 model was used as a pre-trained model through the transfer learning process.

### 2.8.7 Evaluating the performance of the models

The performance of the MLP<sub>i</sub> model and the two state-of-the-art models were evaluated using the Dice coefficient (DC) (Zou et al., 2004), the Jaccard Index (IoU) and the F1-score. The F1-score is a commonly used metric for binary classification problems, e.g., image segmentation and is particularly useful when dealing with class imbalance. The IoU measures the overlap between the predicted image and ground truth images. At the same time, the DC is a similarity metric that measures the overlap between the predicted mask and the ground truth mask. All three metrics vary from 0 to 1, with 1 indicating 100% accuracy and 0 indicating 0% accuracy. Mathematically, the F1-score, the DC and IoU are expressed as shown in equation 2.12- 2.14.

$$DC = \frac{2 * TP}{2 * TP + FP + FN} \quad \text{Equation 2.12}$$

$$IoU = \frac{TP}{TP + FP + FN} \quad \text{Equation 2.13}$$

$$F1\text{-Score} = 2 \frac{Precision + Recall}{Precision + Recall} \quad \text{Equation 2.14}$$

$$\text{Where Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

Where TP are true positives: plant pixels predicted correctly; FP is false positive: pixels that are background pixels predicted as plant pixels; TN is true negative: background pixels predicted correctly; and FN is false negative: background pixels predicted falsely as plant pixels. Predicted.

## 2.8.8 Results and Discussions

### 2.8.8.1 Performance of the MLP classification models

This section compares the performance of the  $MLP_p$  and  $MLP_i$  models using three metric scores: accuracy, F1-score and Recall score. Both models were trained with features selected from quinoa, wheat, and sugar beet species patches. Figure 2.17 shows the performance of the  $MLP_p$  and  $MLP_i$  models in classifying the pixels for segmenting plants. Both models performed well, with the  $MLP_i$  model achieving the highest classification performance with 0.985, 0.974 and 0.982 accuracy, F-score and recall score, respectively. It was observed that the addition of the texture features (entropy, gabor and LBP features) improved the performance of the MLP model. Although not significantly different in terms of performance between both  $MLP_p$  and  $MLP_i$  models, the slight improvement shows the effect of the texture features on the MLP model. However, it is observed that colour features play a dominant role in image segmentation as reported in the published articles and other studies (Adams et al., 2020).

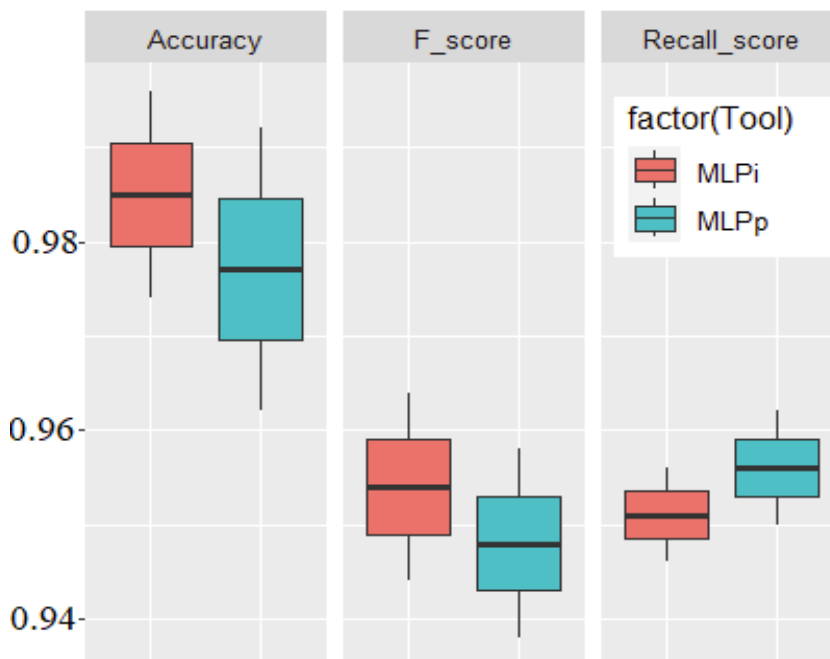


**Figure 2-17** Comparison of  $MLP_p$  and  $MLP_i$  models based on accuracy, F-score, and recall. Note: MLP models were trained with 13 features based on only the colour information of plants, while the  $MLP_i$  were trained with 15 features based on plants' colour and texture information. Error bars showing only the positive component correspond to one standard deviation

### 2.8.8.2 Performance comparison of the MLP and state-of-the-art CNN models

In this work, three evaluation metrics (dice coefficient, IoU and the F-score) were utilized to analyse and compare the performance of the improved MLP model (MLP<sub>i</sub>) and the two state-of-the-art models in the literature. Table 2-7 compares test results of the improved MLP model, U-Net and VGG-16 models on the quinoa, wheat and sugar beet datasets. Although the three datasets had varied topologies with varying background and lighting conditions, all the models had a ceiling effect performance. Figure 2.18 shows the segmentation outcome for the quinoa dataset using the three models. Row B displays the ground truth images that match the RGB images in row A. Predicted masks for MLP<sub>i</sub>, U-Net, and VGG-16 models are shown in rows C, D and E, respectively. Based on visual examination, all three models demonstrated good performance in differentiating between background and plant pixels. The U-Net model had the best performance, with a 0.971 dice coefficient, showing a high spatial overlap between the ground truth and the segmented image. This was followed by the VGG-16 model with a 0.966 dice score and the MLP<sub>i</sub> model with a 0.952 dice coefficient. Despite the models' varied scores, there was no significant difference in their performances on the quinoa datasets. With much complex background and variable light intensities, the models were observed to work well on the wheat datasets (Figure 2.19). Visual observation noted that segmentation was effective on images with less vegetation than those with dense vegetation. The U-Net and MLP<sub>i</sub> models achieved the highest dice score ( $> 0.95$ ), while the VGG-16 model followed with a 0.942 dice score. In a similar work, Kolhar and Jagtap (2021) obtained a similar dice score (0.959) when they applied a U-Net segmentation model on fig field images. Figure 2.19 shows the segmentation results for the sugar beet dataset. The findings demonstrate that U-Net and MLP<sub>i</sub> marginally outperformed the VGG-16 model in segmenting overlapping leaves and can construct exact borders for plant leaves. Although with few training datasets comparably, the improved MLP<sub>i</sub> model had similar scores as the state-of-the-art model, suggesting using this method for small datasets instead of deep learning models, which require more datasets for training. The models were observed to be highly efficient in the glasshouse (quinoa) images and sugar beet datasets than in the wheat datasets. This could be because of the high variability of illumination and light intensity on the field and the several different non-plant materials as compared to the quinoa or sugar beet datasets.

**Table 2-7** Performance of MLPi, U-Net and VGG-16 models for the quinoa, wheat and sorghum dataset segmentation

| Dataset    | Metric |       |                  |         |
|------------|--------|-------|------------------|---------|
|            | Model  | IoU   | Dice coefficient | F-score |
| Quinoa     | MLP    | 0.954 | 0.952            | 0.957   |
|            | U-Net  | 0.966 | 0.971            | 0.962   |
|            | VGG-16 | 0.951 | 0.966            | 0.959   |
| Wheat      | MLP    | 0.951 | 0.953            | 0.947   |
|            | U-Net  | 0.956 | 0.952            | 0.956   |
|            | VGG-16 | 0.942 | 0.940            | 0.948   |
| Sugar beet | MLP    | 0.964 | 0.966            | 0.961   |
|            | U-Net  | 0.973 | 0.979            | 0.971   |
|            | VGG-16 | 0.960 | 0.958            | 0.961   |

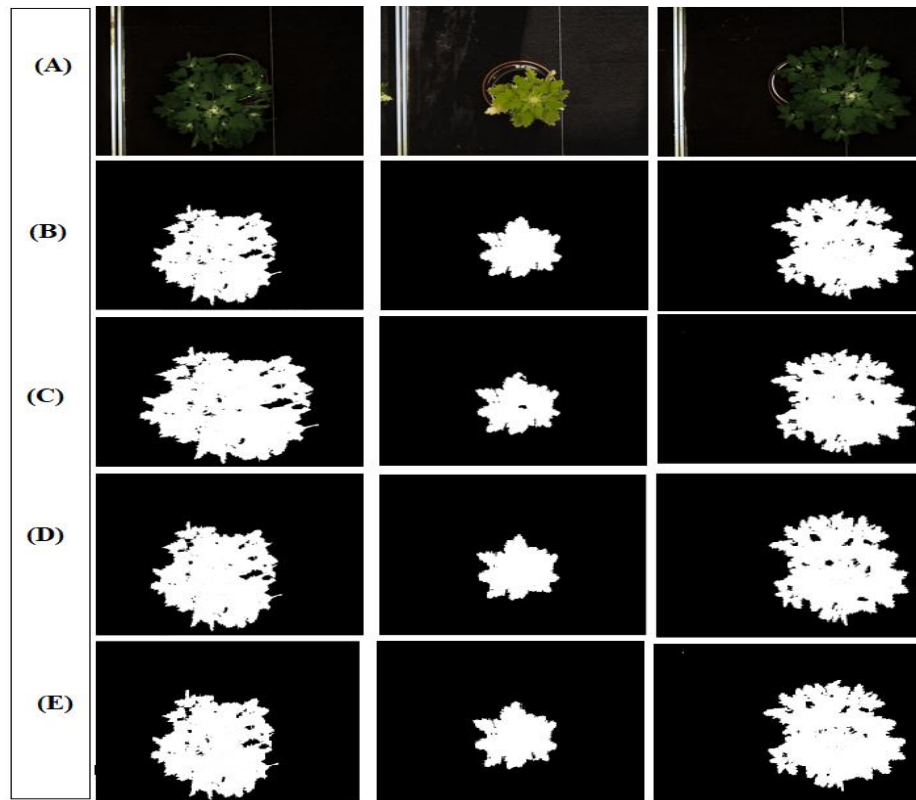


**Figure 2-18** Test results on selected glasshouse images (A) original images (B) Corresponding ground truth images (C) MLPi predicted masks (D) U-Net predicted masks (E) VGG-16 segmented images (masks).



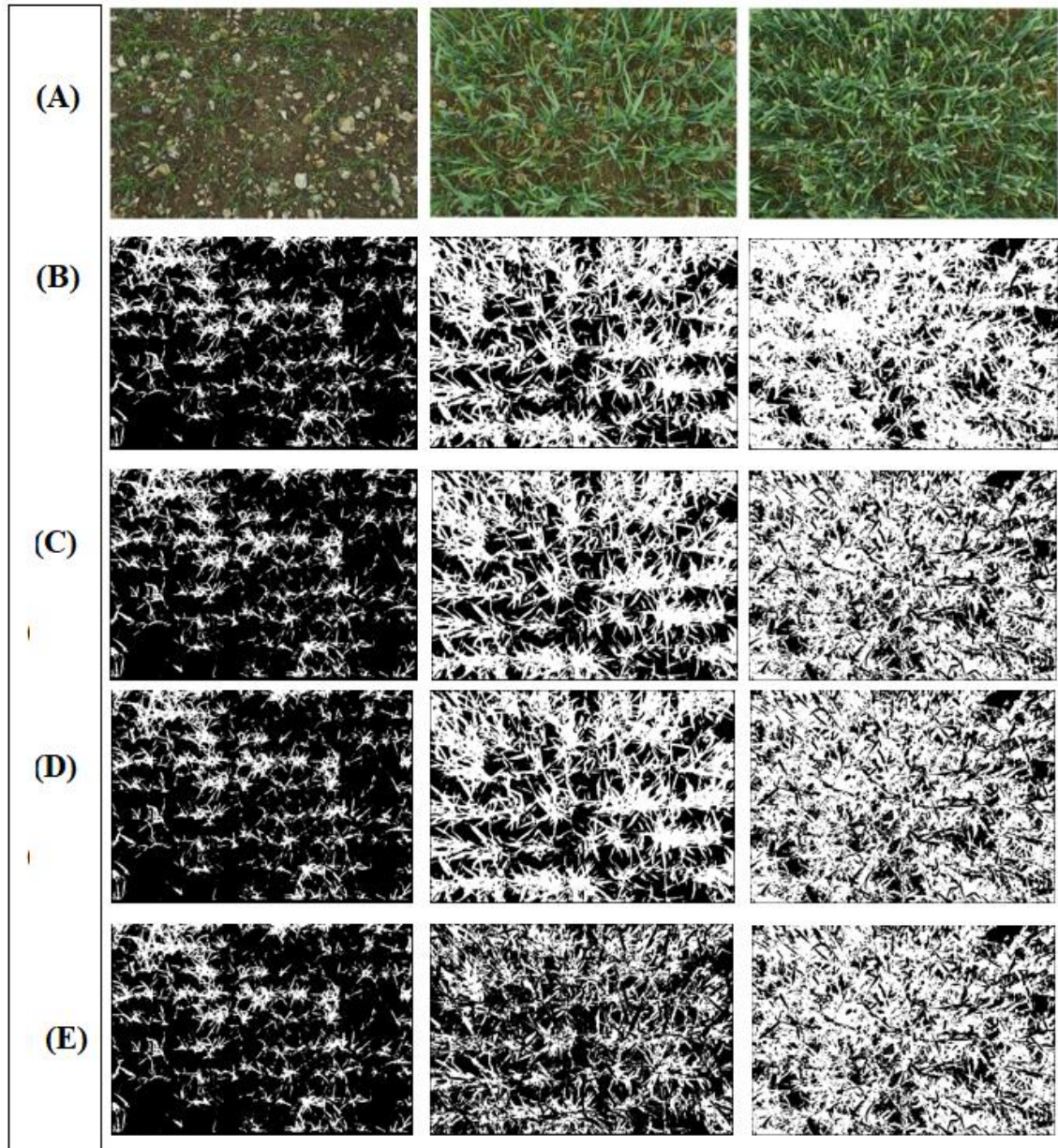


**Figure 2-19** Test results on selected field images (wheat) (A) original images (B) Corresponding ground truth images (C) MLPi predicted masks (D) U-Net predicted masks (E) VGG-16 segmented images (masks).



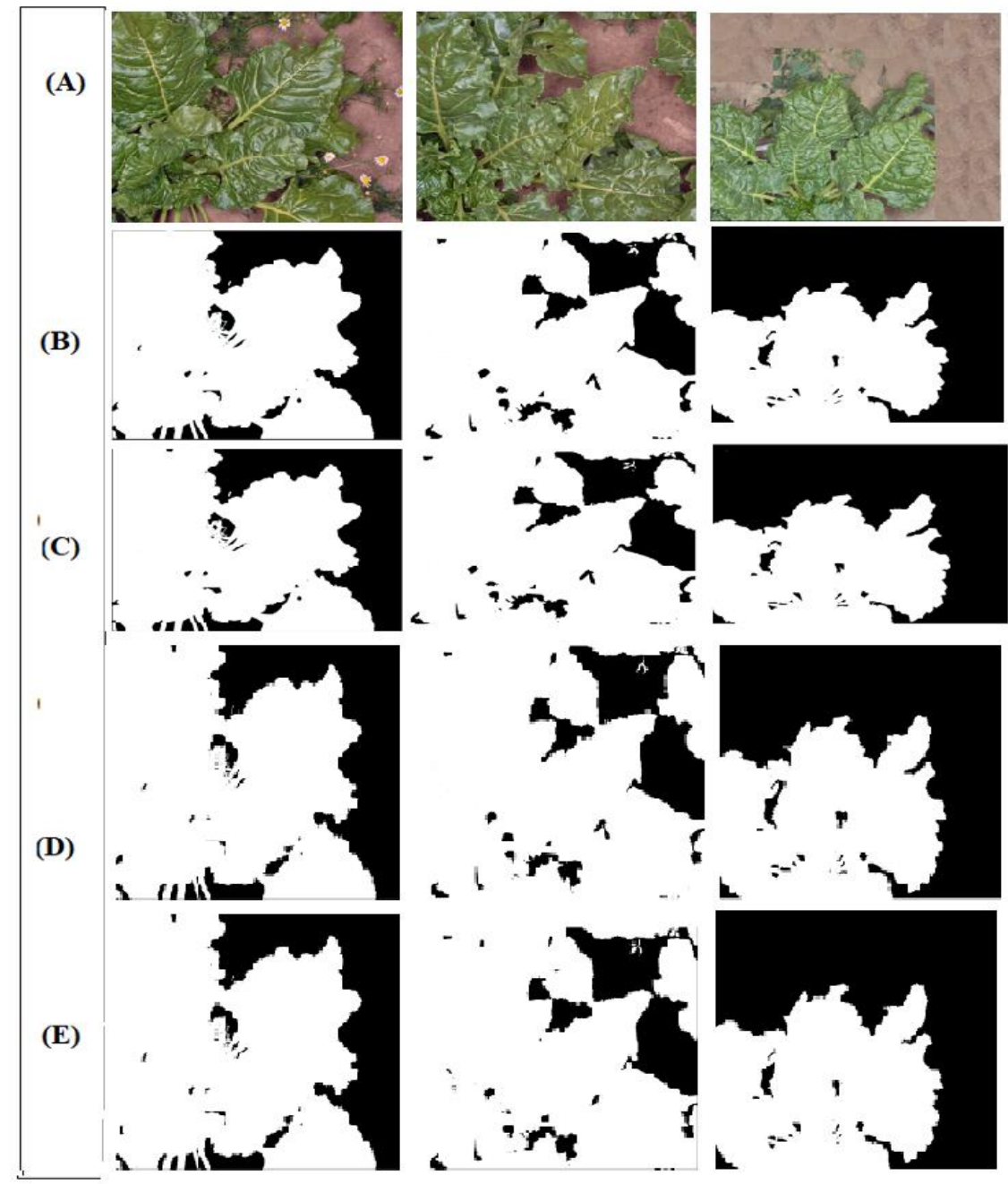


**Figure 2-20 .** Test results on selected sugar beet images (A) original images (B) Corresponding ground truth images (C) MLPi predicted masks (D) U-Net predicted masks (E) VGG-16 segmented images (masks).

### 2.8.9 Conclusion

Previous studies proposed a multi-feature segmentation method to segment plants from different environments. This method was based on extracting and selecting colour features combined with Otsu thresholding to segment images from multiple environments. Although this method works, it faces challenges when plants with visible severe stress symptoms are processed. Hence, an improved version has been proposed that combines selected colour and texture-based features for plant image segmentation. This was compared to state-of-the-art deep learning segmentation tools such as U-Net and VGG-16 models. The results show a high performance of the improved MLPi model comparable to the DL models, indicating a possibility of using this tool in place of the DL models, which requires huge datasets and high computational memory for training.

### 2.8.10 Reference

- Adams, J., Qiu, Y., Xu, Y., & Schnable, J. C. (2020). Plant segmentation by supervised machine learning methods. *Plant Phenome Journal*, 3(1), 1–11. <https://doi.org/10.1002/ppj2.20001>
- Hall-Beyer, M. (2017). Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing*, 38(5), 1312–1338. <https://doi.org/10.1080/01431161.2016.1278314>
- Jolly, P., & Raman, S. (2016). Analyzing Surface Defects in Apples Using Gabor Features. *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 178–185. <https://doi.org/10.1109/SITIS.2016.36>
- Kolhar, S., & Jagtap, J. (2021). Convolutional neural network based encoder-decoder architectures for semantic segmentation of plants. *Ecological Informatics*, 64(November 2020), 101373. <https://doi.org/10.1016/j.ecoinf.2021.101373>
- Nguyen, T. H., Nguyen, T. N., & Ngo, B. V. (2022). A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease. *AgriEngineering*, 4(4), 871–887. <https://doi.org/10.3390/agriengineering4040056>
- Pałczyński, K., Ledziński, D., & Andrysiak, T. (2022). Entropy Measurements for Leukocytes' Surrounding Informativeness Evaluation for Acute Lymphoblastic Leukemia Classification. *Entropy*, 24(11). <https://doi.org/10.3390/e24111560>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). Chapter 2. U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M.

Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015* (pp. 234–241). Springer International Publishing.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.

Turkoglu, M., & Hanbay, D. (2019). Leaf-based plant species recognition based on improved local binary pattern and extreme learning machine. *Physica A: Statistical Mechanics and Its Applications*, 527, 121297.  
<https://doi.org/https://doi.org/10.1016/j.physa.2019.121297>

### **3 Selection of Nitrogen and Drought-Responsive Plant Phenotypic Traits in Wheat Using Machine Learning Methods**

Frank Gyan Okyere <sup>1,2</sup>, Daniel Cudjoe <sup>1,2</sup>, Pouria Sadeghi-Tehran<sup>1</sup>, Nicolas Virlet<sup>1</sup>, March Castle <sup>1</sup>, Andrew B. Riche<sup>1</sup>, Latifa Greche<sup>1</sup>, Fady Mohareb <sup>2</sup>, Daniel Simms <sup>2</sup>, Manal Mhada <sup>3</sup> and Malcolm John Hawkesford <sup>1, \*</sup>

<sup>1</sup>Sustainable Soils and Crops, Rothamsted Research, Harpenden, United Kingdom, 2 School of Water.

<sup>2</sup>Energy and Environment, Cranfield University, Cranfield, United Kingdom, 3 AgroBioSciences.

<sup>3</sup>Department, University of Mohammed VI Polytechnic, Ben Guerir, Morocco

Based on::Okyere FG, Cudjoe D, Sadeghi-Tehran P, Virlet N, Riche AB, Castle M, Greche L, Simms D, Mhada M, Mohareb F and Hawkesford MJ (2024), Selection of nitrogen and drought-responsive plant phenotypic traits in wheat using machine learning methods. Submitted to MDPI plants for consideration for publication

#### **3.1 Abstract**

Plant stress from inadequate nitrogen or water supply can be detrimental to plant health and survival. Hence, image-based methods for the early detection of these stresses is crucial to restoring productivity and ensuring recovery and high plant yields. To analyze the sensitivities of plant features responsiveness to drought stress in variable nitrogen levels, a wheat pot experiment with four treatments made of two levels (high and low) of nitrogen and water was conducted in the glasshouse. Eleven hand-crafted features (manually crafted representations of data) were extracted from the top and side views of images and used to train classical machine learning models. A sensitivity analysis using a one-at-a-time modelling method and a global-sensitivity-analysis based on Morris method was performed to assess the responsiveness of the extracted features to changes in the nitrogen and water levels of plants. A CNN model was proposed and combined with the hand-crafted features to identify drought-stressed plants of variable nitrogen levels. The results show that ML models trained with combined features outperformed

those trained with individual features. The sensitivity analysis identified ASM, HOG as very sensitive to drought stress in plants with variable nitrogen content. A one-way ANOVA analysis of the time series of the most sensitive feature (ASM) shows that drought stress significantly affected the plants on the 6 DADS (day after drought stress). The proposed CNN model outperformed the classical ML models, achieving an 11.2% better accuracy. Also, when trained with the non-segmented dataset, the CNN model had results similar to state-of-the-art ResNet and DenseNet models. Furthermore, the fusion of the hand-crafted features with the CNN model resulted in a 2.5% increase in the performance of the model.

**Keywords:** convolutional neural network, drought, nitrogen, image features, machine learning

### 3.2 Introduction

Wheat is the second most-produced cereal, following maize, with a total global production of 761 metric tons in 2020 (Guarin et al., 2022). Wheat production and yield are expected to double in the coming years to meet the 2050 food security challenge (Serraj & Pingali, 2019). To accelerate this development, plant phenotyping research focuses on nutrient and water stress, two of the most threatening factors affecting wheat production, especially in arid and semi-arid regions. Plant nutrients (especially nitrogen) and water requirements are essential at every growth stage of plants, from seed germination to senescence and the amount required increases with increasing plant growth. In wheat, lack of water can cause a reduced number of tillers and above-ground biomass at the tillering stage (Liwani et al., 2019). It also affects leaf extension rate, leaf number and plant area, inhibiting growth, especially in the roots (Naveed et al., 2014). Similarly, plants deficient in nitrogen manifest chlorosis with yellowish-brown colouration symptoms. Hence, early detection of these stresses is crucial for high crop yield and quality improvement while reducing production costs (Zhou et al., 2021).

Conventionally, breeders assess nitrogen and drought stress in plants by visual scoring of the stay-green morphological properties of plants (Duan et al., 2018). However, this method is subjective and error-prone as the change in colour can depict other stresses. Methods such as assessing leaf stomatal conductance and stem water potential (Berni et

al., 2009) for plant drought stress, as well as chlorophyll measurement for nitrogen deficiency, can assess plant stress non-destructively. Also, single-point field measurement tools such as soil moisture sensors and chlorophyll meters provide consistent approaches to indirectly track plant water and nitrogen status. Although plant breeders can rely on these methods, they are subject to installation and location, accounting for plant spatial variation when measured on the field (Mangus et al., 2016).

Non-invasive imaging techniques such as digital RGB, fluorescence, thermography and spectroscopy are investigated for reliable high-throughput plant phenotyping. They have been used to analyse different stresses such as plant water, diseases and nutrient deficiencies. Among these imaging techniques, digital RGB imaging is the most widely used for phenotyping because the sensor is less costly, portable and produces high-resolution images. Digital images are acquired from the net intensity reflectance of the red, green and blue channels to provide colour, morphological and texture features to estimate nutrient deficiencies (Zermas et al., 2021), water stress (Zhuang et al., 2017b), yield (Khaki et al., 2020) and biomass (Vergara-Diaz et al., 2015). Changes in plant physical traits can express drought or nitrogen stress mechanisms (Briglia et al., 2019). As a representative study, Zhuang et al. (2017) presented a model trained with colour and texture features for drought stress detection in maize. The colour features were the mean and variance of the green-blue (G-B) and green-red (G-R) colour channels, while the texture features were energy-based wavelet transform. Their results showed that the model was an effective detection tool for early water stress. Also, Azimi et al. (2021) extracted SIFT and HOG features for nitrogen deficiency classification. They proposed a CNN model capable of measuring nitrogen stress in sorghum which outperformed classical machine learning methods.

Although research has shown that single or combined phenotypic features can provide comprehensive descriptions of abiotic stress symptoms, there is little knowledge of the sensitivities of these features to plant stress. This makes choosing appropriate stress-related features difficult. Moreover, prolonged nitrogen and drought stress have similar visual symptoms, making it difficult to use a single phenotype to distinguish between these two stresses. Hence, identifying key features related to a particular stress may be helpful.

Machine learning (ML) is a powerful tool for image analysis and can be used to classify or predict changes in plant traits for disease, water and nutrient stress. Approaches like Support Vector Machines (Wilf et al., 2016) and Random Forests (Shah et al., 2019) have high success rates, demonstrating ML's versatility and power as one of the most potent analytic techniques. Although ML methods yield good results in various scenarios (Singh et al., 2016), their accuracy is often insufficient for fully automated systems since they mostly require hand-crafted features for training. Hand crafted features are manually extracted features representing specific aspects of an image relevant to the problem at hand. Deep learning (DL) offers the ability to learn features themselves, which can increase accuracy if enough training data is available. DL architectures, especially convolutional neural networks (CNNs), have convolutional layers capable of automatic feature detection with artificial neural networks for classification and regression purposes (LeCun et al., 1998). This study proposes using a CNN model to identify drought stress plants with different nitrogen concentration. A performance analysis of the proposed model was conducted and compared to the performance of classical ML models and two state-of-the-art DL models (ResNet and Dense models). Furthermore, a sensitivity analysis of the extracted features was performed to assess the effect of nitrogen deficiency and drought stress on selected hand-crafted features. Finally, an analysis to understand the dynamics of the performance of the proposed model when fused with the hand-crafted features was performed. The remainder of the paper is as follows: section 3.3 describes the experiment setup and the proposed methodology. Section 3.4 outlines the results while section 3.5 discusses the outcome of the experiment. The final section provides some perspective and conclusion.

### **3.3 Materials and Methods**

#### **3.3.1 Experiment setup**

The experiment was conducted in a glasshouse plant growth facility (<https://www.cranfield.ac.uk/facilities/plant-growth-facility>, accessed on 10 November 2022) at Cranfield University (Cranfield, Bedfordshire, UK). This fully automated facility has a Lemnatec Scanalyzer system for high throughput image acquisition. The wheat cultivar, Cadenza, was sown in pots (8 cm x 6 cm) filled with low-nitrogen peat

compost (Levington Advance M3, CTS, Garden Supplies, UK). Plants (48 pots) made of four treatments and twelve replicates were structured in a randomised complete block pattern. The treatments were well-watered high-nitrogen (WWHN), well-watered low-nitrogen (WWLN), drought-stress high-nitrogen (DSHN) and drought-stress low-nitrogen (DSLN). At 30, 37 and 44 DAS (days after sowing), the plants were fertilized with two nitrogen (N) levels: high N and low N of 42.5 mM and 4.25 mM concentration, respectively. They were prepared from a modified Letcombe nutrient solution (Masters-Clark et al., 2020).

To prepare the drought stress treatments, all the plants were watered with equal amounts of deionized water till the tillering stage (50 DAS) when the drought stress was induced. At 50 DAS i.e., 0 DADS (day after drought stress), the watering was stopped for the DSHN and DSLN plants. This pattern continued until the end of the experiment which lasted for 15 days. The volumetric soil water content was monitored daily using a portable soil moisture meter (SM150T model, Dela-T Devices Ltd, Cambridge, UK). All the plants were exposed to natural light during the day with an average light intensity of 450-600  $\mu\text{mol}/\text{m}^2/\text{s}$  PAR with a 20 °C to 23 °C and 18 °C to 20 °C temperature range during the day and night, respectively. The soil surface was covered with white pebbles to minimize water loss through evaporation. Supplementary information (Figure C-1) shows samples wheat plants at different drought stress stages in the glasshouse.

### **3.3.2 Low throughput physiological measurement**

Three low throughput physiological measurements that have bearings on nitrogen and drought stress were measured. These include nitrogen content and two gas exchange parameters (stomatal conductance and CO<sub>2</sub> assimilation).

#### **3.3.2.1 Leaf nitrogen content(N)**

Plant leaves were sampled at the 0, 6 and 15 DADS and oven-dried for 72 h. This was milled in a Genogrinder (SPEX SamplePrep®, 2010, USA) at 1750 rpm. The milled sample was combusted using the Dumas combustion method. After complete combustion, reduction and purification, wheat's nitrogen content was estimated using the LECO



combustion method (LECO CN628 Analyzer, LECO Corporation, St Joseph, Michigan, USA).

### **3.3.2.2 Gas exchange measurement**

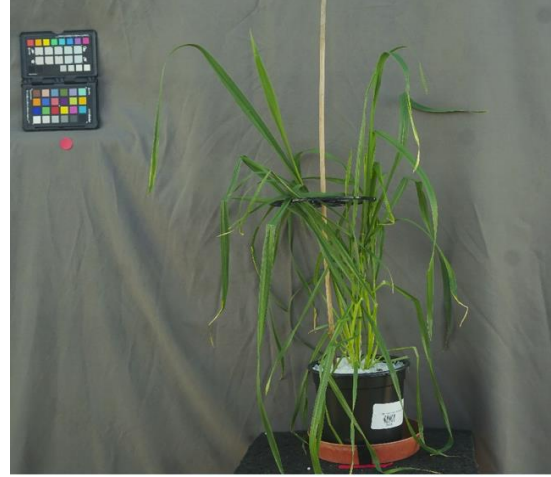
Gas exchange measurements, including net CO<sub>2</sub> assimilation ( $P_n$ ) ( $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ) and stomatal conductance ( $g_s$ ) ( $\text{mmol H}_2\text{O m}^{-2} \text{ s}^{-1}$ ) were measured every three days from the 0 DADS and continued till the end of the experiment using a LI-6400XT portable photosynthesis system (LI-COR Biosciences Inc., Lincoln, NE, USA). The fully expanded leaf of the main stem was marked and readings were taken between midday to 4 p.m. The CO<sub>2</sub> concentration inside the leaf chamber was set at 400  $\mu\text{mol CO}_2 \text{ mol}^{-1}$ . The leaf flow rate, temperature, relative humidity and photosynthetically active radiation were set at 200  $\mu\text{mol s}^{-1}$ , 20 °C, 50–65 % and 1800  $\mu\text{mol m}^{-2} \text{ s}^{-1}$ , respectively, to attain the maximum photosynthetic capacity. The values of  $g_s$  and  $P_n$  were recorded once there was stabilization of the measurement. Due to the relatively large ratio of the whole chamber cuvette and wheat leaf, the leaf area was adjusted during data processing.

### **3.3.2.3 Image data acquisition**

Two sets of images were acquired from the top area (canopy level) and side at 90°. The top images were acquired using an RGB camera (colour 12-bit Prosilica GT3300) with 3296 × 2472 pixel resolution installed on the Lemnatec phenotyping platform in the glasshouse. The camera was set at 2.5 m in the nadir view above the plant canopy, covering an area of 0.48 m<sup>2</sup>. The side images were taken with a Sony A5100 camera (Sony Alpha series, Sony AI Inc. Japan) with resolution of 6058 x 4012 pixels mounted on a tripod and positioned 1.5 m away from the plants. The images captured were within an approximated area of 0.52 m<sup>2</sup>. Imaging commenced on the day the drought stress was induced (0 DADS) and continued daily until the end of the experiment. The camera was set to automatic exposure mode to compensate for the changes in natural light. During scanning, a colour target (24 ColourChecker classic, X-Rite Pantone, Manchester, UK) was placed on the imaging bench to calibrate the colour variations resulting from illumination changes. The camera position and other parameter settings remained constant throughout the experiment. All the images were stored in the PNG format. Figure 3.1 shows sample images from the top and side views.



(a) Top view at  $0^\circ$



(b) Side view at  $90^\circ$

**Figure 3-1** Visualization of image samples representing the top and side views of the plants. The top view images were taken at  $0^\circ$  to the axis of the camera and 3 m from away from the camera whilst the side view images were taken at  $90^\circ$  from the camera axis and 1.5 m away from the ground. A colour checker was placed beside the plants during each image acquisition for colour correction using a custom made algorithm.

### 3.3.3 Image pre-processing

#### 3.3.3.1 Image colour calibration

During imaging, light intensity and illumination changes can cause colour rendering to deviate significantly from the true or perceived colour. Colours are calibrated using a reference ColourChecker (CC) to correct these deviations. The aim is to transform the image pixel values such that the colour checker tiles match the true values after the transformation (Chopin et al., 2018). In this study, the colour correction was conducted in the CIEL\*a\*b colour space. This was adopted because it is device-independent and known to perform better in colour correction than the conventional RGB colour space (Mantiuk et al., 2009).

#### 3.3.3.2 Image segmentation

The images were segmented to remove the background for effective plant feature extraction. We used the method proposed by Okyere et al., (2023), where plant pixels were segmented using the multiple-feature-based machine learning approach. In this

process, image patches of sizes 15 by 15 from 20 randomly selected plant images were transformed into multiple colour spaces. The pixels in the patches were annotated as foreground and background, with each pixel having 24-vector colour properties. A feature selection method was applied to the vector to reduce the dimension from 24 to 12. The selected features were used as input to train a multi-layer perceptron model to segment the images into plant and non-plant pixels. After segmentation, a median filter of window size five was applied to remove the noise generated by the misclassification of the model.

### **3.3.4 Handcrafted feature extraction**

Hand-crafted features are parts or patterns of images with characteristic key points that describe local patches in the neighbourhood of those points. They are manually designed representations taken from raw image data and are used as inputs for ML models. In this study eleven hand crafted features from colour, texture, morphology and local descriptors were employed to assess their sensitivity to nitrogen and drought stress.

#### **3.3.4.1 Colour Features**

Colour features are essential in plant stress detection. During plant growth, non-stressed plant leaves with adequate water content exhibit verdant green colour while stressed plants (primarily nitrogen and drought) show withered signs with yellow to brown colour. This means that colour changes could be used as a proxy for plant stress measurement. However, subtle colour changes are challenging for human eyes to detect. These changes can be evaluated by statistical methods that relate the colour information to the stress under study. Three colour features related to green dominance in plants were extracted. These include dark green colour index (DGCI), green blue Colour Index (GPB) and red, green blue vegetation index (RGBVI).

*Dark Green Colour Index (DGCI):* This colour index introduced by Karcher and Richardson, (2003) expresses the greenness intensity of leaves canopy. DGCI values range between 0 and 1, where 0 denotes light yellow and 1 is the dark green colour. It is derived from the hue, saturation and brightness colour channels using Equation 3.1.

#### **Equation 3.1**

$$DGCI = \frac{(hue - 60)/60 + (1 - saturation) + (1 - brightness)}{3}$$

*Green Blue Colour Index (GPB):*

This feature is the mean of the variance between the green and blue channels of an image. It is derived from equation 3 as

$$GPB = \frac{1}{SP} \sum_{i=1}^{SP} (G_i - B_i)$$

**Equation 3.2**

where  $G_i$  and  $B_i$  are the values of green and blue channels at  $i^{th}$  pixel. SP is the sum of wheat plant pixels in each image.

*Red, Green Blue Vegetation Index (RGBVI):* This is a green dominant feature derived from the difference between the sum of the red and green colour channels and the blue channel of an image. Mathematically, the RGBVI is given as:

$$RGBVI = \frac{1}{SP} \sum_{i=1}^{SP} \left( \frac{G_i - R_i}{R_i + G_i + B_i} \right)$$

**Equation 3.3**

$R_i$ ,  $G_i$  and  $B_i$  are the values of red, green and blue channels at  $i^{th}$  pixel. SP is the sum of plant pixels in the image.

### 3.3.4.2 Texture Features

The texture features in this study are grey-level co-occurrence matrix (GLCM) driven features. GLCM, first proposed by Haralick (1973), are well-known for extracting second-order statistical texture features that describe the variations in the spatial relationships of pixels in an image. The basic idea of GLCM is to look at the second-order texture relationship between neighboring pixels. It assesses the probability distribution  $P(x, y)$  for the grayscale values of an image, where  $x$  is the grayscale value at any random position in the image and  $y$  is the grayscale value at another position relative to a specific vector distance  $d$  and angular direction  $\theta$  from the first pixel. An element in GLCM is obtained determining the frequency a pixel with intensity value  $I$  appears in a specific spatial relationship to the pixel with the value (Jiang et al., 2018). To calculate the adjacent pixels for GLCM, four directions are mostly considered (i.e.  $[0^\circ, 45^\circ, 90^\circ, 135^\circ]$ )

and one is regarded as the distance between a pair of pixels. Each element  $(i, j)$  in the resulting GLCM is simply the sum of the number of times a pixel with value  $i$  occurred in the given spatial relationship to a pixel with value  $j$  in the input image. To illustrate, Figure 3.2 shows the calculation of the first three values in a GLCM matrix. The pixel  $s(1,1)$  has an output 1 in the GLCM because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1. Also, the horizontal pixels  $(1,2)$  in the image outputs a value of 2 because there are two instance of two horizontally adjacent pixels with the values of 1 and 2.

In this study, three components of the GLCM texture features, namely angular second moment (ASM), contrast and dissimilarity were used.

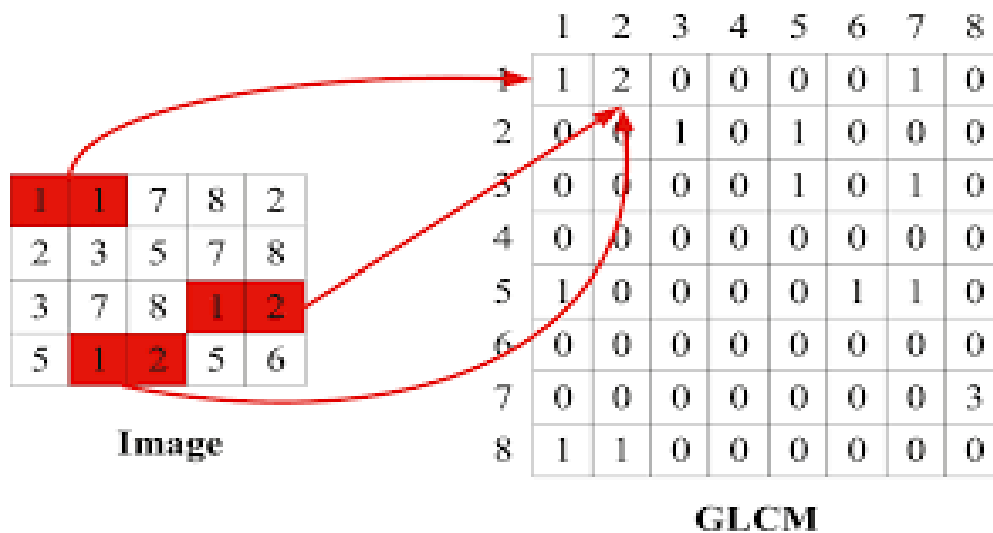


Figure 3-2 Illustration of example of Gray Level Co-occurrence Matrix (GLCM) Calculations form an image matrix. The image is represented by a grid of pixel values, with each number representing a gray-level intensity ranging from 1 to 8. The GLCM matrix is constructed by counting the frequency of a specific pixel intensity value is next to a pixel with a different intensity value (e.g., 2). This sample appears to depict a 0-degree angle (horizontal adjacency). Each GLCM element represents the number of times a given intensity pair (for example, 1 and 2) occurs in the specified direction.

*Angular second moment (ASM):* The ASM measures the uniformity or evenness of pixel pairs in the GLCM of an image and is calculated from the sum of the squares of all element values in the GLCM. The higher the ASM, the more uniform the grayscale in the image and vice versa. This means that with a high ASM, the intensity values in the image are more uniform and less textured. It is expressed mathematically as:

$$ASM = \sum_{p,q=0}^{n-1} Y_{p,q}^2 \quad \text{Equation 3.4}$$

Where  $Y_{p,q}$  is the element values in the GLCM,  $n$  is the previous pixel levels.

*Contrast*: This measures the local variations in an image reflecting image clarity and the depth of texture. It accounts for the highest and lowest variation values of the adjacent set of pixels. The higher the contrast, the greater the difference in intensity between neighbouring pixels, indicating a coarser appearance in the image. On the other hand, a low-contrast image presents GLCM concentration around the principal and outputs low spatial frequencies. The farther a GLCM element is from the diagonal element value, the greater the contrast (Hall-Beyer, 2017). Contrast is given by the expression as:

$$Contrast = \sum_{p,q=0}^{n-1} Y_{p,q} (p-q)^2 \quad \text{Equation 3.5}$$

Where  $Y_{p,q}$  is the matrix value at position  $(p,q)$  in the output GLCM matrix,  $n$  is the grey levels in the output image

*Dissimilarity (DISS)*: This feature measures the absolute variation in intensity of grey-level pairs in an image. It is similar to contrast except that the element weighting is increased linearly in dissimilarity, while in contrast, the weighting is increased exponentially (Park and Guldman, 2020). High dissimilarity value shows great heterogeneity in texture and vice versa. DISS is given by equation (6) as:

$$DISS = \sum_{p,q=0}^{n-1} Y_{p,q} (|p-q|) \quad \text{Equation 3.6}$$

### 3.3.4.3 Morphological Features

Morphological changes in plants play essential roles in tracking stress development. In this study, three morphology features of plants including plant aspect ratio, plant area density and convex hull area were studied.

*Convex hull area (HULL):*. This is a valuable metric that quantifies the spatial coverage of plant canopy. It is the smallest convex polygon (area) that encloses the plant canopy when projected in 2D space (Pound et al., 2013). Analysis of the convex hull area of plants enables researchers to characterize and quantify plant growth. It can be used to assess response of plants to environmental conditions or stress.

*Plant Aspect Ratio (PAR):* This is the measure of the relationship between the height and diameter of the top canopy of plants. It is obtained from the side view's bounding box and the top view's minimum enclosing circle. It provides information on plant canopy architecture which has significant impact assessment on plant performance, adaptation and yield . It is given by the expression in equation (3.7):

$$PAR = \frac{\text{Plant height (side view)}}{\text{Diameter of enclosing circle(top view)}}$$

**Equation 3.7**

*Plant Area Density(PAD):* This is the total plant pixels to the convex hull area of the side view image. It provides essential information on plant biomass within a specific area and can be used to access plant health, yield and stress on plants. It is given equation 3.8 as:

$$PAD = \frac{\text{Total plant pixels (side view)}}{\text{Convex hull area (side view)}}$$

**Equation 3.8**

#### **3.3.4.4 Local Image Descriptors**

In image processing, local image descriptors are techniques that retrieve detailed information from specific patches or areas in an image. Unlike global descriptors, local descriptors concentrate on smaller regions enabling the analysis of texture, shape and other local features. Some common local image descriptors include Scale Invariant Feature Transform (SIFT) (Lowe, 2004), Speeded- Up Robust Features (SURF) (Bay et al., 2006), Local Binary Patterns (LBP) (Ojala et al, 2002) and Histogram of Oriented gradient (HOG) (Dalal & Triggs (2005). In this work, we used HOG and SIFT features for nitrogen and drought stress analysis.

*HOG features* This is a feature descriptor based on the frequency of gradient orientation in a local neighbourhood of an image (Dalal & Triggs, 2005). It captures the orientation and magnitude of gradients in the image. The computation of HOG features starts by generating a non-overlapping square grid of size  $n \times n$ , resulting in  $m$  ( $n \times n$ ) grids. The spatial gradient is further computed in the vertical and horizontal directions using a 1-D discrete derivatives mask with  $dx = [-1, 0, 1]$  and  $dy = [-1, 0, 1]^T$  as the horizontal and vertical filtering kernels (Greche et al., 2020). This is followed by the computation of the grid's gradients by combining the magnitude  $M$  and orientation angle  $\theta$  from the image, which is given by the equation:

$$M(x,y) = (h_x^2 + h_y^2)^{1/2}$$

**Equation 3.9**

$$\tan \theta(x,y) = \left(\frac{h_y}{h_x}\right)$$

**Equation 3.10**

where  $h_x$  and  $h_y$  are pixel gradients along the  $x$  and  $y$  directions of the image.

Finally, the edge orientations are divided into  $q$  bins with angles between 0 and 180 degrees. The number of bins and grids are the main parameters that determine the dimensions of the HOG features. It is effectively used in object recognition and has recently been used in image analysis for phenotyping plant nutrient status (Azimi et al., 2021a) and plant canopy cover (Keller et al., 2019).

*SIFT features:* SIFT, proposed by Lowe (2004), is an image descriptor used for image matching and object recognition. It has distinctive key points that are robust to changes in scale, rotation, affine transformation and illumination. The SIFT feature extraction process involves scale space construction, key point localization, orientation assignment and key point descriptor calculation. The first step identifies the location and scale of key points across different scales using the difference of Gaussian (DoG) filters. Local extrema in the DoG pyramid are identified as key points in key point localisation. At this stage, unstable key points are eliminated by discarding low contrast and poorly localized ones to get more accurate extrema locations. The next step is orientation assignment,



where the orientation of each key point is obtained by considering the local image gradient directions within the neighbourhood of the key point. A descriptor is finally generated for each key point, representing the local appearance around the key point. The computation of the key point is based on the image gradient magnitude and orientation of each image sample point, resulting in a 3D histogram of gradient location and orientation. This creates an  $N \times 128$ -dimension image descriptor, where  $N$  is the number of features.

### **3.3.5 Sensitivity analysis**

While the extracted features and plant stress may be related, these features have different effects on the dynamics of stress in plants. In this work, sensitivity analysis was performed to evaluate the relative impact of each feature on the nitrogen deficiency and water stress of plants. Sensitivity analysis considers the interactions of the input variables and output of a model, which helps quantify how changes in the input parameters affect the behaviour of the model (Pianosi et al., 2015). The sensitivity analysis in this work is based on two commonly known approaches.

The first, based on a one-at-a-time (OAT) method, involves systematically holding out one variable at a time while keeping the rest as input to the model and repeating for each of the remaining variables in the same way. In this case, the random shuffling of a single independent variable causes an increase or decrease in the score of the model (Breiman, 2001). This breaks the relationship between the dependent and independent variables, indicating the dependency of the model on the sensitive features. To perform the OAT sensitivity analysis, a random forest (RF) classifier was trained using the entire set of features. Gridsearch cross validation (CV) was performed to determine the optimal set of combinations of hyper-parameters for the RF classifier. The number of trees and the number of features for each split were the main hyperparameters considered while the remaining hyperparameters were kept constant. The best parameters provided by the gridsearchCV were used to train the model with different sets of features based on the OAT method for performing classification tasks.

The OAT method entails varying one feature while keeping the others constant and observing the subsequent variations in the output. However, it requires a baseline (for instance, the mean values of the features) to compute the variance, which can affect the

resulting sensitivity index if different baselines are used. The second approach, Morris sensitivity analysis ( Morris, 1991) was performed to address this issue. It is a global sensitivity method that estimates the effects of the input variables on the model output using a sampling-based approach. It analyses the input parameters to estimate elementary effects (EE) distribution. For each input feature  $x$  with a minute change in the model performance while keeping the other features constant, the change in the model output  $Y$  resulting from the small changes called elementary effect (EE) is given by the expression

$$EE_i = \frac{[Y(x_1, x_2, \dots, x_{i-1}, x_i + \Delta, \dots, x_k) - Y(x_1, x_2, \dots, x_k)]}{\Delta} \quad \text{Equation 3.11}$$

The EE shows the relationship between the input and output of the model (Morandage et al., 2019) measuring the local variation of the model response  $Y$  in response to changes in the model parameters. In the Morris sensitivity method, two sensitivity measures, the absolute values of the mean ( $\mu^*$ ) and the standard deviation ( $\sigma$ ) of the elementary effects, are computed to indicate the sensitivity and overall influence of the input parameters on the output. A high standard deviation of the input variable indicates a significant influence on the output of the model and vice versa.

For both methods, three sets of analyses were done: nutrient deficiency sensitivity (treatments: WWHN, WWLN), drought stress sensitivity (treatments: WWHN, DSHN), and combined stress sensitivity (all treatments) analysis.

### 3.3.6 Model development

This study proposed a deep learning model that is trained on raw images as well as combined features from the raw and extracted handcrafted features. The performance was compared to classical machine learning models and state-of-the-art deep learning models. The details of the models used in this work are provided in the following sub-sections.

#### 3.3.6.1 Classical ML methods

Classical ML models can be effectively utilized in various plant phenotyping tasks by learning and interpreting patterns in meaningful extracted features for classification and regression of the phenotyping traits under study. Three classical ML algorithms were

trained with the extracted features, including support vector machine (SVM), decision tree (DT) and random forest (RF).

*SVM model:* SVM classifier is a non-probabilistic binary classifier, which can also work as a multi-class classifier. It has different kernel functions, such as linear, polynomial, radial basis, or sigmoid functions, to process nonlinear data effectively based on structural risk minimization. SVM works by separating binary or multiclass feature space using a hyperplane (Melgani and Bruzzone, 2004).

The optimal hyperplane is constructed such that a large distance to any class's nearest training data point is made, which results in a good separation between the various classes. SVM aims to maximize the margin between classes with the support vectors being the data points closest to the hyperplane. The hyperplane is generated based on the equation 3.12:

$$W^T X + b = 0$$

**Equation 3.12**

Where  $X$  is the feature vector used as input for the model,  $W$  is the weight vector and  $b$  is the bias (Burges., 1998)

*Decision Tree (DT):* Decision trees are non-parametric supervised machine learning algorithms that can perform classification, regression and even multiclass tasks. DT works by building a decision tree of a dependent variable as a function of several independent variables. It has a flowchart-like tree structure that includes a root node, branches, nodes and leaves. The roots initiate the process and the leaves make the final decision. In this work, the DT model was built by first assigning all variables to the root node in a top-down manner, which grows by minimizing an error metric of the related nodes based on a specific splitting rule (Azimi et al., 2021a).

*Random Forest (RF):* RF is a powerful ML tool that operates as an ensemble combining the output of several decision tree models. It is widely used for classification problems due to its high performance on high dimensional data with small training samples as well as high computation efficiency (Rodriguez-Galiano et al., 2012). The trees are created by

drawing a subset of training samples through a bagging or voting approach. Each tree in the RF splits out a class prediction and the most voted class becomes the RF model's prediction. We tuned the two main parameters of RF: the number of decision trees and the maximum number of features to optimize the model for high performance.

### 3.3.6.2 Deep learning methods

#### 3.3.6.2.1 The proposed CNN models.

The proposed CNN model is a 13-layer architecture with feature extraction and classification capabilities. It consists of convolutional, batch normalization, pooling, dropout and fully connected SoftMax layers. The model takes in a 70x70x3 image as input to the convolutional layers. These layers have a set of 3x3 kernels containing multiple learnable weights that operate on the dot product to learn their patterns and trends in the input image. The convolutional layer simultaneously learns several different features from the input image. Each convolutional layer is followed by a batch normalization and a max pooling layer of kernel size (2x2) with a stride of 2 to down-sample the spatial dimensions of the feature maps. The operation of the max pooling layer is given by the equation:

$$Pool_{i,j,k} = \max_{m,n} X_{i*s_x+m, j*s_y+n, k}, \text{ for } 1 \leq k, 1 \leq m$$

**Equation 3.13**

Where X is the input, (m, n) are the output indices, k is the channel index and  $s_x$  and  $s_y$  are the stride values in the horizontal and vertical directions, respectively. The filter of size  $f_x$  defines the pooling window and  $f_y$  centred at the output index.

The batch normalization layers lie between the different convolutional layers and work by applying a transformation technique to standardise the inputs in the range of either [0,1], [-1,1], or to mean=0 and variance =1(whitening). Batch normalization (Bn) is mathematically given as

$$Bn = \frac{y_i - \mu_\beta}{\sqrt{\varphi_\beta^2 + \epsilon}}$$

**Equation 3.14**

Where  $y_i$  is the input,  $\mu_\beta$  and  $\varphi_\beta$  are the mean and the variance over each input channel and over a batch.  $\epsilon$  is added to improve the numerical stability of small batch variance. Batch normalization in CNN produces a faster and more stable network.

Furthermore, the model has an activation function that updates the weights of the input features, allowing for a nonlinear relationship between the input and final output features. This layer uses the rectified linear unit (ReLU), which minimizes the problem of gradient vanishing in CNN models based on the equation.

$$\text{ReLU}(x) = \max(0, x) \quad \text{Equation 3.15}$$

The ReLU function returns 0 for any negative input; however, for any positive value  $x$ , it returns that value. Supplementary information Table C-1 shows the number of parameters, output dimensionality, and layer types of the proposed CNN structure.

#### 3.3.6.2.2 State of the art deep learning models

##### *ResNet and DenseNet models*

The performance of the proposed model was compared with existing deep learning architectures such as DenseNet (G. Huang et al., 2017) and ResNet18 (Penatti et al., 2015) as they show high reliability among reported deep learning models for plant phenotyping analysis.

ResNet (Residual network) is a CNN model consisting of stacked residual blocks, each concatenated by three convolutions of 1x1, 3x3 and 1x1 kernels. The residual blocks converged by average pooling and were classified using the SoftMax layers. ResNet solves the challenge of vanishing gradient associated with deeper neural networks using a skip connection technique (He et al., 2016). This is done by connecting activations of a layer to further layers by skipping some layers in between.

DenseNet is a form of CNN model consisting of dense blocks of several convolutional layers. A transition layer follows the dense block, linking one dense block to another. The convolutional layers in the dense block involve three main units: batch normalization, activation function and convolutional layers. In this study, we used a dense net-12

network, which typically has five convolutional and pooling layers, three transition layers (6,12,24), one SoftMax layer (16) and two dense blocks (1x1 and 3x3 conv). Each layer receives feature maps and additional input information from the previous layer and performs concatenation to enable the next layer to receive the cumulative knowledge of all the previous layers. DenseNet was used in this work because it improves network performance and does not require deepening (e.g. ResNet) or widening (GoogleNet) the network architecture (Z. Xu et al., 2020).

#### *3.3.6.2.3 Combination of CNN features and hand-crafted features*

Like other traditional CNN models, the proposed model automatically extracts features and results in better performance than conventional ML models. However, it is difficult to understand what features are learned (low interpretability). From other studies, some global features, such as SIFT and HOG, contribute effectively to the performance of the conventional ML models. However, some CNN architectures cannot learn these global features (Lee et al., 2017), suggesting the need to fuse the hand-crafted features to the proposed CNN model to improve its performance and offer better generalization. The hand-crafted features were merged with the CNN architecture by extending the last two fully connected layers with additional neurons and reducing the weights to the given class probabilities (Bogacsovics et al., 2022). First, the hand-crafted features (Hf) were standardized to have scale measures equal to the CNN features (Cf). Next, we concatenated these two features (CNN and hand-crafted features) and passed them to the two dense layers. The SoftMax layer was finally used to obtain the predictive class probabilities. Figure 3.3 is a diagrammatic representation of the altered CNN model fused with the hand-crafted features. Section A (Figure 3.3) is the complete architecture for the proposed CNN, while the shaded area is the CNN feature extraction component. Also, the section B is the hand-crafted feature extraction component.

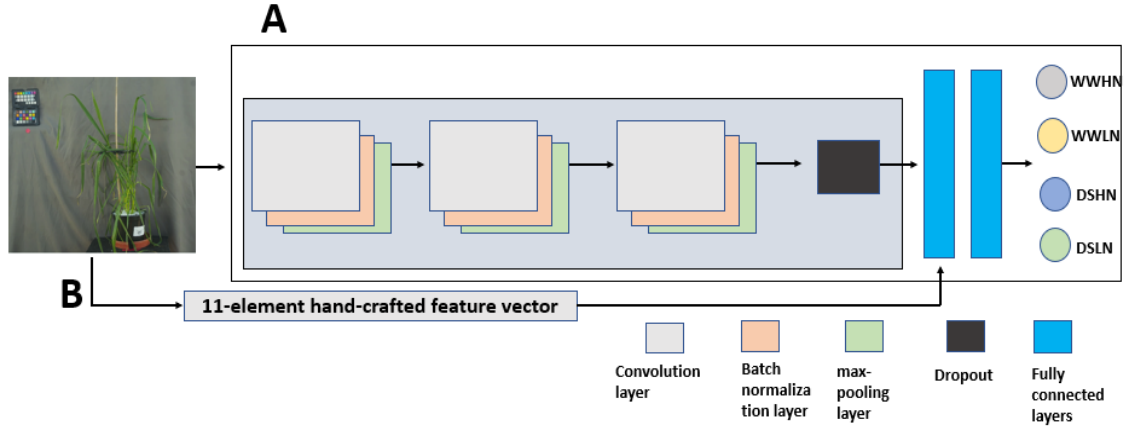


Figure 3-3 A representation of the proposed convolutional neural network (CNN) architecture used for this study. This is made of two sections; (A) the CNN section where plant images serves as input to the model and (B) an 11-element handcrafted feature vector extracted from the plants are integrated with the CNN based features to train the fully connected layers.

### 3.3.7 Statistical analysis and model evaluation

One-way analysis of variance (ANOVA) with post hoc multiple comparisons using Tukey's HSD approach was applied at a 95% confidence level to evaluate the response of plants' different trait expressions to different water stress in plants. This was done to investigate differences between means of the treatments and assess the response of the extracted plant features to water and nitrogen stress.

For all the models in this study, the input dataset was divided into train, validate and test samples for classification and regression models. 80% of this data was used to train and validate each model under a 5-fold cross-validation routine. The remaining 20% was used to test the models to identify the best-performing model. The performance of the classification models was assessed using three metric scores: average accuracy (AA), F-score and Recall as shown in equations 3.16 -3.18, while the regression model performance was evaluated using the coefficient of determination ( $R^2$ ) the mean absolute error (MAE) and root mean square error (RMSE) (equation 3.19-3.21).

$$AA = \frac{TP+TN}{TP + TN + FP + FN}$$

**Equation 3.16**

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{Equation 3.17}$$

$$\text{F-Score} = 2 \frac{P \times R}{P + R} \quad \text{Equation 3.18}$$

Where  $P = \frac{TP}{TP + FP}$ , TP: true positive, TN: True negative, FP: false positive, FN: false negative

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{Equation 3.19}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad \text{Equation 3.20}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad \text{Equation 3.21}$$

where n is the number of data,  $x_i$  is the observed values,  $y_i$  is the predicted values and the bar denotes the mean of the variable. Statistical calculations were performed with the Statistical package in RStudio.

## 3.4 Results

### 3.4.1 Hand-crafted features and physiological measurement

Figure 3.4 depicts a plot of the selected plant physiological traits ( $N$ ,  $P_n$  and  $g_s$ ) for the final day (15 DADS) of the experiment based on nitrogen deficiency and drought stress variations. There was an observable significant difference ( $P < 0.011$ ) between the high and low N treatments at the end of the experiment (Figure 3.4a). Nevertheless, there were no appreciable variations across the treatments for the same N levels with varied water levels. Figure 3.4b demonstrated that drought stress caused substantial differences in stomatal conductance ( $P < 0.008$ ) among the four treatments. While the well-watered treatments had significantly high stomatal conductance—with the WWHN treatment



recording the highest—the drought-stressed treatments had low stomatal conductance with DSHN having the lowest.

Furthermore,  $g_s$  could differentiate treatments of same nitrogen level but varying water levels; WWLN was significantly different from DSLN treatments ( $p < 0.005$ ), and WWHN was significantly different from DSHN treatments ( $p < 0.002$ ). In Figure 3.4c, the control group (WWHN) was observed to have the highest  $P_n$  while the DSLN with  $0.003 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$  had the lowest  $P_n$  on the final day (15 DADS) of the experiment. The results show that  $P_n$  could distinguish between the drought stress and well-watered treatments. The DSHN was observed to be significantly different from the WWHN ( $p < 0.006$ ) and WWLN ( $p < 0.003$ ). Similarly, the DSLN treatment was different from the WWHN ( $p < 0.021$ ) and the WWLN ( $p < 0.002$ ). However, the means of the two drought stress treatments (DSHN and DSLN) were not significantly different. Because the three physiological traits varied between the four treatments, we investigated their correlations with the extracted features. Here, the N,  $P_n$  and  $g_s$  were used as ground truth measurements in subsequent analysis.

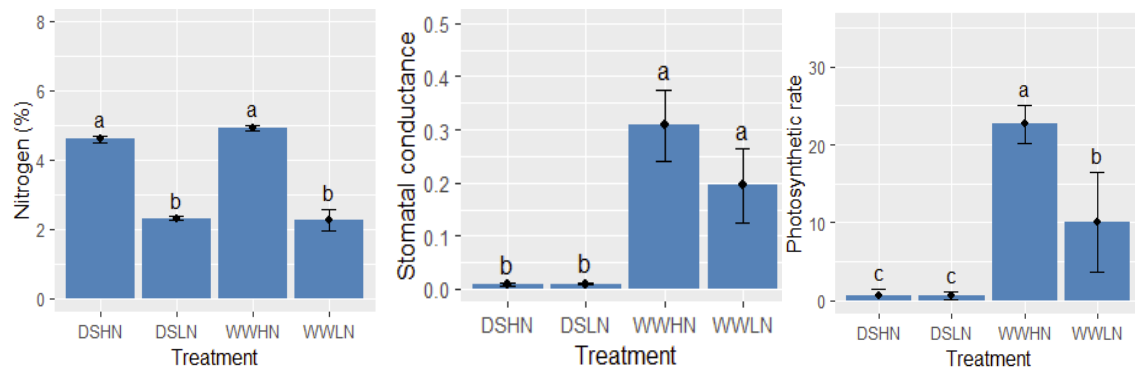
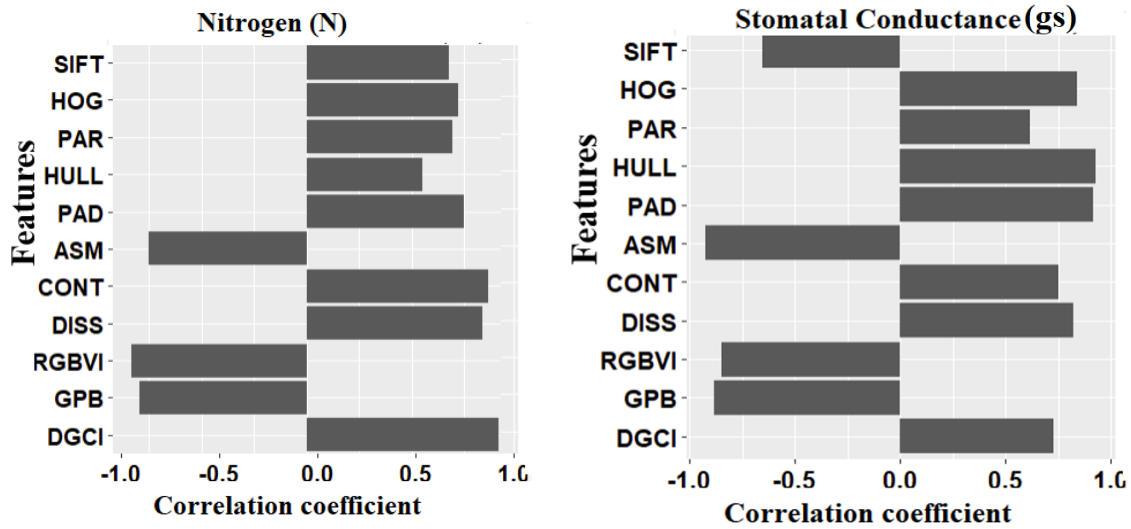


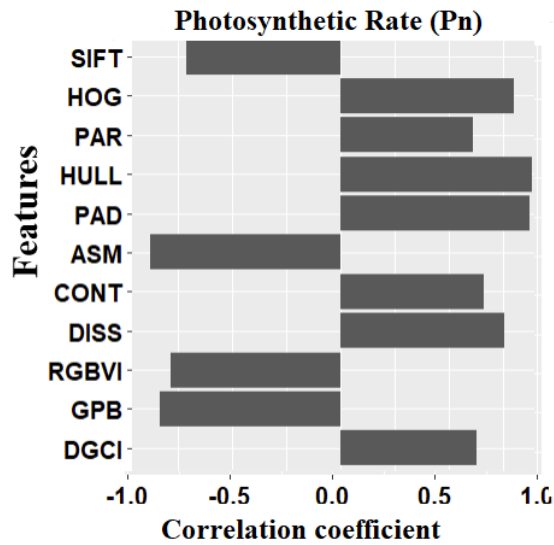
Figure 3-4 Measured physiological traits of plants: (a) nitrogen content, (b) stomatal conductance, (c) photosynthetic net assimilation. Groups represented by bars with the same letter (e.g., a, b, c) are not statistically different from each other at a significance level of 5% ( $p = 0.05$ ). Bars with different letters indicate statistically significant differences between groups.

Figure 3.5 shows the correlation (Pearson) between the eleven hand-crafted features and the physiological traits (N,  $g_s$  and  $P_n$ ) at 15 DADS of the experiment. The hand-crafted features had relatively strong correlations with all three physiological traits ( $> \pm 0.5$ ). All the features were positively correlated to the N trait except the GBP, RGBVI and ASM,

which were negatively correlated with a strong correlation with the three physiological traits. The colour features exhibited the strongest relationship with the N trait with DGCI having the highest  $r$  (0.912). Conversely, the HULL feature had the lowest correlation (0.549) with the N trait. The relationship between the extracted features and the two gas exchange measurements,  $g_s$  and  $p_n$ , were similar. All the features were positively correlated to both traits except the GBP, RGBVI, ASM and SIFT, which were negatively correlated. The colour and texture features exhibited the highest correlation with  $>0.8$   $r$  values. The HULL and ASM features had the best correlation with the  $g_s$  obtaining  $>0.920$   $r$ -value. The  $g_s$  trait was correlated to both the PAR and SIFT features, despite their comparative relatively low  $R$ -values ( $>0.52 < 0.55$ ). On the other hand, the morphology features PAD and HULL had the highest correlation ( $>0.870$ ) with the  $P_n$  trait. The colour and texture feature exhibited variable relationship with the GBP and RGBVI strongly correlating negatively to the  $P_n$  and the DGCI with relatively good positive correlation (0.625) to the  $P_n$ . The three texture features also exhibited good correlation with the  $g_s$  trait with  $r > 0.66$  and  $< 0.87$ .



(a) Correlation for nitrogen content      (b) Correlation for stomatal conductance



(b) Correlation for photosynthesis rate

**Figure 3-5** Pearson correlation coefficients of extracted features with (a) stomatal conductance, (b) nitrogen content and (c) CO<sub>2</sub> net assimilation. The extracted features include ASM: Angular second moment; CONT: contour area; DISS: Dissimilarity; PAR: Plant arial, HULL: Hull area, PAD: Plant area density; SIFT: Scale-invariant feature transform; HOG: Histogram of oriented gradient; GPB: Green Blue; RGBVI: Red, green blue vegetation index; DGCI: dark green colour index.

The relationships between the selected features are revealed by closely examining the principal component analysis (PCA) biplot in Figure 3.6. The PAD, DGCI, CONT, DISS HULL and HOG were negatively correlated and the loading vectors were clustered in the opposite diagonal quadrant. The closeness the ASM, GBP and RGBVI arrows suggests a positive correlation between them, however, these features are negatively correlated to the the HOG, HULL and DISS features. Furthermore, two-colour features (GPB and RGBVI) and texture features (DISS and ASM) contributed the most to the variance in the PC1 and PC2 components. In contrast, the morphological features PAD and PAR had the least influence on the two principal components.

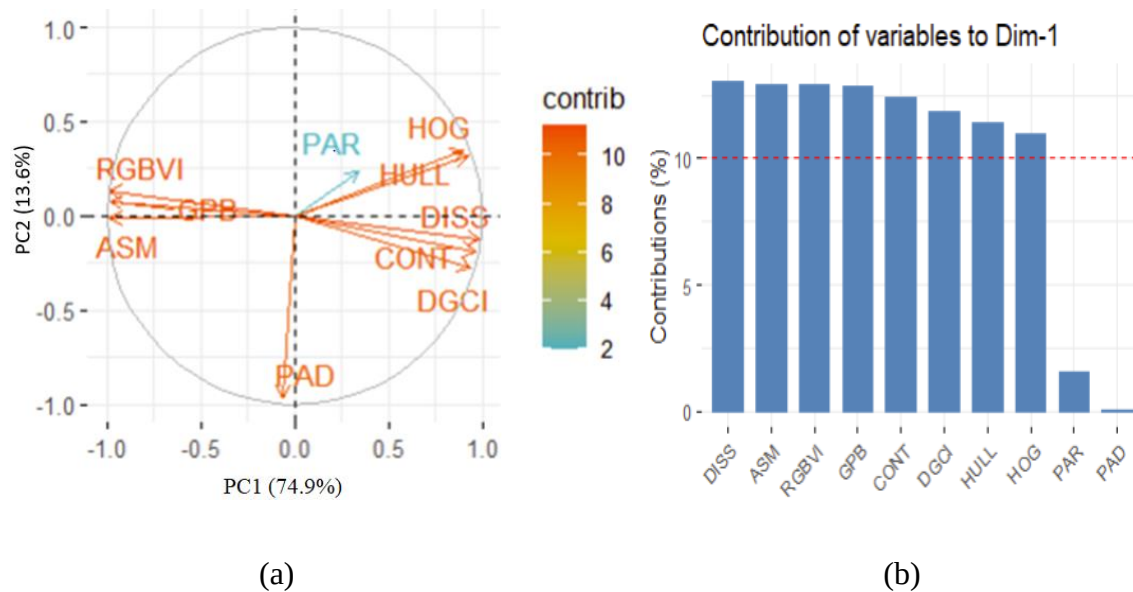


**Figure 3-6** The biplot of the principal component analysis (PCA) of the hand-crafted features (ASM: Angular second moment; CONT: contrast ; DISS: Dissimilarity; PAR: Plant arial ratio, HULL: Hull area, PAD: Plant arial density; SIFT: Scale-invariant feature transform; HOG: Histogram of oriented gradient; GMB: Green Blue; RGBVI: Red, green blue vegetation index; DGCI: dark green colour index) (a) is the biplot showing distribution loading vectors (b) trait importance depicting the contribution of variance to the principal component PC1 vs PC2 accounting for 74.9% of the variability

### 3.4.2 Sensitivity analysis results for the extracted features

Feature sensitivity analysis was performed on nitrogen deficiency, water stress and a combination of nitrogen deficiency and water stress (combined stress). This was based on the OAT and Morris method. The treatments for the nitrogen deficiency sensitivity analysis comprised WWHN and WWLN, while the drought stress comprised the WWHN and DSHN treatments. The combined stress includes all four treatments.

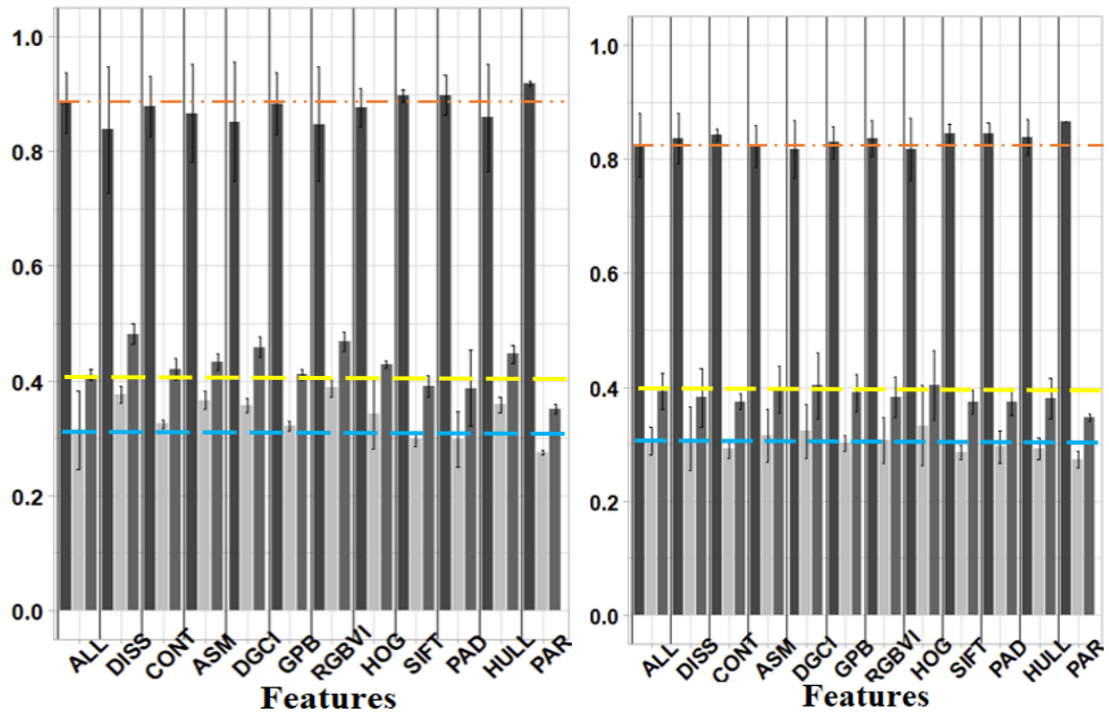
#### 3.4.2.1 OAT-based sensitivity.

Figure 3.7 shows the simulated effect of feature variation on nitrogen deficiency, drought and combined stress using the OAT sensitivity method. The OAT works on the principle that a feature which causes a decrease in the performance of the model when excluded in the training phase is sensitive to the stress under study and vice versa. For nitrogen deficiency analysis, the model with all features as input (ALL) had strong predictive power, achieving an  $R^2$  of 0.884 for nitrogen deficiency (Figure 3.7a). However, the

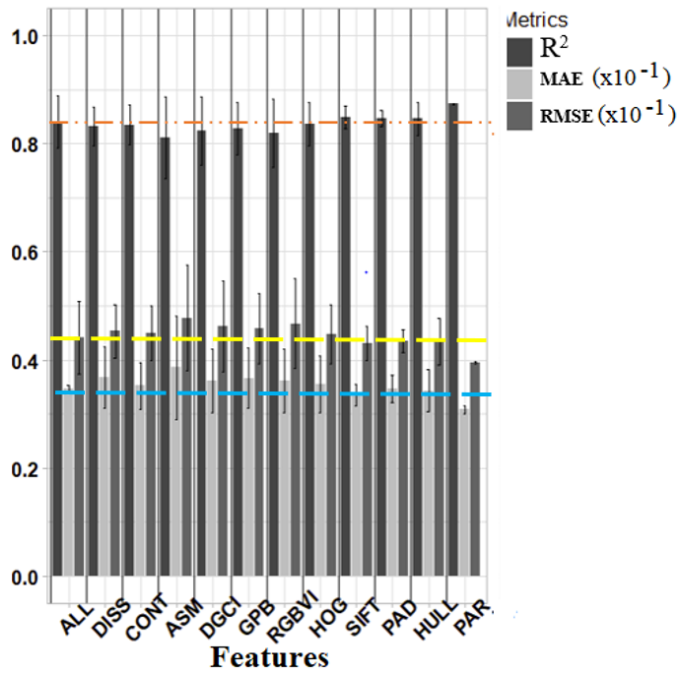
predictive power decreased ( $> 0.03$ -unit drop) when trained without the DISS, ASM, DGCI, RGBVI, and HULL features with  $R^2 = 0.838, 0.851, 0.847$  and  $0.859$ , respectively. This demonstrates that those features had the most bearing on wheat N-level prediction. Although there were subtle changes in the performance of the model when trained without CONT and GPB features (slight decrease in the  $R^2$ , MAE and RMSE of the model), these were not as strong as those mentioned earlier. Similarly, DGCI had the lowest RMSE ( $0.29 \times 10^{-1}$ ) and MAE among the features with the most important bearing on N-level production. This gives more insight into the variability of the error distribution of DGCI for N level treatments, Although PAR feature had the lowest MAE score ( $0.38 \times 10^{-1}$ ), the increase in model performance ( $R^2 = 0.911$ ) when PAR is omitted shows that PAR is relatively not important to the model as the others. It was also observed that the performance of the model increased when trained without the PAD and SIFT features where PAR exclusion achieved the highest  $R^2$  value ( $0.918$ ) suggesting the less effect of these features on the performance of the model.

Results from Figure 3.7(b) demonstrate that the absence of the ASM, and DGCI features caused a decrease in the model's performance for drought stress. The exclusion of the HOG features resulted in the lowest performance of the model (a 1% drop) with  $R^2$  of  $0.817$ , MAE =  $0.332$  and RMSE= $0.402$ . The small number of features causing a decrease in performance indicates the dependency of the model on these features. In contrast, excluding the PAR, HULL, HULL, PAD, SIFT ,GPB and CONT features resulted in an increase in the model performance. The PAR feature with  $R^2=0.836$ , MAE =  $0.272$  and RMSE= $0.346$ , when excluded, resulted in the model having the highest performance. Although the model's performance increased slightly when trained without GPB and RGBVI features, the increase was not significantly different ( $p$ -value =  $0.445$ ) from the performance when trained with all the features.

Lastly, as shown in Figure 3.7c, there was an increase in the  $R^2$  value and a decrease in the MAE and RMSE of the model when PAR was removed. This shows that this feature was redundant in the performance of the model. Also, PAD, ASM and RGBVI features each caused a 7.20, 12.71 and 6.43 unit drop in performance when not used in the modelling of the  $P_n$  traits, indicating a high sensitivity power for analyzing nitrogen or drought stress.



(a) Nitrogen stress sensitivity analysis      (b) drought stress sensitivity analysis



(c) Combined stress sensitivity analysis

Figure 3-7 Sensitivity analysis of the input variables on the three levels of stresses (a. nitrogen stress, b. drought stress, c. combined stress) using RF based on the OAT method. The features of each model are instances where those features were not part of the training dataset, and the ALL is the combined features used for training the model. The short dashed red, yellow and blue lines indicate the baseline for comparing the different models using the  $R^2$ , MAE AND RMSE metrics. DGCi: dark green colour index, GPB: green blue colour Index, RGBVI: red, green blue vegetation index, ASM: angular second moment, CONT: contrast, DISS: dissimilarity. PAR: plant aspect ratio, PAD: plant area density, HULL: convex hull area, SIFT: scale invariant feature transform. HOG: histogram of oriented gradient. All: all the features combined.

### 3.4.2.2 Morris' Method Sensitivity

The Morris sensitivity results were evaluated by comparing the mean and standard deviation of the distribution function of each input feature (Figure 3.8). The mean of the absolute value of the elementary effect gives the overall sensitivity of the *ith* input variable on the output response. The higher the absolute mean value( $\mu^*$ ), the more sensitive the input variable is to the measured stress. A high standard deviation of the elementary effect means that the variable may interact with other variables or have a nonlinear effect on the output (Campolongo & Braddock, 1999). The plot in Figure 3.8a suggests that RGBVI, ASM and DGCi were the most sensitive features for nitrogen deficiency detection, with ASM being the most sensitive ( $\mu^* > 0.06$ ). This agrees with the results obtained on nitrogen deficiency sensitivity using the OAT method where ASM and DGCi were highly sensitivity. The RGBVI and ASM had high standard deviations (0.515 and 0.482 respectively), indicating variability in their sensitivity to nitrogen deficiency and their interactions with other variables. The PAD feature with  $\mu^*$  and  $\sigma < 0.01$ ) had the least effect on the nitrogen deficiency. As shown in Figure 3.8(b), HOG and ASM were identified as the most sensitive parameters to drought stress with  $\mu^* > 0.05$ . The HOG feature's high mean and standard deviation show the high sensitivity and nonlinear interactions this feature has with other features. Compared to the OAT, although the ASM feature was identified as sensitive to drought stress, the HOG showed less sensitivity to drought in the OAT method. Figure 3.8(c) shows that DISS, CONT and DGCi are less sensitive to the combined stress. The nitrogen and drought stress combination were principally driven by the ASM trait, which had the highest  $\mu^*$  score (0.077). However, it had a low  $\sigma$  as compared to RGBVI and HOG features, indicating less variability and interactions with other features. Moreover, since the magnitude of  $\sigma$  was less than the  $\mu^*$  score, the ASM and combined stress can be identified as having an

approximately linear relationship (Menberg et al., 2016). In all the three stress scenarios (nitrogen deficiency, drought stress and combined stress), ASM feature was identified as highly sensitivity using both sensitivity methods indicating the possibility of using the ASM to measure nitrogen and drought stress in wheat.

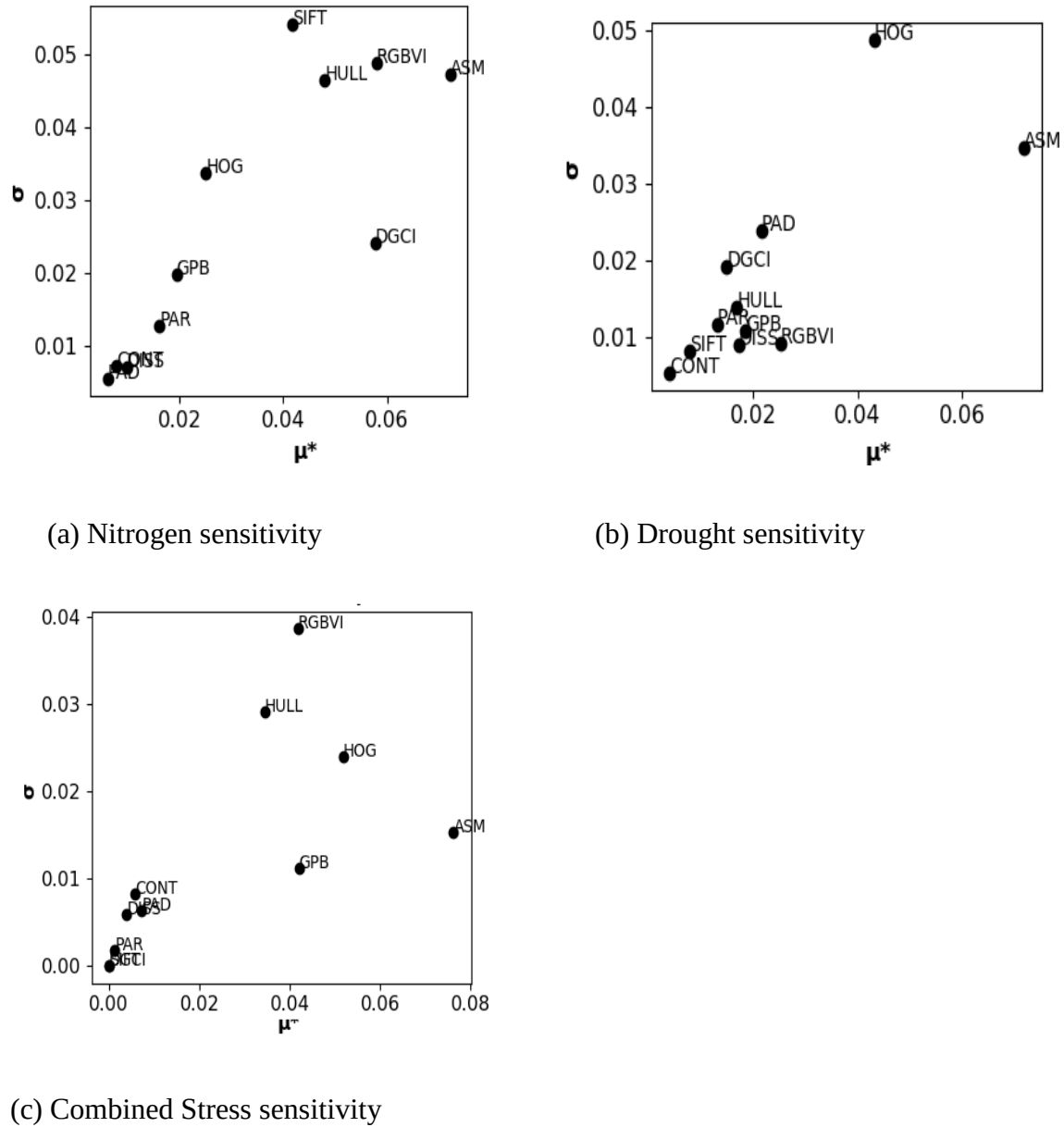


Figure 3-8 Morris based method for sensitivity analysis of plants under different stress conditions: (a) Nitrogen sensitivity, (b) Drought sensitivity, and (c) Combined Stress sensitivity. Each plot uses two variables on its axes:  $\mu^*$  (x-axis): represents the mean sensitivity whiles  $\sigma$  (y-axis) represents the standard deviation or variability in sensitivity. DGC: dark green colour index, GPB: green blue colour Index, RGBVI: red, green blue vegetation index, ASM: angular second moment, CONT: contrast, DIS: dissimilarity. PAR: plant aspect ratio,



PAD: plant area density, HULL: convex hull area, SIFT: scale invariant feature transform. HOG: histogram of oriented gradient

### **3.4.3 Time series analysis of sensitivity for early stress detection.**

Identifying stress before it results in irreparable crop damage is one way to improve crop management systems. This aids in the selection of stress-tolerant crops for high yield. From the sensitivity analysis above, the ASM feature was the most sensitive to nitrogen deficiency and drought stress. As mentioned earlier, a mixed anova analysis was performed to analyze the individual variability, test interactions effects of the groups and timepoints of this data. Supplementary Table C-2 provides a detailed output for a mixed anova analysis with group and timepoint effects. In this case the group include WWHN, WWLN, DSHN and the timepoints are 0,3,6,9,12 and 15 DADS. The intercept (0.833) is statistically significant ( $p < 0.001$ ), with a 95% confidence interval ranging from 0.803 to 0.863. At time point 0, the DSLN group had a significantly lower value (-0.027) as compared to the reference group (WWHN). This effect is substantial, with a p-value = 0, suggesting that DSLN is significantly different from the WWHN. However, there is no significant difference between DSHN and WWHN at timepoint 0 ( $p = 0.767$ ). Similarly, the DSLN and DSHN groups were not significantly different at timepoint 0. It was also observed that the difference between WWLN and WWHN is marginally significant ( $p = 0.05$ ), with WWLN having a slightly lower value. The timepoint was observed not to be statistically significant ( $p = 0.116$ ), indicating that there is a no general significant change in value over time for all the groups. There is an observable significant interaction between DSLN and timepoint ( $p < 0.001$ ), suggesting that the DSLN group shows a decrease in the response variable over time relative to WWHN. Similarly, there is a positive interaction between DSHN and the various time points ( $p < 0.001$ ) where the DSHN group shows a significant increase in value over time as compared to the WWHN. Similarly, WWLN exhibits a substantial positive increase in responsiveness over time compared to WWHN ( $p = 0.002$ ).

Since the timepoint was statistically not significant, a further post-hoc analysis using Tukey's Honest Significant difference (HSD) test for each time point was performed to determine which specific treatments' means are significantly different from each other.

A time series analysis was performed on this feature to assess the early onset of drought stress. As shown in Figure 3.9, there was a clear distinction between treatments of the

same water level with different nitrogen levels at 0DADS. That is, DSHN and DSLN were significantly different from WWHN and WWLN. At 3 DADS, the control group (WWHN) was significantly different from the other treatments, with p-values = 0.015, 0.019 and 0.044 for DSHN, DSLN and WWLN, respectively. At this stage, the ASM feature could differentiate between DSHN and WWHN, DSHN and WWLN, as well as DSLN and WWHN. The treatments were observed to be significantly different at the 9 and 12 DADS. At the end of the experiment (15 DADS) where there were no observable differences between DSHN and DSLN. It was also observed that drought stress caused an increase in ASM when we compare the ASM values at the beginning (0 DADS) and the end of the experiment (15 DADS). At 0 DADS, all the treatments had < 0.84 ASM values. However, the ASM value of the drought stress treatments (DSHN and DSLN) increased with increasing drought stress, where both had >0.90 ASM values at 15 DADS.

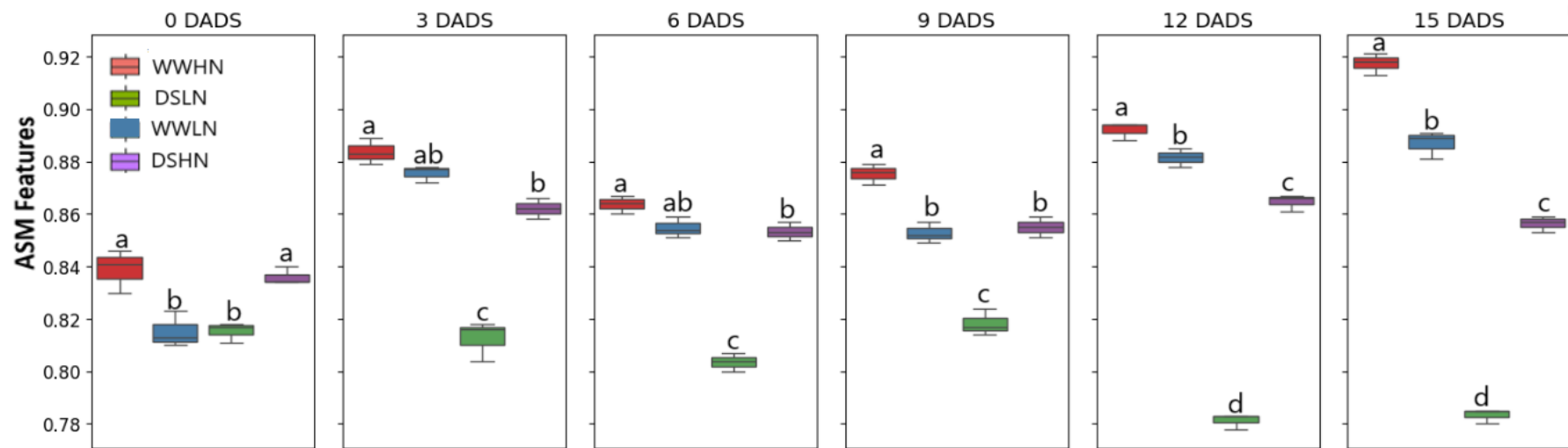


**Figure 3-9** Time series analysis of the ASM feature from 0 DADS to 15 DADS. This was done to understand the dynamic effect of the feature as when plants were subjected to water stress under varying nitrogen conditions. The significant tests are conducted at 3-day and 6-day intervals. Bars followed by the same letter do not differ statistically by Tukey's test at  $p=0.05$

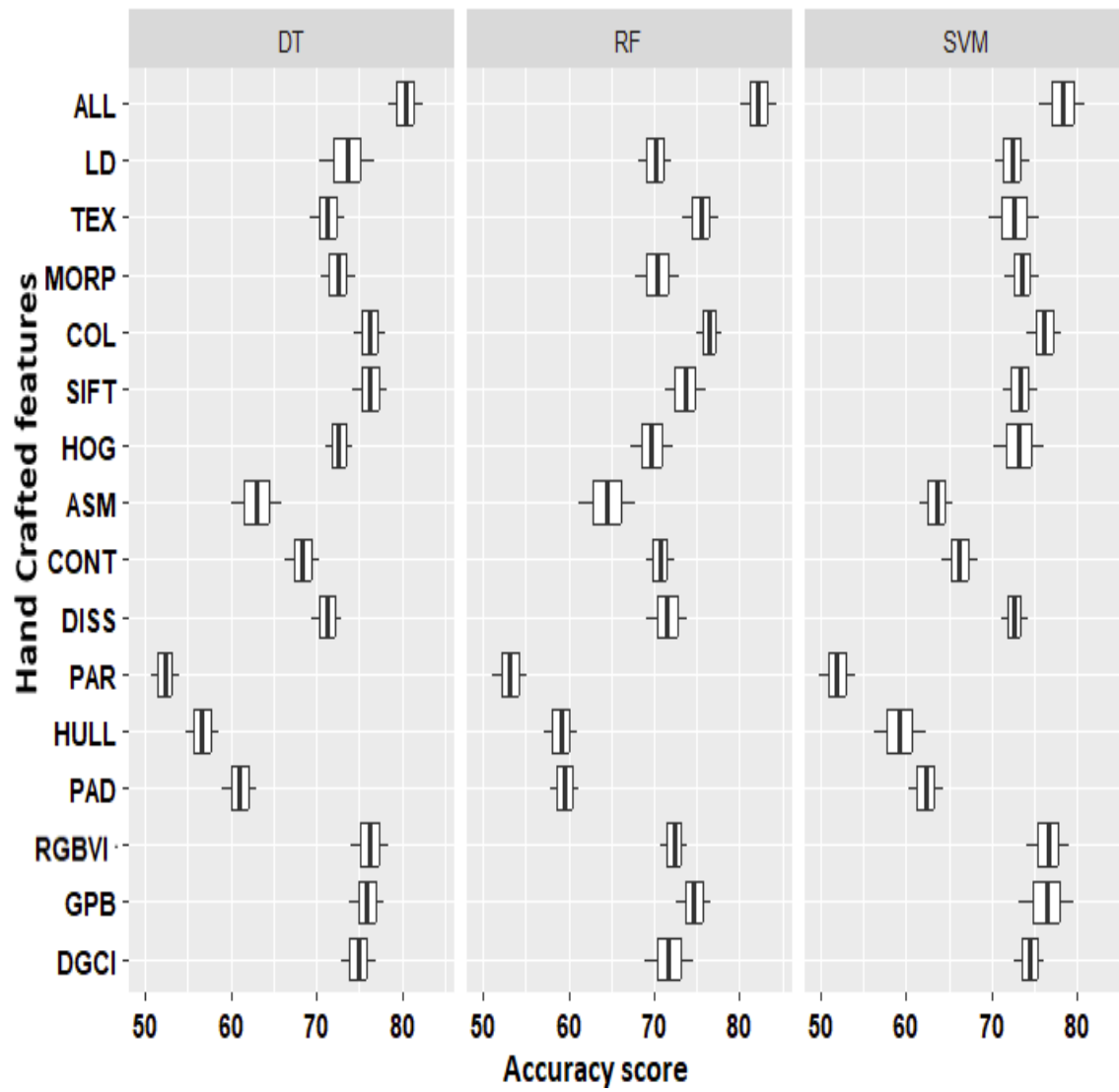
#### **3.4.4 Evaluating the performance of classical ML models.**

In this section, we evaluated the classification performance of the classical ML techniques for both top and side-view extracted features for the classification of different plant stress levels. That is to identify plant drought stress under different nitrogen levels. The performance of these models when trained with single and combined features was also assessed. This was done to analyze the effect of drought stress on single extracted features and the combined features. Figure 3.10 is a box and whisker plot of the AA score distribution, whereas Tables 3.1 and 3.2 show the AA, F-score and Recall-score for the three models based on the individual and combined extracted features. Training the models with each single feature outputs a relatively good performance for both top and side views with accuracy and an F-score between 50-70%. It was generally observed that the side view single features exhibited high classification performance as compared to the top view features, with AA ranging between 61% to 76%, while the top view achieved performance ranging from 51% to 76%. The RGBVI produced the highest classification score (76.5% AA) produced by the SVM model for the top view. On the other hand, the DISS feature was observed to have the best classification performance for the single features on the side view, achieving 76.81%, AA for the side view.

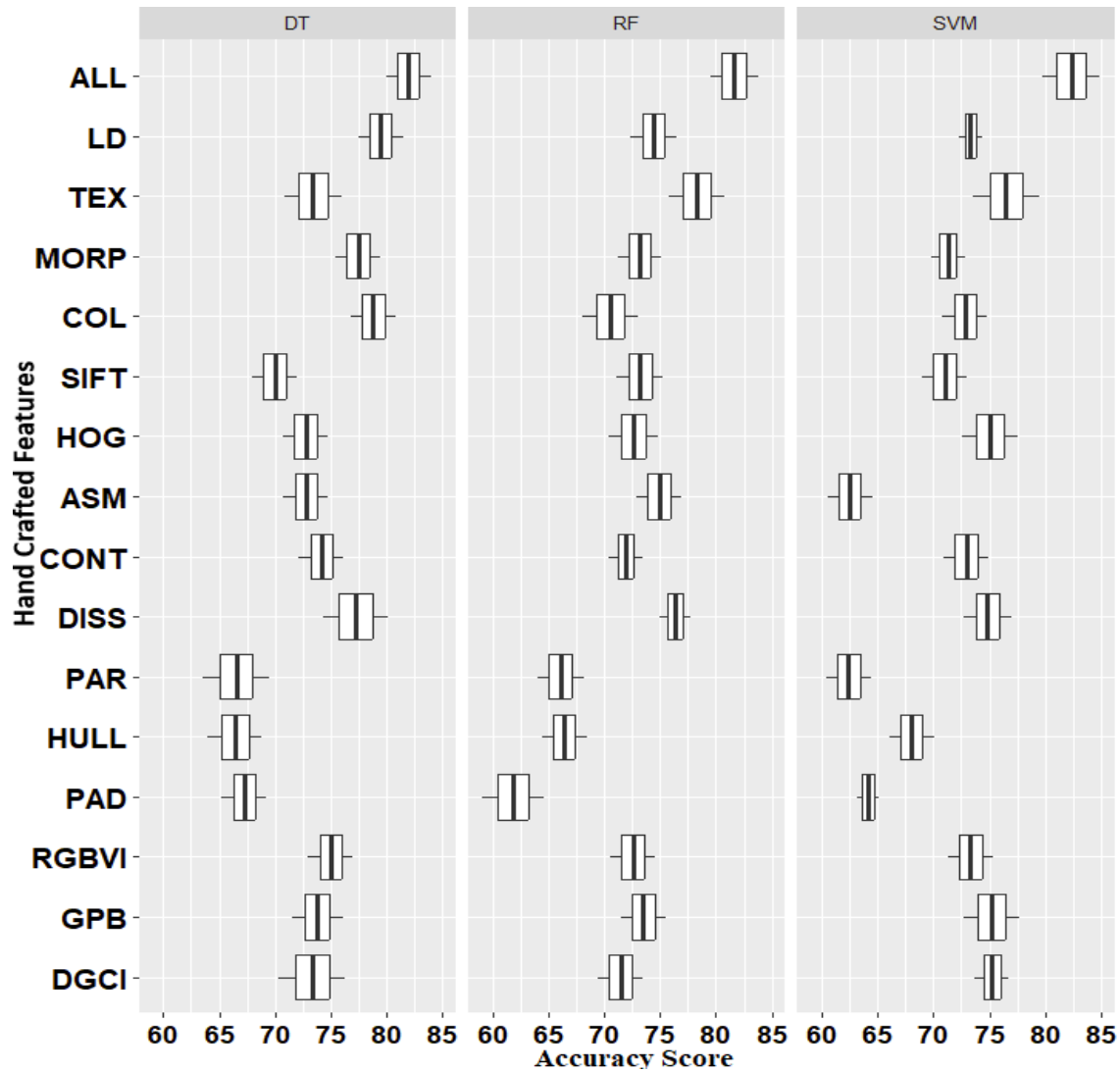
Analyzing the combined features, the local image descriptors (HOG and SIFT) had the best performance for the side view, obtaining 79.46% from the DT model, followed by the colour features with 78.80% AA also produced by the DT model. On the other hand, colour features had the best performance in the top view with 76.43% AA. It should be noted that the side view combined features produced relatively high performance as compared to the top view features. Furthermore, It was observed that while the models achieved a comparable low performance when trained with the individual features, their performances improved when trained with the combined features. However, training the models with all the features combined produced a superior performance to models trained with the individual or combined features. The high accuracy is an indication of good generalization using these features.

Among all the models with the single feature dataset, SVM produced the highest  $R^2$  value (76.55%) using the RGBVI feature dataset for the top view imaging. While the three

models yielded varying outcomes based on the input feature, their overall performances were comparable. Conversely, the RF model had the highest AA (76.81) when trained with the DISS feature dataset. Likewise, the SVM followed with a 75.72  $R^2$  value when trained with the DGCI feature dataset. When trained with the combined features, the DT model exhibited the best performance for the side view images while the RF model also produced the highest AA (76.43%) for the top view images. However, when the models were trained with all the data, the SVM model showed the best performance ( $R^2= 82.20$ ) for both views, although the difference in performances between the different models was relatively not significant. However, this performance is limited by the high computation time. As shown in Table 3-1, SVM had the highest computation time for individual and combined features. Due to the increased number of features, the computation time of the SVM model significantly increased (49.58 s) when trained with all features for both views. This shows that to apply the SVM model in nitrogen or drought stress phenotyping, the trade-off between computational time and model performance needs to be considered (Azimi et al., 2021a). Similar significance results were found for the other model performance metrics, F-score and Recall scale, which clearly showed a linear trend with the AA score.



(a) Top view features



(b) Side view features

Figure 3-10 Plot of model performances (accuracy) for top and side views of plant features. Accuracy range is from 0-100% with models close to 100% indicating high accuracy and close to 0% indicating low accuracy. DGCI: dark green colour index, GPB: green blue colour Index, RGBVI: red, green blue vegetation index, ASM: angular second moment, CONT: contrast, DISS: dissimilarity. PAR: plant aspect ratio, PAD: plant area density, HULL: convex hull area, SIFT: scale invariant feature transform. HOG: histogram of oriented gradient. MORP features comprise all the morphology features (PAD, PAR and HULL) whereas COL represents all the colour features (DGCI, GPB and RGBVI). TEX represents the texture features (CONT, DIS and ASM) whilst LD is the local image descriptors (HOG and SIFT). Finally, the All features represents all the features combined.

|                | Top view |       |       |       |        |       |       |       |       |       |       |       |
|----------------|----------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|-------|
|                | DT       |       |       |       | RF     |       |       |       | SVM   |       |       |       |
|                | AA       | FS    | RC    | T (s) | AA     | FS    | RC    | T (s) | AA    | FS    | RC    | T (s) |
| DGCI           | 74.93    | 74.42 | 73.23 | 0.68  | 71.70  | 70.50 | 71.18 | 3.761 | 74.35 | 73.82 | 73.70 | 29.52 |
| GPB            | 75.90    | 75.17 | 75.71 | 0.79  | 74.65  | 73.81 | 73.52 | 3.88  | 76.35 | 74.71 | 75.92 | 29.28 |
| RGBVI          | 76.25    | 76.60 | 75.25 | 0.64  | 72.30  | 71.34 | 70.73 | 3.80  | 76.55 | 75.11 | 75.33 | 29.35 |
| PAD            | 61.04    | 58.10 | 59.90 | 0.64  | 59.54  | 60.38 | 59.45 | 3.72  | 62.25 | 60.35 | 60.15 | 28.87 |
| HULL           | 56.64    | 56.80 | 56.25 | 0.75  | 59.10  | 57.71 | 59.80 | 4.26  | 59.22 | 55.62 | 56.82 | 27.71 |
| PAR            | 52.32    | 51.91 | 51.12 | 0.76  | 53.13  | 51.45 | 50.51 | 3.36  | 51.88 | 47.20 | 47.53 | 29.26 |
| DISS           | 71.20    | 69.43 | 70.26 | 0.77  | 71.52  | 69.21 | 70.73 | 4.26  | 72.60 | 69.43 | 70.28 | 37.73 |
| CONT           | 68.40    | 68.12 | 68.62 | 0.73  | 70.68  | 69.70 | 70.14 | 3.88  | 66.20 | 64.75 | 64.10 | 32.08 |
| ASM            | 63.00    | 61.52 | 61.50 | 1.55  | 64.53  | 62.25 | 62.82 | 4.10  | 63.50 | 61.10 | 61.28 | 31.25 |
| HOG            | 72.56    | 72.80 | 72.21 | 1.44  | 69.67  | 70.00 | 69.25 | 2.11  | 73.09 | 73.14 | 73.50 | 30.40 |
| SIFT           | 76.32    | 75.73 | 75.32 | 0.98  | 73.625 | 73.19 | 73.24 | 2.00  | 73.25 | 72.29 | 71.38 | 28.45 |
| <b>Colour</b>  | 76.11    | 75.90 | 76.10 | 1.09  | 76.43  | 75.50 | 74.42 | 4.51  | 76.11 | 76.81 | 76.12 | 40.84 |
| <b>Morph</b>   | 72.50    | 71.34 | 71.25 | 0.83  | 70.33  | 71.15 | 70.51 | 4.87  | 73.50 | 73.10 | 72.96 | 33.11 |
| <b>Texture</b> | 71.30    | 72.00 | 71.44 | 1.37  | 75.43  | 72.88 | 72.60 | 3.95  | 72.63 | 72.52 | 70.44 | 33.67 |
| <b>LD</b>      | 73.53    | 72.10 | 72.90 | 1.25  | 70.10  | 70.21 | 70.80 | 3.99  | 75.25 | 73.25 | 74.75 | 38.66 |
| <b>All</b>     | 80.00    | 79.53 | 79.38 | 1.01  | 82.20  | 82.35 | 81.63 | 5.10  | 78.26 | 78.17 | 78.10 | 42.97 |

**Table 3-1 A**

Performance of the three classical Machine Learning techniques for stress classification (average accuracy (AA), F-score (FS), Recall score (RS), Time (T))



| Side View         |       |       |       |       |       |       |       |       |       |       |       |       |
|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                   | DT    |       |       |       | RF    |       |       |       | SVM   |       |       |       |
| Features          | AA    | FS    | RC    | T (s) | AA    | FS    | RC    | T (s) | AA    | FS    | RC    | T (s) |
| DGCI              | 73.30 | 74.47 | 73.27 | 0.55  | 71.46 | 71.80 | 70.70 | 2.41  | 75.72 | 74.51 | 74.73 | 33.45 |
| GPB               | 73.78 | 73.15 | 72.53 | 0.72  | 73.51 | 74.70 | 73.91 | 2.57  | 75.21 | 73.82 | 73.53 | 38.28 |
| RGBVI             | 75.00 | 73.60 | 74.29 | 0.69  | 72.56 | 71.14 | 71.73 | 2.40  | 73.34 | 73.35 | 73.72 | 30.71 |
| PAD               | 67.26 | 66.13 | 66.91 | 0.60  | 61.82 | 60.38 | 61.45 | 2.72  | 64.15 | 64.34 | 63.21 | 32.66 |
| HULL              | 66.42 | 66.87 | 65.90 | 0.78  | 66.42 | 65.61 | 62.86 | 2.55  | 68.05 | 67.72 | 68.90 | 28.26 |
| PAR               | 66.51 | 64.94 | 65.93 | 0.74  | 66.05 | 67.26 | 66.42 | 2.44  | 62.41 | 61.97 | 60.70 | 30.11 |
| DISS              | 72.22 | 71.49 | 71.20 | 0.76  | 76.81 | 75.47 | 76.22 | 3.70  | 74.83 | 73.76 | 73.38 | 34.52 |
| CONT              | 74.15 | 73.10 | 73.68 | 0.61  | 71.92 | 70.40 | 71.66 | 2.17  | 72.93 | 72.43 | 72.36 | 33.22 |
| ASM               | 72.79 | 70.54 | 69.50 | 0.81  | 74.90 | 74.11 | 74.57 | 3.88  | 62.52 | 62.23 | 62.32 | 34.29 |
| HOG               | 72.75 | 72.82 | 72.25 | 0.88  | 72.62 | 72.19 | 72.50 | 3.45  | 75.03 | 75.16 | 75.33 | 32.28 |
| SIFT              | 69.98 | 68.72 | 68.33 | 0.98  | 73.21 | 72.52 | 72.31 | 2.20  | 71.00 | 70.14 | 69.21 | 32.00 |
| <b>Colour</b>     | 78.80 | 76.95 | 76.18 | 1.44  | 70.34 | 70.82 | 70.62 | 4.35  | 72.84 | 70.52 | 71.45 | 44.65 |
| <b>Morphology</b> | 77.43 | 75.38 | 75.22 | 1.64  | 73.10 | 72.26 | 72.95 | 3.92  | 71.32 | 71.63 | 70.46 | 42.28 |
| <b>Texture</b>    | 73.41 | 72.63 | 73.49 | 1.75  | 78.55 | 75.58 | 76.40 | 3.71  | 76.50 | 74.67 | 73.25 | 43.71 |
| <b>LD</b>         | 79.46 | 77.10 | 77.94 | 1.68  | 74.32 | 74.20 | 73.72 | 2.46  | 73.33 | 72.66 | 73.13 | 43.45 |
| <b>All</b>        | 81.94 | 81.53 | 80.31 | 2.45  | 81.62 | 79.65 | 80.11 | 4.98  | 82.35 | 81.16 | 81.61 | 49.58 |

**Table 3-2**

Performance of the three classical Machine Learning techniques for stress classification. (average accuracy (AA), F-score (FS), Recall score (RS), Time (T))

### **3.4.5 Evaluating the performance of the proposed CNN models.**

The proposed CNN model was trained with segmented and non-segmented plant images, and the performance was compared to state-of-the-art ResNet and DenseNet models. Table 3-3 shows the performance of the proposed CNN, ResNet and DenseNet models using the AA, F-score and Recall scores as metrics. All the models had > 90% accuracy for both segmented and non-segmented images. However, the models trained with the segmented images slightly outperformed those with non-segmented images. This agrees with Azimi et al. (2021b) who propose the use of segmented images in CNN modelling for small training samples. For the segmented images, the DenseNet had the best performance with 97.83, 96.77 and 96.30 for the AA F-score and Recall score, respectively. The proposed CNN model followed with a 97.50% AA score outperforming the ResNet model with 0.08% more AA. Again, the DenseNet model showed higher performance than the proposed CNN and ResNet models when trained with the non-segmented images achieving 95.77%, 95.43% and 95.02% AA, F-score, and Recall score.

A confusion matrix visualization method was plotted to compare the performances of the three CNN models on the test dataset (Figure 3.11). The plots show that the Densenet model had the highest correct recognition rate for WWHN and WHLN treatments with misclassification errors of 5.8% and 1.5% for the non-segmentated images. Although the proposed CNN model had the highest classification rate for DSLN (98.07), the comparatively low classification accuracy for the other treatments affected the overall performance of the model. Interestingly, the proposed model had the highest performance in predicting the DSLN treatment, although the Desnet and ResNet models also had a relatively good performance comparably. It should be noted that DSLN treatment could be most difficult to distinguish because it is made of two stresses (drought and nitrogen deficiency), which could have multiple stress characteristics.

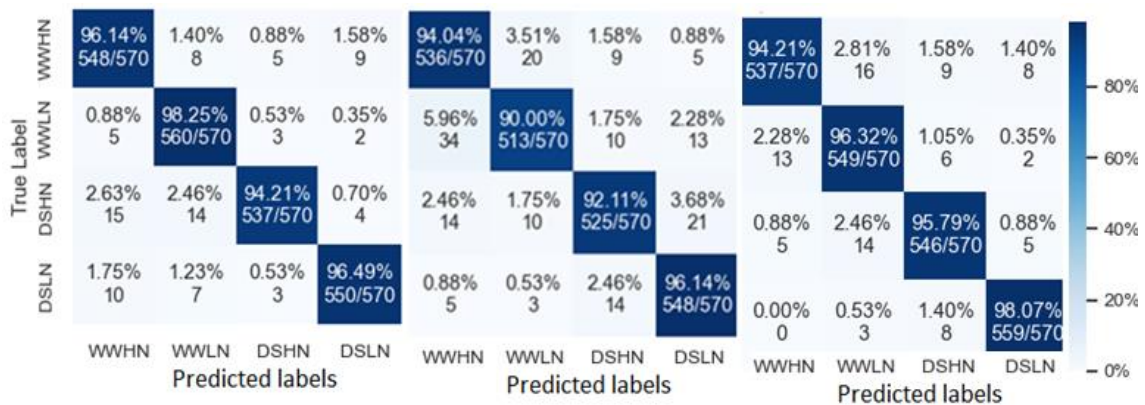
Generally, there was an improvement in classification for the three models when trained with the segmented images. However, the differences weren't significant. The Desnesnet model was observed to have the highest correct recognition rate for all four treatments, recording the highest classification for WWHN (99.30%) and the lowest for DSHN (98.07%). The proposed CNN followed achieving the highest classification of DSLN

(98.42) and lowest for WWHN (97.72%). The proposed model had some challenges identifying differentiating between DSHN and DSLN as 1.58% of DSHN were identified as DSLN and 1.05% DSLN were identified as DSHN

To analyse the stability of the proposed model, we compared the accuracy and loss rate curves under variable number of iterations for the training and validation set. From supplementary information (Figure C-2), at 40 epochs, the proposed model begins to converge and remains relatively stable even after 100 iterations. At this point, the model accuracy and loss stabilized, showing the ideal classification performance. The proposed model achieved a high performance and required fewer training parameters and a shorter training time than the DenseNet and ResNet models, making it ideal for drought and nutrient stress identification.

**Table 3-3** Performance analysis for proposed CNN, ResNet-18 and DenseNet

| Model        | Data type       | Metrics |         |        |         |
|--------------|-----------------|---------|---------|--------|---------|
|              |                 | AA      | F-Score | Recall | Time(s) |
| Proposed CNN | Segmented       | 97.50   | 96.10   | 97.30  | 35.29   |
|              | with background | 93.11   | 93.72   | 93.34  | 44.11   |
| ResNet       | Segmented       | 97.31   | 97.00   | 96.81  | 117.21  |
|              | with background | 93.16   | 92.62   | 93.19  | 123.55  |
| DenseNet     | Segmented       | 97.83   | 96.77   | 96.30  | 88.41   |
|              | with background | 95.77   | 95.43   | 95.23  | 110.62  |



(i) Densenet model (ii) ResNet model (iii) Proposed CNN model  
(a) Confusion matrices for non-segmented images

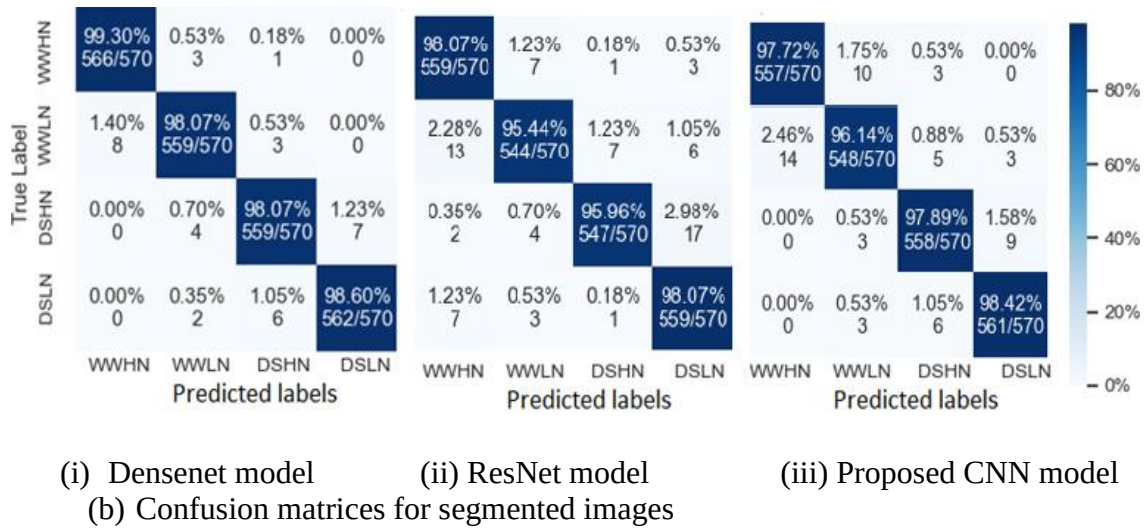


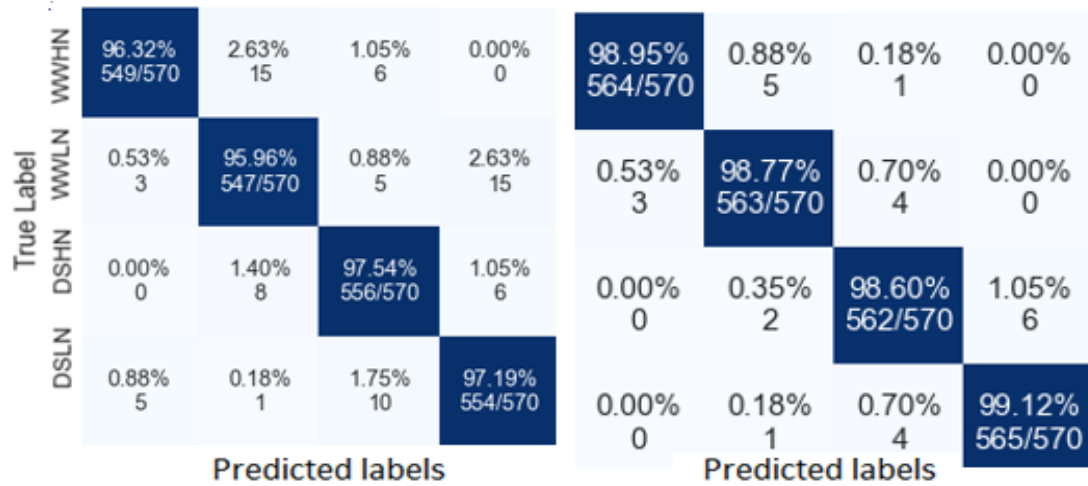
**Figure 3-11** Confusion matrices depicting classification results of the three models (the proposed CNN, ResNet and DenseNet). This shows the accuracy of the models in classifying plants based on their stress levels. The true class (ground truth) are on the y-axis, while the predicted class are on the x-axis. Each matrix shows the percentage and count of correctly and incorrectly classified samples for four different classes: WWHN WWLN, DSHN and DSLN. (a) shows results for non-segmented plant images while (b) shows results for segmented plant images

### 3.4.6 Performance of the proposed model with hand-crafted features

The proposed CNN model was fused with the hand-crafted features to understand the impact of the features on identifying drought-stressed plants under varying nitrogen levels. The model was trained on both the segmented and non-segmented image datasets. Table 3-4 presents the results of the performance of the proposed CNN model when fused with the hand-crafted features. Compared to the standalone CNN models for both datasets, a slight improvement in the model's performance was observed when the hand-crafted features were fused to the model, achieving a 98.90% and 96.71% (This is a 1.4% and 3.6% increment in AA scores for the segmented and non-segmented dataset respectively). Also, there was an appreciable increase in true positive predictions in each class (Figure 3.12), which justifies the robustness of the proposed model with hand-crafted features. Although additional features were added to the model, this did not affect the processing time.

**Table 3-4** Performance analysis for proposed CNN, model trained with hand-selected features for segmented and non-segmented images.

| Model                                     | Data type     | Metrics  |         |        |          |
|-------------------------------------------|---------------|----------|---------|--------|----------|
|                                           |               | Accuracy | F-Score | Recall | Time (s) |
| Proposed Model +<br>Hand-crafted features | Segmented     | 98.90    | 98.20   | 98.50  | 35.29    |
|                                           | Non segmented | 96.71    | 96.11   | 96.03  | 44.11    |



a. Non-segmented images

(b) Segmented Images

**Figure 3-12** A confusion matrix depicting the classification results for the proposed CNN combined with hand-crafted features. This shows the accuracy of the models in classifying plants based on their stress levels. The true class (ground truth) are on the y-axis, whilst the predicted class are on the x-axis. Each matrix shows the percentage and count of correctly and incorrectly classified samples for four different classes: WWLN WWLN, DSHN and DSLN. (a) shows results for non-segmented plant images while (b) shows results for segmented plant images.

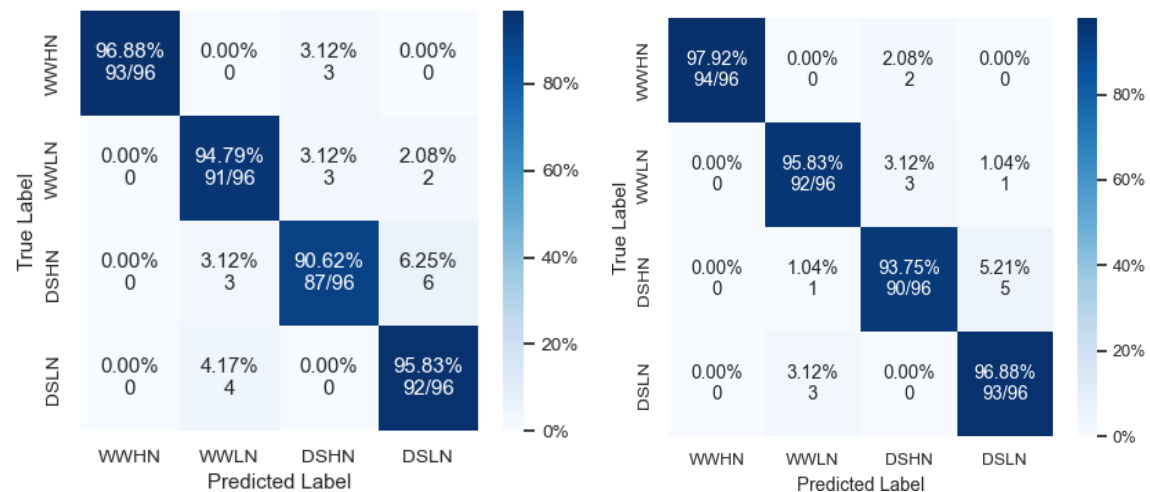
### 3.4.7 Performance of the proposed model at different time points

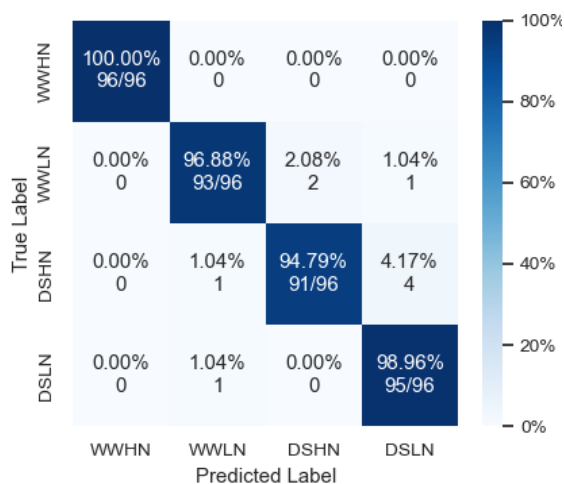
To ascertain the robustness of the proposed model, it was tested on plants at various levels of drought stress. The phases under consideration are 0, 3, 6, 9, 12, and 15 DADS. For each drought stress stage, 48 plants from each of the four treatments were employed, totaling 192 plants. Table 3-5 presents the results of the performance of the proposed CNN fused with the handcrafted model at the selected drought stress stages. Figure 3.13 is the confusion matrices showing performance of the model for different treatments at three selected drought stress stages (0, 9 and 15 DADS). Generally, the model had good performances across the selected stages. The model was better at the latter growth stages (the 12 and 15 DADS) as compared to the early growth stages (0 and 3 DADS) with the accuracy steadily increasing with each drought stage, reaching a plateau of 98% at

drought stages 9, 12 and 15 DADS. This suggests that as the drought stress increased, the model becomes more accurate in its classification of the different treatments. The F-score followed a similar pattern to the accuracy metric starting from 0.959 to 0.982 between stages 6 and 9, then settling at 0.985 by stage 15 DADS. This demonstrates that both precision and recall improve as DADS increases, indicating a balanced improvement in both sensitivity and precision of the model. There is consistent performance of the model from stages 9 to 15, with high values for all measures, indicating that the model performs especially effectively at higher drought stages. From the confusion matrix, the model had differentiating between the low N and drought stress plants at the early drought stress stage (0 DADS). This is especially so for the WWLN and DSHN plants with 94% and 90% accuracies respectively. However, the model had a perfect accuracy for the WWLN treatment identification. The model's challenges were improved on the 9 DADS where it attained 93% and 95% accuracies for WWLN and DSHN identification.

**Table 3-5** Performance of the proposed CNN model for each drought stress stage

| drought stage | Metrics  |         |        |
|---------------|----------|---------|--------|
|               | Accuracy | F-score | Recall |
| 0             | 0.952    | 0.949   | 0.950  |
| 3             | 0.966    | 0.961   | 0.963  |
| 6             | 0.981    | 0.978   | 0.981  |
| 9             | 0.987    | 0.982   | 0.983  |
| 12            | 0.989    | 0.983   | 0.985  |
| 15            | 0.992    | 0.985   | 0.981  |





(c) 15 DADS

**Figure 3-13** Confusion matrix depicting the classification of the four treatments (WWHN, WWLN, DSHN, DSLN) at selected drought stress stages (0, 9, and 15 DADS)

### 3.5 Discussion

This study aimed to understand the sensitivities of plant physiological features to drought stress under different nitrogen levels and to develop a lightweight deep learning model to differentiate plants with different water contents. The analysis was carried out using digital images acquired from two different views (top canopy and side view). The features were extracted from the colour, texture, morphology, and local image descriptors. The association between the hand-crafted features and the stress under investigation was investigated using the two sensitivity analysis techniques. While the OAT method provides a global view of the feature interactions, the Morris modeling method detailed the magnitude of the input features' effect on the model's output. The sensitivity analysis shows that the colour features, especially DGCI and RGBVI, as well as texture (DISS, ASM) and morphology features (HULL), were highly sensitive to nitrogen deficiency. The DGCI feature sensitivity to nitrogen deficiency agrees with the work of Caturegli et al., 2020 who compared NDVI acquired from the ground-based multispectral sensor with DGCI features from digital images of UAV to quantify the colour of turfgrass fertilized with different nitrogen levels. Their results revealed a high correlation between DGCI and the nitrogen clipping content of turfgrass, suggesting DGCI -features to be used as supporting decisions for nitrogen estimation. Moreover, DGCI was found to be very sensitive to chlorophyll content in the work of Hassanijalilian et al., (2020b). who

assessed the link between soybean chlorophyll content with SPAD metre conversion, proposing this feature to be used as a proxy tool to quantify chlorophyll and nitrogen in plants. Both sensitivity methods identified DGCIRGBVI (colour) and ASM (texture) as very sensitive to drought stress. This finding agrees with the work of Zhuang et al. (2017), who identified green-related colour features and energy features from plant texture properties as dominant in expressing drought stress in maize species. This demonstrates how the dynamics of drought stress in plants are highly expressed by both colour and textural features. The ASM feature was found to be highly sensitive to drought stress of plants in different nitrogen content. Thus, plant breeders can use this feature's statistical analysis to estimate the early onset and progress of nitrogen deficiency or drought stress in plants.

From the PCA plot in section 3.4.1, the PC1 and PC2 collectively account for 88.5% of the variations, which means that these two components capture most of the data's variability. From the colour gradient in the PCA plot, most of the variables contributed effectively to the variance indicating likely redundancy in the variables. In future research, instead of using the individual variables as inputs to a model, it is recommended that the PC1 and PC2 be used as inputs. The hand-crafted features were used to train three classical ML models, and the performances were compared. The side view features trained models were slightly better in performance than the top view models. This could be due to the high number of plant leaves on the side view, which increases the probability of getting more variations in pixel features for the model training. In contrast, the top view with a low spatial ratio had fewer plant pixels which reduced the probability of getting more pixels with information on nitrogen deficiency or drought stress for the model training. Although there were differences in the different model performances of the top and side views, they were not significant. It was also observed that the model had a better performance when trained with the combined features than the single features. While the combined features resulted a high dimensional training vector which increases the processing time, it had a high discriminative power for models to learn variance, patterns and trends for stress identification. A few combined features as used in this work is adequate as more feature combinations could result in curse of dimensionality, increasing space and computational time which risk overfitting the model during testing.



Hence, it is proposed that a few highly sensitive features, such as colour, texture and local descriptors, should be used.

The outcome of the deep learning modelling demonstrates that the proposed CNN model had relatively equal performance to the two state-of-the-art ResNet and Densenet models. The good performance of the proposed model shows that it can automatically learn the desired features for efficient classification of stress levels in plants. As shown in multiple works (Khaki et al., 2020) deeper layers of CNN models require massive datasets for high performance, which is not readily available. The proposed model, due to its lightweight architecture and high performance, could be used in place of the state-of-the-art DL models. Moreover, unlike the ResNet and Desnet, the proposed CNN requires few trainable parameters, making it ideal for installation on edge devices for the detection and tracking of drought stress in plans.

This study also shows that segmented images could improve the performance of DL when trained on limited samples. The possible increase in performance of the segmented image-based CNN models could be because segmentation enables extracting features only from the plant rather than from the plant and background. This ensures the CNN finds learning features relevant to the plant easier, facilitating fast and robust modelling with less misclassification. Fusing the hand-crafted features with the proposed CNN model resulted in an increase in the performance of the proposed model (3.6% better accuracy score) indicating the importance of the carefully selected features in identifying nitrogen stress from drought stress. However, the increment in performance rate could also be due to overfitting of the model. The hand-crafted features were selected from the plant images which were the same images used to train the CNN model. The likelihood of the proposed model extracting identical features to the hand-crafted features is significant, rendering the hand-crafted features redundant. Understanding the relationship between the hand-crafted features and the DL based features is required to understand the importance of the hand-crafted features to the performance of the model

### **3.6 Conclusions**

In this study, a hybrid CNN, which is made of the fusion of a lightweight CNN and hand-crafted features, was developed to identify nitrogen and drought-stress in wheat. The

hand-crafted features, selected from colour, texture, morphology and local image descriptors, were used to train three classical ML models: RF, SVM and DT. Although features from the side canopy performed better than those from the top view, there was no significant difference in the model's performance. Also, models trained with the single features resulted in a relatively good performance achieving 51-76% AA score. . However, an improvement in the model performance was observed when trained with the combined features. A sensitivity analysis of the extracted features shows that colour features were more sensitive to nitrogen stress, while colour and texture features were sensitive to drought stress. Four of the extracted features were sensitive to both stresses with ASM being the most sensitive feature. A time series analysis of the ASM feature shows the possibility of tracking nitrogen and drought stress dynamics in plants, hence, recommended to be used as a proxy tool to measure nitrogen or drought stress in wheat species. The proposed CNN model outperformed the classical ML models and had comparatively similar performance to the ResNet and DenseNet models.

## Acknowledgements

This research was supported by the OCP S.A. under the University of Mohammed VI Polytechnic, Rothamsted Research and Cranfield University project (FPO4). Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) through the Delivering Sustainable Wheat program (BB/X011003/1)

## 3.7 References

- Azimi, S., Kaur, T., & Gandhi, T. K. (2021a). A deep learning approach to measure stress level in plants due to Nitrogen deficiency. *Measurement: Journal of the International Measurement Confederation*, 173(September), 108650. <https://doi.org/10.1016/j.measurement.2020.108650>
- Azimi, S., Kaur, T., & Gandhi, T. K. (2021b). CHAPTER 3: A deep learning approach to measure stress level in plants due to Nitrogen deficiency. *Measurement: Journal of the International Measurement Confederation*, 173(June 2020), 108650. <https://doi.org/10.1016/j.measurement.2020.108650>
- Bogacsovics, G., Toth, J., Hajdu, A., & Harangi, B. (2022). Enhancing CNNs through the use of hand-crafted features in automated fundus image classification. *Biomedical Signal Processing and Control*, 76(October 2021), 103685. <https://doi.org/10.1016/j.bspc.2022.103685>

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Campolongo, F., & Braddock, R. (1999). The use of graph theory in the sensitivity analysis of the model output: A second order screening method. *Reliability Engineering and System Safety*, 64(1), 1–12. [https://doi.org/10.1016/S0951-8320\(98\)00008-8](https://doi.org/10.1016/S0951-8320(98)00008-8)
- Christopher J.C. Burges. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2, 121–167.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893 vol. 1. <https://doi.org/10.1109/CVPR.2005.177>
- Duan, L., Han, J., Guo, Z., Tu, H., Yang, P., Zhang, D., Fan, Y., Chen, G., Xiong, L., Dai, M., Williams, K., Corke, F., Doonan, J. H., & Yang, W. (2018). Chapter 3. Novel digital features discriminate between drought resistant and drought sensitive rice under controlled and field conditions. *Frontiers in Plant Science*, 9(April). <https://doi.org/10.3389/fpls.2018.00492>
- Greche, L., Akil, M., Kachouri, R., & Es-sbai, N. (2020). A new pipeline for the recognition of universal expressions of multiple faces in a video sequence. *Journal of Real-Time Image Processing*, 17(5), 1389–1402. <https://doi.org/10.1007/s11554-019-00896-5>
- Hassanijalilian, O., Igathinathane, C., Doetkott, C., Bajwa, S., Nowatzki, J., & Haji Esmaeili, S. A. (2020). Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Computers and Electronics in Agriculture*, 174(April), 105433. <https://doi.org/10.1016/j.compag.2020.105433>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.
- Keller, K., Kirchgessner, N., Khanna, R., Siegwart, R., Walter, A., & Aasen, H. (2019). Soybean leaf coverage estimation with machine learning and thresholding algorithms for field phenotyping. *British Machine Vision Conference 2018, BMVC 2018*, 1–12.
- Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10(January), 1–14. <https://doi.org/10.3389/fpls.2019.01750>
- Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, 71, 1–13. <https://doi.org/10.1016/j.patcog.2017.05.015>

- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8), 1778–1790. <https://doi.org/10.1109/TGRS.2004.831865>
- Menberg, K., Heo, Y., & Choudhary, R. (2016). Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information. *Energy and Buildings*, 133, 433–445. <https://doi.org/10.1016/j.enbuild.2016.10.005>
- Morandage, S., Schnepf, A., Leitner, D., Javaux, M., Vereecken, H., & Vanderborght, J. (2019). Parameter sensitivity analysis of a root system architecture model based on virtual field sampling. *Plant and Soil*, 438(1–2), 101–126. <https://doi.org/10.1007/s11104-019-03993-3>
- Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2), 161–174. <https://doi.org/10.1080/00401706.1991.10484804>
- Penatti, O. A. B., Nogueira, K., & Dos Santos, J. A. (2015). Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2015-Octob*, 44–51. <https://doi.org/10.1109/CVPRW.2015.7301382>
- Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling and Software*, 70, 80–85. <https://doi.org/10.1016/j.envsoft.2015.04.009>
- Pound, M. P., French, A. P., Atkinson, J. A., Wells, D. M., Bennett, M. J., & Pridmore, T. (2013). RootNav: Navigating images of complex root architectures. *Plant Physiology*, 162(4), 1802–1814. <https://doi.org/10.1104/pp.113.221531>
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Shah, S. H., Angel, Y., Houborg, R., Ali, S., & McCabe, M. F. (2019). A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sensing*, 11(8). <https://doi.org/10.3390/rs11080963>
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, 21(2), 110–124. <https://doi.org/10.1016/j.tplants.2015.10.015>
- Vergara-Diaz, O., Kefauver, S. C., Elazab, A., Nieto-Taladriz, M. T., & Araus, J. L. (2015). Grain yield losses in yellow-rusted durum wheat estimated using digital and conventional parameters under field conditions. *Crop Journal*, 3(3), 200–210. <https://doi.org/10.1016/j.cj.2015.03.003>

- Wilf, P., Zhang, S., Chikkerur, S., Little, S. A., Wing, S. L., & Serre, T. (2016). Computer vision cracks the leaf code. *Proceedings of the National Academy of Sciences of the United States of America*, 113(12), 3305–3310. <https://doi.org/10.1073/pnas.1524473113>
- Xu, Z., Guo, X., Zhu, A., He, X., Zhao, X., Han, Y., & Subedi, R. (2020). Using deep convolutional neural networks for image-based diagnosis of nutrient deficiencies in rice. *Computational Intelligence and Neuroscience*, 2020. <https://doi.org/10.1155/2020/7307252>
- Zermas, D., Nelson, H. J., Stanitsas, P., Morellas, V., Mulla, D. J., & Papanikolopoulos, N. (2021). A Methodology for the Detection of Nitrogen Deficiency in Corn Fields Using High-Resolution RGB Imagery. *IEEE Transactions on Automation Science and Engineering*, 18(4), 1879–1891. <https://doi.org/10.1109/TASE.2020.3022868>
- Zhuang, S., Wang, P., Jiang, B., Li, M., & Gong, Z. (2017). Chapter 3. Early detection of water stress in maize based on digital images. *Computers and Electronics in Agriculture*, 140, 461–468. <https://doi.org/10.1016/j.compag.2017.06.022>

## **4 Modelling the spatial-spectral characteristics of plants for nutrient status identification using hyperspectral data and deep learning methods.**

Frank Gyan Okyere <sup>1,2</sup>, Daniel Cudjoe <sup>1,2</sup>, Pouria Sadeghi-Tehran<sup>1</sup>, Nicolas Virlet<sup>1</sup>, March Castle <sup>1</sup>, Andrew B. Riche<sup>1</sup>, Latifa Greche<sup>1</sup>, Fady Mohareb <sup>2</sup>, Daniel Simms <sup>2</sup>, Manal Mhada <sup>3</sup> and Malcolm John Hawkesford <sup>1, \*</sup>

<sup>1</sup>Sustainable Soils and Crops, Rothamsted Research, Harpenden, United Kingdom, 2 School of Water.

<sup>2</sup>Energy and Environment, Cranfield University, Cranfield, United Kingdom, 3 AgroBioSciences.

<sup>3</sup>Department, University of Mohammed VI Polytechnic, Ben Guerir, Morocco

Based on: Okyere FG, Cudjoe D, Sadeghi-Tehran P, Virlet N, Riche AB, Castle M, Greche L, Simms D, Mhada M, Mohareb F and Hawkesford MJ (2023) Modeling the spatial-spectral characteristics of plants for nutrient status identification using hyperspectral data and deep learning methods. *Front. Plant Sci.* 14:1209500. doi: 10.3389/fpls.2023.1209500

### **Abstract**

Sustainable fertilizer management in precision agriculture is essential for both economic and environmental reasons. To effectively manage fertilizer input, various methods are employed to monitor and track plant nutrient status. One such method is hyperspectral imaging, which has been on the rise in recent times. It is a remote sensing tool used to monitor plant physiological changes in response to environmental conditions and nutrient availability. However, conventional hyperspectral processing mainly focuses on either the spectral or spatial information of plants. This study aims to develop a hybrid convolutional neural network (CNN) capable of simultaneously extracting spatial and spectral information from quinoa and cowpea plants to identify their nutrient status at different growth stages. To achieve this, a nutrient experiment with four treatments (high and low levels of nitrogen and phosphorus) was conducted in a glasshouse. A hybrid CNN model comprising a 3D CNN (extracts joint spectral spatial information) and a 2D CNN (for abstract spatial information extraction) was proposed. Three pre-processing

techniques, including second-order derivative, standard normal variate and linear discriminant analysis, were applied to selected regions of interest within the plant spatial-spectral hypercube. Together with the raw data, these datasets were used as inputs to train the proposed model. This was done to assess the impact of different pre-processing techniques on hyperspectral-based nutrient phenotyping. The performance of the proposed model was compared with a 3D-CNN, a 2D-CNN and a Hybrid Spectral Network (HybridSN) model. Effective wavebands were selected from the best-performing dataset using a greedy stepwise-based correlation feature selection (CFS) technique. The selected wavebands were then used to retrain the models to identify the nutrient status at five selected plant growth stages. From the results, the proposed hybrid model achieved a classification accuracy of over 94% on the test dataset, demonstrating its potential for identifying nitrogen and phosphorus status in cowpea and quinoa at different growth stages.

**Keywords:** convolutional neural network, hyperspectral imaging, plant nutrition, machine learning, spectral curve

## 4.1 Introduction

Nitrogen (N) and phosphorus (P) are two essential macronutrients that play a significant role in the normal functioning and growth of plants. They are involved in vital plant metabolic processes, such as cell division, protein formation and photosynthesis (Siedliska et al., 2021b). Plants with adequate nitrogen nutrition display green leaves, whilst nitrogen deficiency manifests as chlorosis, starting from light green and progressing to yellow and eventually brown. Phosphorus deficiency inhibits shoot growth and shows decolourized leaves, transitioning from pale green to yellow in severely affected regions (McCauley et al., 2011). When plants are insufficient in nutrients, fertilizer application is required to enrich the soil. However, the application of nitrogen and phosphorus fertilizer often relies on farmer's experience and intuition, which may result in over- or under-application. This practice can lead to soil quality degradation, crop yield reduction, environmental pollution and loss of biodiversity (Gao et al., 2021). Hence, accurate estimation and tracking of plant N and P status is essential to promote good agronomic practice and effective fertilizer management.

The tracking of plants' N and P status traditionally involves visual inspection or laboratory-based chemical analysis, which can be destructive, labour intensive and prone

to error. To indirectly measure nutrient status, breeders and researchers use contact and remote sensing tools. One such tool is the SPAD (Soil Plant Analysis Development) meter, which estimates nitrogen content by measuring the light transmittance of the red (650 nm) and infrared (940nm) wavelengths through plant leaves (Uddling et al., 2007). While this technique is simple and fast, it has limitations. As a leaf contact instrument, it only captures a small area of contact (2 x 3 mm), which may not provide an accurate representation of the spatial variation of nitrogen in plants. Additionally, when these tools are used on a large scale, users may introduce errors and obtain false measurement (Hassanijalilian et al., 2020b). To mitigate some of these challenges, image-based non-invasive phenotyping techniques offer a potential solution.

Image-based techniques have proven valuable in plant phenotyping, allowing for the measurement of various plant phenotypic properties such as biomass, leaf area and plant height (Hairuddin et al., 2011). These techniques utilize images acquired from digital RGB cameras, hyperspectral and multispectral sensors and 3D laser scanners to extract non-invasive features, trends and patterns that demonstrate the dynamics of phenotyping traits in response to physiological and chemical changes in plants. Among these imaging sensors, digital cameras are the most widely used for plant phenotyping due to their low cost, portability and high spatial resolution (Sadeghi-Tehran et al., 2017). However, their spectral limitations, capturing only broad spectral bands of red, green and blue, present a major challenge. To overcome this challenge, hyperspectral imaging (HSI) has emerged as a promising technique. By combining imaging and spectroscopy, HSI allows for the acquisition of spectral and temporal information from plants, useful for estimating plant physiological parameters (Mishra et al., 2017). HSI samples the reflective areas of the electromagnetic spectrum spanning from the visible regions (400- 700 nm) to the short-wave infrared regions (1100-2500 nm). This approach has been successfully applied in non-destructive phenotyping of plant leaf area index (J. Zhang et al., 2021), plant biomass (Ma et al., 2020), plant nutrient estimation (Ye et al., 2020) and detection of diseases and fungal infections (Siedliska et al., 2018).

HSI data are presented in a hypercube format, with the first two dimensions providing spatial information and the third dimension representing spectral information. Extracting relevant information from these hypercubes requires advanced pattern recognition algorithms, such as machine learning (ML) (Singh et al., 2016). ML algorithms learn



patterns and trends from data without relying on explicit programming. In a study by Yamashita et al., (2020), ML regression models were employed to analyze differences in spectral reflectance and estimate N and chlorophyll contents in tea plants. A data-based sensitivity algorithm was applied to select the most informative spectral bands. The results showed that the integration of spectral data with machine learning models is a promising technique for accurate plant nutrient estimation. To further improve the application of ML models to HSI, researchers are currently exploring deep learning methods.

Deep neural networks, particularly convolutional neural networks (CNNs), have emerged as state-of-the-art machine learning techniques capable of detecting, classifying and predicting plant phenotypes. In the context of HSI, CNNs have been successfully combined with HSI for various applications such as crop and weed classification (Dyrmann et al., 2016), plant seedlings classification (Nkemelu et al., 2018) and plant disease and pest detection (Sladojevic et al., 2016; Fuentes et al., 2017). Most traditional deep learning methods employed for HSI extract either spectral or spatial information separately (Hong & He, 2020). For instance, 2D-CNN models focus on spatial information, while 1D-CNN models primarily utilize spectral information from HSI. In Roy et al., (2020) and Mu et al., (2021) work, combining information from both 2D and 1D CNN models produced better results than using only the spectral or spatial information. However, since these two are extracted separately, there is difficulty in fully utilizing all information from the two simultaneously. To address this limitation and effectively utilize the structural traits of the HSI data, researchers have introduced 3D-CNN methods (Mirzaei et al., 2019; Pi et al., 2021). Indeed, 3D-CNN modeling naturally suits the extraction of spatial-spectral information from the HSI hypercube and presents a promising approach for modeling different scenarios of data information. For example, (2020) proposed a novel 3D-CNN model to extract combined features to classify meat using HSI imaging, while Jung et al. (2022) developed a 3D-CNN classification model for diagnosing grey mould disease in strawberries. In both studies, the models employed 3D convolutions, filters and kernels to extract relevant spatial and spectral information from the hypercube. However, homogeneous 3D models pose difficulties in optimization due to increased training parameters, resulting in lengthy computational time, over-fitting and gradient disappearance. These challenges are attributed to the increased complexity

of the models. To further understand CNN modeling using HSI data, readers are referred to Kumar et al., (2020); Paoletti et al., (2019); Huang et al., (2022); Mu et al., (2021) and Ma et al., (2020) for more details.

To mitigate some of these challenges, researchers have proposed the use of hybrid CNN models, which fuse two-unit deep learning blocks together to simultaneously extract both spectral and spatial information from HSI data. For instance, Mohan & Venkatesan (2020) proposed a multiscale hybrid CNN model for hyperspectral image classification. This model extracts spatial-spectral features from different window sizes using a 3D-CNN block. The output from the 3D CNN block is concatenated and fed into a 2D CNN block for abstract spatial feature extraction. The performance of the model showed satisfactory performance when compared to state-of-the-art CNN based models. Although this method works effectively in simultaneously extracting spectral and spatial features from hypercubes, there is a considerably high number of trainable parameters, which affects the processing time.

In this study, we developed a hybrid model capable of extracting spatial and spectral information simultaneously from a hypercube with reduced computational complexities and time. The proposed model is to identify plant nutrient status (N and P) at different growth stages. Specifically, hypercube regions of interest (patches) measuring 15x15-pixels were extracted from four nutrient treatments. Different spectral transformation techniques were applied to obtain four distinct datasets. The resulting datasets were each used to train the proposed model to classify plants based on their nutrient status. Furthermore, a waveband selection method was applied to the dataset produced by the best performing pre-processing technique. The selected wavelengths were used as inputs to retrain the proposed model to identify plant nutrient status at different growth stages.

## **4.2 Materials and Methods**

Figure 4.1 shows the workflow of this study. It involves generating the spectral library of hypercube patches to build four different datasets. These datasets were used to train four deep learning models to classify plant nutrient status. The most informative dataset was then used to retrain the models based on the selection of nutrient deficiency-sensitive wavebands from the full plant data (not patches) to classify plants at selected growth stages.

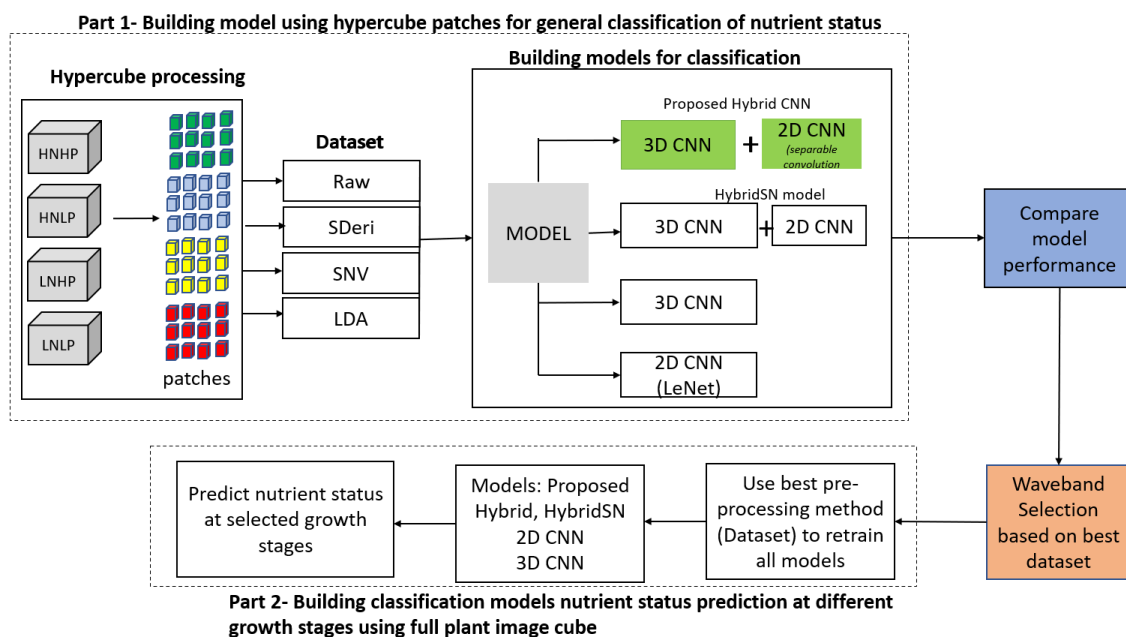


Figure 4-1 Flowchart outlining the process of building the proposed 3D-2D hybrid model and comparing it to the state-of-the-art models for plant nutrient status classification using hyperspectral imaging data. The diagram is divided into two main sections. Part 1 describes the development and training of different models using hypercube patches for plant nutrient status classification. Part 2 describes the building of different models using the best training datasets for plant nutrient status classification at different growth stages.

#### 4.2.1 Plant materials and treatments

*Vigna unguiculata* (cowpea) and *Chenopodium quinoa* (quinoa) were cultivated in a Plant Growth Facility (<https://www.cranfield.ac.uk/facilities/plant-growth-facility>) at Cranfield University (Cranfield, UK). The plants were grown in pots filled with compost with reconstituted nutrients. The compost was first washed with deionized water following the works of Masters-Clark et al.(2020) with some modifications. This was done by flooding one-part compost with five-part deionized water. A double-headed 0.5  $\mu\text{m}$  sieve was used to drain off the soluble solution. The steps were repeated five times, ensuring the plant nutrients were depleted and the washed compost was dried in an oven at 70°C for 72 hours. Based on the Letcombe nutrition solution (Masters-Clark et al., 2020), the required treatments were prepared as high nitrogen-high-phosphorus (HNHP), high-nitrogen-low-phosphorus (HNLP), low-nitrogen-high-phosphorus (LNHP) and low-nitrogen-low-phosphorus (LNLP), with five replicates for each treatment. The treatments had concentrations of 49.1 mM and 14.6 mM N for HN and LN and 13.4 mM and 3.3 mM P for HP and LP, respectively. By visual inspection, the different treatments

resulted in changes in the leaf pigments, resulting in variations in the spectral characteristics during plant development. For both species, the leaves were green when the nitrogen and phosphorus levels were adequate, while deficiency in nitrogen was characterized by chlorosis progressing from light green to yellow to brown. Phosphorus deficiency inhibited shoot growth and decolourized leaves from a blue-green colour to pale green/yellow in severely affected regions (McCauley et al., 2011).

#### **4.2.2 Hyperspectral imaging system and data collection**

Spectral images were collected 3.0 m above the canopy level using a Lemnatec Scanalyzer installed in the glasshouse. The Scanalyzer has a push broom hyperspectral camera (hyperspec® inspector™ Headwall Photonic) that operates within the 390 nm – 2500 nm region covering the visible–near-infrared (VNIR) and short-wave infrared (SWIR) regions. The sensor collects data with a 0.7 nm step (at the VNIR region), and 6 nm step (within the SWIR region), and an FWHM (full width at half maximum) image slit of 2.5 nm. This results in a hypercube of 1015 spectral bands for the VNIR data and 275 spectral bands for the SWIR data, with a dynamic range of 16 bits. Data collection was performed twice per week throughout the full development of the plants. Five different growth stages of cowpea and quinoa, hereafter referred to as growth stages I, II, III, IV and V, were considered for this study. They represent the 19, 29, 36, 50 and 69 DAT (days after transplanting) and 13, 30, 51, 60 and 72 DAT for the cowpea and quinoa plants, respectively. Supplementary information (Table D-1) describes the selected growth stages based on the BBCH system for coding the phenological growth stages of plants (Meier et al., 2009).

#### **4.2.3 Pre-processing of raw spectral data**

Prior to any analysis, the raw data were pre-processed to normalize the spectral data from ambient illumination and to reduce noise and other artefacts that were produced during scanning. The preprocessing steps comprised (i) radiometric calibration to remove the illumination system effect, (ii) spectral down-sampling to remove redundant wavelengths and (iii) noise reduction.

*Radiometric calibration:* Hyperspectral image acquisition suffers from radiometric errors caused by illumination from varying light exposure. Radiometric correction is essential to reduce the variable illumination effect and the influence of dark current on the spectral

data. During scanning, a white panel (Zenith Lite™ Ultralight Targets 95%R, Sphereoptics®) was imaged as the white reference data. Dark reference data with approximately 0% reflectance were collected in the night without any light source by completely covering the camera lens with an opaque cap (Zhu et al., 2013). The calibrated data were calculated using equation (4.1):

$$I_c = \frac{I_{\text{raw}} - R_{\text{dark}}}{R_{\text{white}} - R_{\text{dark}}} \quad \text{Equation 4.1}$$

Where  $I_c$  is the calibrated data,  $I_{\text{raw}}$  is the raw spectral data,  $R_{\text{dark}}$  is the dark reference and  $R_{\text{white}}$  is the white reference.

**Down-sampling:** Although the hyperspectral data contain essential plant information, the huge dataset poses computational challenges. Down-sampling helps to reduce these computational complexities and the noise generated during scanning (Sadeghi-Tehran et al., 2021). Moreover, down-sampling is conducted to map the spectral resolution to its reference target. In this study, an averaging window with a 2 nm spectral width was used to down sample the spectral data. It is a common practice to remove the first and last few bands of hyperspectral data before processing to reduce noise and artifacts effects. Also, at the extreme ends of the spectrum, the signal-to-noise ratio tends to decrease, leading to less useful information in those bands. Hence it is prudent to remove those bands before processing. Additionally, removing the extreme ends of the spectrum reduces the effect of atmospheric distortions that are captured during image acquisition which can introduce errors in the analysis. In this work, the first and last 15 wavelengths of the VNIR datasets were removed a two level-down-sampling technique (first level: 3 nm window and second level: 2 nm window width) was applied to the resulting datasets giving a final VNIR hypercube with 165 wavelengths. Similarly, the first 25 and the last 15 wavelengths were removed from the SWIR data resulting in hypercube of 235 wavelength. A two stage down-sample technique (first stage uses a 3 nm spectral width and second stage uses a 2 spectral width) was employed reducing the data to 58 wavebands. The combined dataset before preprocessing was 1290 wavelengths (1015 VNIR and 275 SWIR) and the hypercube after pre-processing was 223 wavelengths (165 for VNIR and 58 for SWIR).

datasets). The removal and down-sampling technique ensured that the resulting datasets is of high quality and more suitable data for analysis.

**Spectra smoothing and denoising:** This is a common preprocessing practice that involves some numerical operations on the raw spectral data to reduce spectral noise levels. This eliminates spikes and smooths spectral curves while isolating essential features, which may be obscured by the presence of noise across the different wavelengths. During spectral smoothing and denoising, the original shape and features of the spectra are normally preserved. In our study, we applied the Savitzky–Golay filter (a commonly used low-pass filter), an effective and computationally fast filter, to smooth the spectra. For each data point in the spectrum, the filter selects an odd-sized window of spectral points and fits a least-square regression with a polynomial of higher order. During this operation, the data points in question are eventually replaced with the corresponding values of the fitted polynomial. We used a window of size 13- and second-degree polynomials as the optimal parameters. It should be noted that a small window size leads to the emergence of large artefacts in the smoothed spectra, while the larger the window size, the smaller the distinction between full and moving window processing (Rinnan et al., 2009).

**Segmentation:** After pre-processing, the HSI data were segmented using a selected spectral ratio combined with Otsu thresholding. To segment the VNIR dataset, 705 nm and 750 nm wavelengths were extracted to create a red-edge normalized difference vegetation index (RENDVI) image, which was then combined with automatic Otsu thresholding for segmentation. The RENDVI generally differentiates plant vegetation from non-vegetation regions. Hence, the combination with Otsu thresholding creates a binary image where the plant images are labeled as one and all other items labeled as zeros. The SWIR dataset segmentation used a similar approach, where the NDVI images were derived from wavelengths 1375 nm and 1141nm, as proposed by Williams et al (2017).

#### **4.2.4 Data transformation**

Although HSI contains important information that relates objects to their absorptance and reflectance, the measured spectra are subject to spurious signals such as multiplicative and additive effects. Hence, further pre-processing steps are applied to minimize the

effect of undesirable occurrences on spectral measurement including artifacts, particle size effects and light scattering (Amigo, 2010). In this study, the effects of different pre-processing techniques on the performance of some CNN classification models were studied. The datasets were created by normalizing or transforming the raw HSI data into different dimensions using techniques such as standard normal variate (SNV), second derivative (SDeri) and linear discriminant analysis (LDA). In addition to the pre-processed raw data, these new datasets (SNV, SDeri and LDA) were modelled for plant stress identification.

#### 4.2.4.1 Standard normal variate

SNV is a pre-processing technique aimed at limiting the multiplicative effect of scattering and particle size. It accounts for the variation in baseline shift and curvilinearity in the reflectance spectra and reduces the difference in the global intensities of the reflectance spectra (Luo et al., 2019). SNV transformation was performed on each individual spectrum, requiring no reference spectrum. The raw spectra were transformed using the SNV technique by finding the mean centre of each spectrum and dividing by its standard deviation (Barnes, et al., 1989), as shown in equation 4.2:

$$x_{i,j}^{\text{SNV}} = \frac{\sum_j^y x_{i,j} - \bar{x}_i}{\sqrt{y-1}} \quad \text{Equation 4.2}$$

Where  $x_{i,j}^{\text{SNV}}$  is the element of the transformed spectrum,  $x_{i,j}$  is the corresponding original element of the spectrum  $i$  measured at wavelength  $j$ ,  $\bar{x}_i$  is the mean reflectance of the uncorrected spectra  $i$  and  $y$  is the number of variables or wavelengths in the spectrum.

It should be noted that SNV is performed independently for each pixel (has zero mean and variance equal to one). Hence, it gives an advantage over averaging methods such as MSC (Multiple Scattering Correction), where the presence of non-plants could influence the averaging process.

#### 4.2.4.2 Spectral derivatives

Spectral derivative aims to normalize the differences between two continuous narrow bands and remove or suppress image artefacts that result from non-uniform illumination (Pandey et al., 2017). It increases the spectral resolution of overlapping peaks while

accounting for the baseline correction of reflectance spectra (Perez-Sanz et al., 2017). According to Cloutis et al., (1996), the spectral derivative is sensitive to the signal-to-noise ratio (SNR), such that the higher the spectral derivative, the higher the noise and vice versa. From the literature, the first and second derivatives are more effective in managing different spectral disparities and improving data modeling. In this study, we employed the second derivative (SDeri) using the Savitzky–Golay smoothing and polynomial derivative package in Python. Supplementary information (Figure D-1) shows the hypercube of cowpea with examples of the raw, SNV and SDeri spectral curves.

#### **4.2.4.3 Linear discriminant analysis**

The high spatial and spectral resolution of the data in this study posed a computational challenge to the analysis of the hypercube. Hence, it was desirable to apply a dimensionality technique to reduce the depth of the hypercube while maintaining the core informative features of the data. To do so, the hypercube defined as ( $w \times h \times l$ ) was rearranged into a 2-d spectral matrix ( $M \times l$ ), where  $M$  is the product of  $w$  and  $h$ . A linear discriminant analysis (LDA) was applied to find the linear combination of spectral features that characterize the different treatments. LDA is a probabilistic generalization technique that aims at projecting features in higher dimensions to lower dimensions for solving classification and regression problems (Xia et al., 2019). It reduces the size of a dataset while retaining the relevant information that discriminates the different classes. Unlike principal component analysis (PCA), which takes in only the spectral features and its variances irrespective of their classes, LDA makes use of the different class labels to maximize the differences in the classes. Supplementary information (Figures D-2 and D-3) are scatter plots of the first and second discriminants using LDA for cowpea and quinoa respectively. Comparing Supplementary information Figures D-2, and D-3 for LDA and PCA, respectively, it is observed that LDA performed better in separating the four classes. In the PCA, the classes were not as clearly separated, even though together the first two principal components contained over 90% of the classes information.



## 4.2.5 Dataset for model training

### 4.2.5.1 Generating hypercube patches

The original HSI hypercube obtained from the Scanalyzer contained 1397 spectral dimensions (1015 and 275 for VNIR and SWIR, respectively) for each plant. As mentioned earlier, spectral band removal and down-sampling techniques were applied on both species, reducing the datasets to 223 (155 VNIR and 68 SWIR) spectral dimensions for further analysis. To build the classification model, an automated algorithm was created to extract hypercubes from the four datasets (raw, second derivative, LDA and SNV data). This was done by building a spectral library of regions of interest (patches) from the plant leaves. A patch (Supplementary information Figure D-4) is defined as a square area of size 15 x 15 pixels and has a spectral dimension with  $\lambda$  number of wavelengths. From a hyperspectral image ( $H \in I^{(h \times w \times y)}$ ),  $n$  number of patches ( $P \in I^{(m \times n \times k)}$ ) can be produced depending on the patches size required, where  $h \times w$  and  $m \times n$  represent the width x height dimensions of the hyperspectral image and patch hypercube, respectively, while  $y$  and  $k$  denote the spectral bands for patches and the image respectively. To produce the patches, the hypercube containing the whole plant was divided into four equal segments. A function that utilizes the patchy library in Python was developed to create patches on each segment, as shown in Supplementary Information Figure D-4. Then, 20 non-zero patches from each segment were extracted, producing 80 patches per image and resulting in 2800 patches per treatment. The patches for each plant species were generated from 35 (7 time points and 5 replications) hyperspectral images per treatment.

### 4.2.5.2 Developing the Training Dataset

Since the hyperspectral data for both species from each treatment were divided into 2800 sample patches, there was a total of 11,200 patches used for the experiment. Before extracting the hypercube patches for training the models, the original hyperspectral data were first divided into a 3:1 train–test dataset. The training dataset was further split into an 80%-20% train–validation dataset. The splitting was done on the hypercube patches and not the pixels. The number of samples of the training, validation and test set for each model is shown in Table 4-1. To introduce spatial variations into the dataset, the 40% of

the train–validation dataset was rotated at different angles between -20° and 20°. All this was done for both species.

**Table 4-1** The number of training, validation and testing samples

| Number of samples (patches) | Number of Treatment | Training set | Validation set | Test set | Dataset Type |
|-----------------------------|---------------------|--------------|----------------|----------|--------------|
| 2800                        | 4                   | 1680         | 420            | 700      | Patches      |
| 2800                        | 4                   | 1680         | 420            | 700      | Patches      |
| 2800                        | 4                   | 1680         | 420            | 700      | Patches      |
| 2800                        | 4                   | 1680         | 420            | 700      | Patches      |

## 4.2.6 Developing models for spatial-spectral characterization of HSI data

### 4.2.6.1 Hybrid CNN framework hyperspectral classification

In this study, we employed a hybrid 3D-2D CNN method to model the spatial–spectral variations and interclass appearance to improve the power of accurately differentiating the N and P variations in cowpea and quinoa. The next paragraph explains the architecture and operations of the two units of the model.

*3D Convolutional Unit:* This block employs a 3D convolutional technique where the kernel slides in three dimensions convolving input data in their spatial and spectral dimensions, resulting in a 3D data output. It is made of three 3D convolutional layers interspersed with layers of filters ranging from 64, 32 and 16, constituting a (3 x 3x 3), (3 x3 x 5) and (3 x 3 x 7) kernel size, respectively. In addition, the filters and the generated feature maps are all configured in a 3D format. The 3D convolutional operation is given by equation (3) (Mohan & Venkatesan, 2020a):

$$y_{qrs} = f(\sum_r \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \sum_{l=0}^{b-1} k_{ijl} x_{(i+m)(j+n)(l+p)} + b_{mnp})$$

**Equation 4.3**

where  $y_{qrs}$  denotes the extracted feature at position (q, r, s) and  $k$  is the kernel at dimension  $ijl$ .  $S$  is the kernel size in the spectral dimension.

*2D CNN Unit:* In this block, the input is convolved with 2D kernels and filters where the convolution is mathematically obtained by the sum of the dot product between the kernels and the input data. The kernel strides over the entire data to acquire the spatial information. As mentioned by Mohan & Venkatesan (2020), the 2D convolution process can be mathematically expressed as:

$$T_{mn} = \sum_l \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} k_{ij} x_{(i+m)}(j+n) + bias_{mn}$$

**Equation 4.4**

where  $T_{mn}$  is the feature extracted at position (m, n) and k is the kernel (2D) at dimension  $h \times w$ .

This convolution operation is applied on all feature maps ( $l$ ) in the receiver area summing all the values for non-linear activation. This block has one 2D convolutional layer (separable in nature), a max pooling layer and two fully connected (FC) layers and SoftMax layers. The separable convolutional is a transformation form of conventional 2D convolution in which a single convolution is divided into two or more convolutions to get the same output. There are two steps in separable convolutions: spatial and depthwise separable convolution (Firat et al., 2022). Spatially separable optimizes the performance of the convolution networks to help preserve the spatial information of the data, while the depthwise convolution decreases the number of trainable weight parameters while increasing representation efficiency. To prevent the model from overfitting, a 0.5 probability dropout was used on each fully connected layer. The output is fed into a SoftMax layer. Figure 4.2 summarizes the hybrid CNN architecture for this study. Table 4-2 also shows the output dimension and number of parameters used in each layer for the hybrid model.

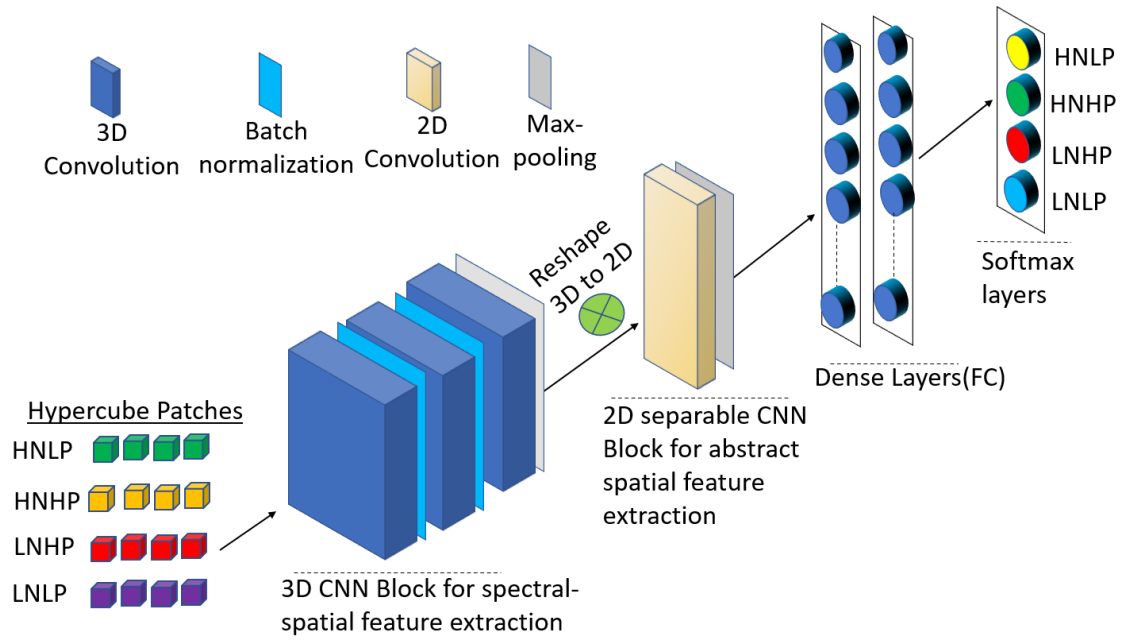


**Figure 4-2** Schematic representation of the proposed 3D-2D hybridCNN model frameworks with the full-spectrum bands as inputs. The green, yellow, red and violet boxes represents the hypercube patches extracted from the HNLP, HNHP, LNHP and LNLP classes respectively. The 3D block has stacks of 3D convolutional layers interspersed with batch normalization and max pooling layers. The 2D component is seperable in nature with max-pooling layers. The dense layers transform the feature maps into a vector suitable for classification.

#### 4.2.6.2 Other models for comparisons (3D CNN, 2D CNN and HybridSN)

The performance of the proposed method was compared to other methods, such as the hybrid spectral convolutional neural network (HybridSN), 3D-CNN and 2D-CNN models. The HybridSN is a modification of the model proposed by Roy et al. (2020) for HSI classification. It is a 3D-2D CNN joint model that combines the complementary spectral-spatial information from the 3D-CNN component and abstract-level information from the 2D CNN model. In this model, the first three layers are homogeneous 3D convolutions, with a 3x3x3 kernel size in the first and third layers and 3x3x5 filter kernels in the second layer. Each layer has 8, 16 and 32 filters, respectively. The output from the last layer of the 3D block is reshaped and fed into the 2D CNN block. This block is made of a classical 2D convolution with a 3 x 3 kernel size and 64 filters. The output is flattened and passed through two dense layers with Dropout. This is followed by a single SoftMax layer corresponding to the output. Supplementary information Table D-2 shows the detailed architecture of HybridSN used in this study. The hybrid SN is similar in operation to the proposed hybrid CNN model. Their similarities lie in the 3D CNN model followed

by the 2D CNN model. However, the proposed hybrid CNN is different in the use to the 2D CNN model which is highly efficient and reduces computational complexity making it ideal for practical applications especially in areas requiring low latency or resource constrained environments.

The 3D-CNN model for hyperspectral data modeling is known for extracting spatial and spectral information simultaneously from hyperspectral data. It is especially useful when relevant information is localized in both the spatial and spectral dimensions and exhibits good correlations in both domains (Nagasubramanian et al., 2019). The 3D-CNN model used in this study has four 3D convolutional layers, with 3D kernels of size  $3 \times 3 \times 3$  used for the input of the first convolution layer and three kernels of sizes  $3 \times 3 \times 5$  in between the other three layers. The convolutional layers are interspersed with maxpooling and batch normalization layers between every two convolutional layers. It uses the rectified Linear Input (ReLU) activation function for each convolution output (Glorot et al., 2011). There are two fully connected layers that follow the last convolution layer. Dropout with a 0.5 probability was performed after the first max pooling operation. The dropout process was used to prevent overfitting during training. The last fully connected layer was fed to the SoftMax layer for classification. Supplementary information Table D-3 is the detailed architecture of 3D CNN for the classification of nutrient status.

The 2D-CNN model is a modified version of LeNet-5 deep learning architecture introduced by LeCun et al. (1998). This model was chosen due to its simple and straightforward architecture. It is made of two convolutional layers with two pooling layers interspersed between the convolutions. It also has a flattening layer, two fully connected layers and a SoftMax layer that classifies the resulting features. The convolutional layer is responsible for generating feature maps by sliding the given filters over the input data and recognizing patterns and trends. The first convolutional layer has a  $3 \times 3$  kernel size and a stride of one, which outputs feature maps of sizes  $11 \times 11 \times 6$ , while the second convolutional layer with a  $3 \times 3$  kernel size and a stride of one takes in  $11 \times 11 \times 6$  input feature maps and outputs  $7 \times 7 \times 12$  feature maps. The pooling layer at each end of the convolutional layers is made of filters of size  $2 \times 2$  and a stride of two which downsamples the feature maps by calculating the average value of the patches of each feature map. The first pooling layer halves the size of the feature map from  $11 \times 11 \times 6$  to  $5 \times 5 \times 12$ , while the second pooling layer halves the size of the feature map from  $5 \times 5 \times 12$

12 to 2 x 2 x 24. The fifth and sixth layers are fully connected layers with 120 feature maps of sizes 1 x 1 and 140 feature maps of sizes 1 x 1, respectively. In addition, the ReLU activation function was used in this architecture to help mitigate the vanishing gradient problem and speed up computation and convergence in the model. Supplementary information (Table D-4) shows the detailed architecture of the 2D-CNN model (LeNet) used in this study.

The four models were each trained with RAW, SNV, SDeriv and LDA datasets, resulting in four models for each dataset. For the validation models, a stratified 10-fold cross validation with five repeats was conducted.

All the models were programmed in Python 3.8 and implemented based on TensorFlow and Keras open-source framework. The operating platform was on a PC with Intel (R) Core (TM) i7-370K U CPU with 3.50 GHz and 16 GB RAM. All the classification algorithms were established using the full spectrum (390–2500 nm) and the same parameters (window size, training sample, testing sample) were set for a fair comparison.

**Table 4-2** The proposed hybrid-CNN algorithm structure.

| Layer (Type)                         | Output Shape   | Number of Parameters |
|--------------------------------------|----------------|----------------------|
| Input_layer1(input layer)            | 15, 15, 223, 1 | 0                    |
| Conv3D (Conv3D)                      | 13, 13, 221, 8 | 224                  |
| batch normalization                  | 13, 13, 221, 8 | 32                   |
| Conv3D-1 (Conv3D)                    | 11, 11, 219, 8 | 1736                 |
| batch normalization                  | 11, 11, 219, 8 | 32                   |
| Conv3D-2 (Conv3D)                    | 9, 9, 217, 16  | 3472                 |
| MaxPooling3D                         | 4, 4, 108, 16  | 0                    |
| Reshape 3D-2D                        | 4, 4, 1728     | 0                    |
| separable_conv2d-2(Separable Conv2d) | 2,2,32         | 70880                |
| MaxPooling3D                         | 1, 1, 32       | 0                    |
| Flatten                              | 32             | 0                    |
| Dense                                | 256            | 8448                 |
| Dropout                              | 256            | 0                    |
| Dense-1(Dense)                       | 128            | 32869                |
| Dropout                              | 128            | 0                    |
| Dense-2(Dense)                       | 4              | 516                  |

#### 4.2.7 Optimal wavelength selection

Hyperspectral data contain a large amount of information that could be highly redundant and multi-collinear within adjacent wavelengths, resulting in computational complexities

during processing and application (Hennessy et al., 2020). Feature mining methods that extract the most relevant and sensitive informative wavelengths from the spectral data are important to reduce multi-collinearity in the data and enhance model robustness. In this work, a correlation-based feature selection (CFS) method was applied to the best group of spectral datasets. CFS is one of the most popular data engineering methods for selecting a set of sensitive features to build a discriminative model for a specific purpose. It works on the principle that good and informative wavelengths are those that are highly correlated with a particular class but uncorrelated with each other. To use the CFS algorithm, a heuristic search algorithm was applied along with a correlation function (Pearson's correlation) to assign high scores and select the best subset of features that had high predictive ability of the class labels with poor correlation with each other. A greedy stepwise (GS) search strategy was applied during the application of the CFS algorithms to select the space attributes of the variable subsets.

#### **4.2.8 Evaluation metrics for model performance**

To evaluate the performance of the proposed model against the classification models, the overall accuracy (OA), the F- Score and the kappa coefficient (Kappa) evaluation metrics were used. OA represents the number of correctly classified samples in the overall test samples, which is given by equation 4.5. The F-score evaluates the accuracy of the model on the entire dataset and is useful when there is an uneven class distribution. It is given by the expression shown in equation 4.6. The kappa coefficient is a statistical measure that describes the mutual information between the ground truth map and the predicted classified map and is expressed mathematically as shown in equation 4.7. Kappa value ranges between 0 and 1, such that a coefficient of 1 means perfect agreement with predicted class and ground truth and vice versa when the coefficient is 0.

$$OA = \frac{\sum_{y=1}^x d_{ii}}{N}$$

**Equation 4.5**

where OA is the overall accuracy, N is the total number of samples, x is the number of class labels present in the dataset,  $d_{ii}$  is the obtained from the diagonal element of the confusion matrix.

$$\text{F-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Equation 4.6}$$

$$\text{Where Precision} = \frac{TP}{TP + FP}, \text{ Recall} = \frac{TP}{TP + FN}$$

Where TP= True positive, FN= False negative and FP= False Positive,  $P_o$  is the

$$K = \frac{P_o - P_c}{1 - P_c}$$

$$\text{Equation 4.7}$$

Where  $P_o$  is the probability of observed agreements and  $P_c$  is the expected agreement.

## 4.3 Results

### 4.3.1 Characteristics of spectral curves for different nutrient treatment

Figures 4.3 and 4.4 represent the average spectral curves for cowpea and quinoa, respectively, at growth stage III. For both species, although the spectral curves are similar in shape, there are visible differences between the four treatments. There is a characteristic peak in the visible spectra observed, especially for wavelengths between 550 nm and 560 nm, for both plant species. This indicates a high amount of chlorophyll absorption in this region (Siedliska et al., 2021b), which shows that chlorophyll could be a responsive feature for the dynamics of N and P in both plants. The region between 600 nm and 700 nm (red band) shows a clear and continuous distinction between some variants of the experiment for both species. The low-nitrogen treatments exhibited higher reflectance in contrast to the high-nitrogen treatments, which had low reflectance in this region. HNLP had the lowest reflectance, while LNLP had the highest reflectance in cowpea. Similarly, LNLP and HNHP had the highest and lowest reflectance in the red-edge region for quinoa. This agrees with Clevers & Gitelson (2013), who demonstrated rapid change in the reflectance of the plant canopy in the red-edge region in response to high levels of nitrogen. The regions between 750 and 1100 nm (near infrared) exhibited the highest differences between the spectral curves of the two species. The HNHP had the highest reflectance in both species at above 0.62 nm and above 0.64 nm for cowpea and quinoa, respectively. This agrees with Zhai et al. (2013), who showed that reflectance



of healthy plants (in this case, HNHP plants) in the NIR region has a strong correlation with the chemical and cellular architecture of plants and exhibits high reflectance. In the SWIR region, there was a characteristic drop in reflectance from 1250 to 1500 nm and a gradual rise and fall in reflectance across the spectra from 1500 to 2000 nm. Moreover, there was comparably low reflectance in the two P- related treatments (HNHP and HNLP) at 1450 nm, which agrees with Siedliska et al., (2021), who suggest that this region has a quantitative relationship between light reflectance and P treatments. From wavelengths 2000 nm to 2500 nm, there was an observable increase in reflectance for both species.

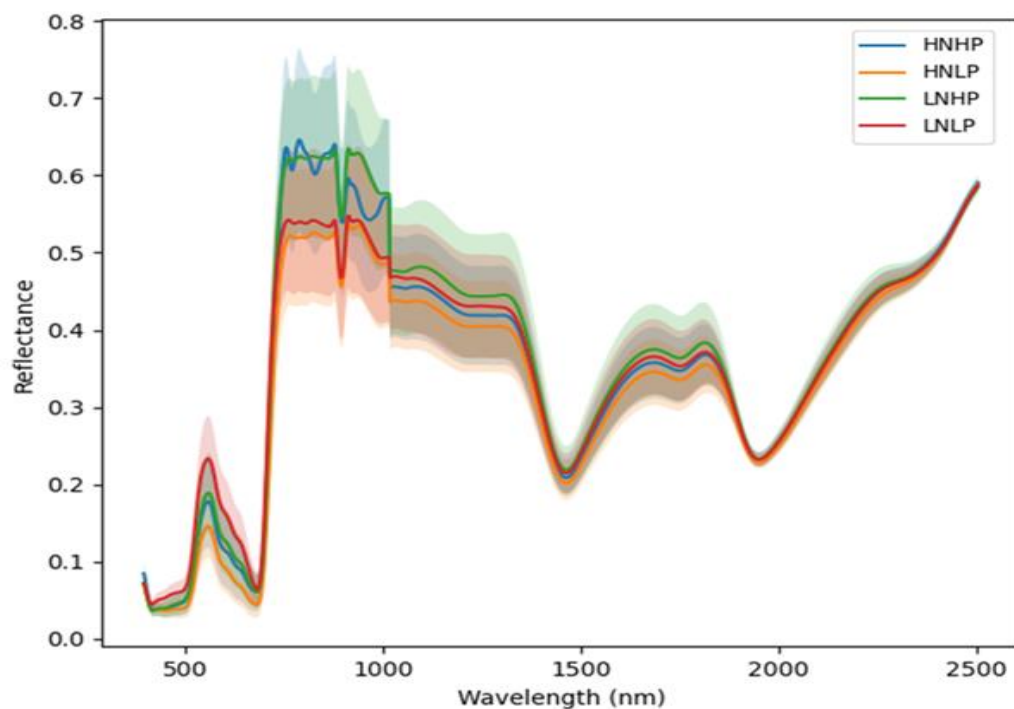


Figure 4-3 Average reflectance spectra for cowpea under different N and P levels obtained at growth stage III. The x-axis represents the wavelengths ranging from approximately 400 nm to 2500 nm, covering the visible to near-infrared (NIR) spectrum whiles the y-axis represents reflectance, indicating how much light is reflected by the plant material at each wavelength. Each line represents the average spectral signature for five plants from each treatment of the experiment.

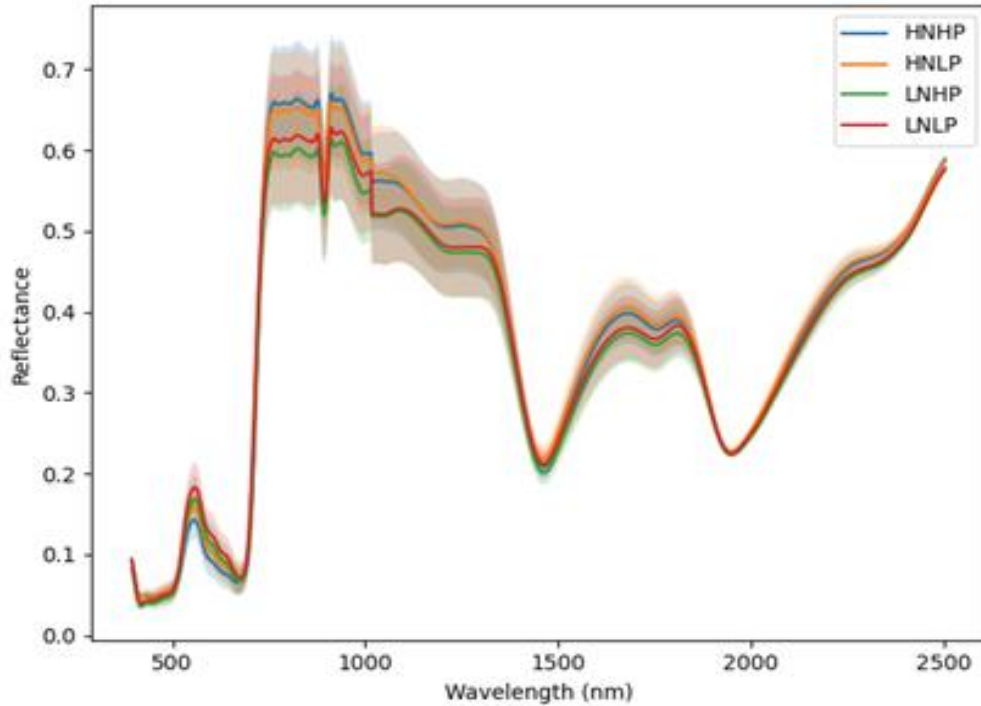


Figure 4-4 Average reflectance spectra for quinoa under different N and P levels obtained at growth stage III. The x-axis represents the wavelengths ranging from approximately 400 nm to 2500 nm, covering the visible to near-infrared (NIR) spectrum whilst the y-axis represents reflectance, indicating how much light is reflected by the plant material at each wavelength.

Each line represents the average spectral signature for five plants from each treatment of the experiment.

#### 4.3.2 Model performance based on different pre-processing methods

Figures 4.5 and 4.6 are box and whisker plots summarizing the accuracy distribution scores for the different models of the two species. Tables 4.3, 4.4 summarize the performance of each model based on the evaluation metrics used in this study. From Figure 4.5, the pre-processing methods influenced the performance of the different models. SDeri datasets had higher accuracies in all four models, achieving the highest OA (99.24%), kappa (98.64%) and F-score (99.19%) for cowpea and 99.18% OA, 98.85% kappa and 98.76% F-score for quinoa. The high OA and F-score show the good generalization capacity of the models in classifying the nutrient status. In contrast, the RAW dataset produced the lowest performance for all the models, with the hybrid CNN model achieving the highest performance for cowpea (98.39%, 98.37% and 97.81% OA, F-score and kappa coefficient, respectively). In quinoa, the HybridSN model achieved the

best performance when trained on the raw dataset with 98.57%, 98.43% and 97.76% OA, F-score and kappa coefficient, respectively. On the contrary, the 2D-CNN model (trained with raw dataset) had the lowest performance with 92.77% OA for cowpea. Furthermore, the 2D-CNN model had the lowest performance (91.15% OA) when trained with the LDA dataset for quinoa. There was a 3.58% and 1.57% reduction in accuracy performances when the hybrid model was trained with the SNV and LDA datasets, respectively, for cowpea as compared to the performance when trained with the SDeri dataset. Similarly, the proposed model decreased in performance when trained with the SNV and LDA datasets, achieving 96.32% and 91.45% OA scores, respectively, for quinoa. This suggests that the SDeri dataset had important learnable wavelengths to discriminate the treatments for both species. Although the LDA dataset had reduced training samples, all the LDA-based models displayed low performance compared to the SDeri dataset-trained model's performance.

The proposed model for both species exhibited good performance when trained with SNV- and SDeri-transformed datasets. Although the HybridSN and 3D CNN models had good performances when trained with SDeri datasets, their training time (caused by the high number of trainable parameters) made the implementation of this model a challenge. The proposed hybrid CNN model had reduced trainable parameters ( $< 200,000$ ), which subsequently affected the computational processing time. The SDeri-trained hybrid CNN model had a processing time reduced by over 1000 s as compared to the 3DCNN-SDeri model, which had a 2141.63 s processing time for cowpea. For quinoa, the SDeri-hybrid CNN model had about 54.16% reduced training time when compared to the 3D CNN model (trained with a raw dataset, which had the highest training time of over 2000 s). For both species, the HybridSN-based models had a significantly high training time irrespective of the training dataset.

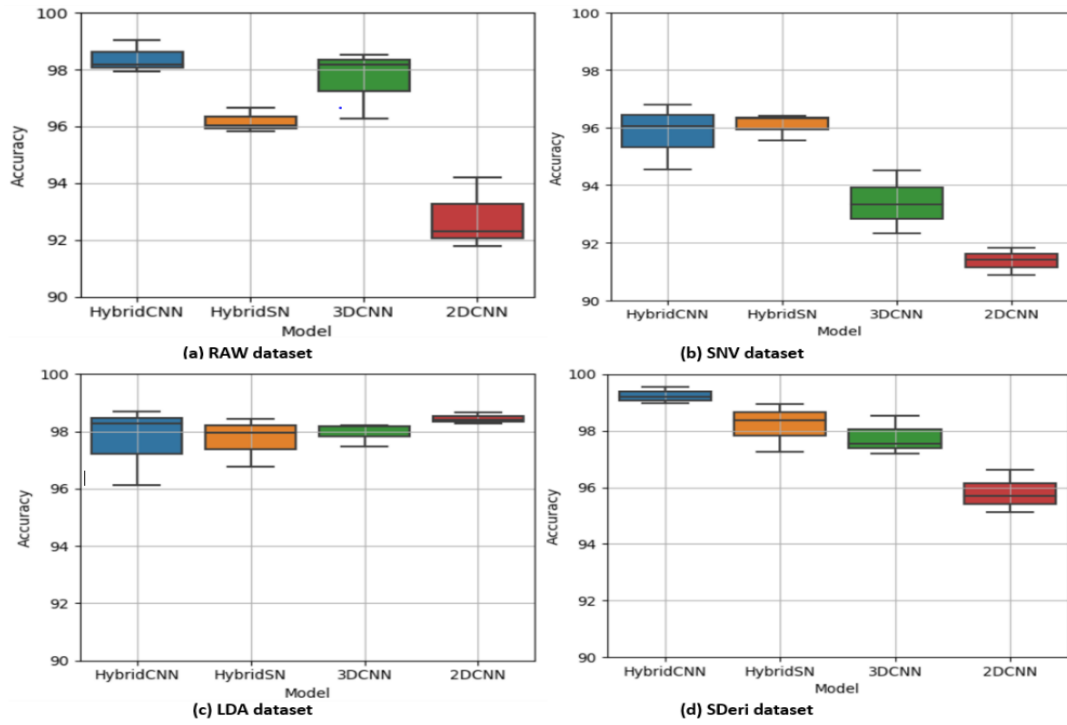


**Figure 4-5** Box and whisker plot of classification accuracy scores for cowpea for the four models: Hybrid-CNN, 3DCNN, 2D CNN and HybridSN. (A) Raw dataset, (B) SNV dataset, (C) LDA dataset, (D) Hybrid CNN dataset

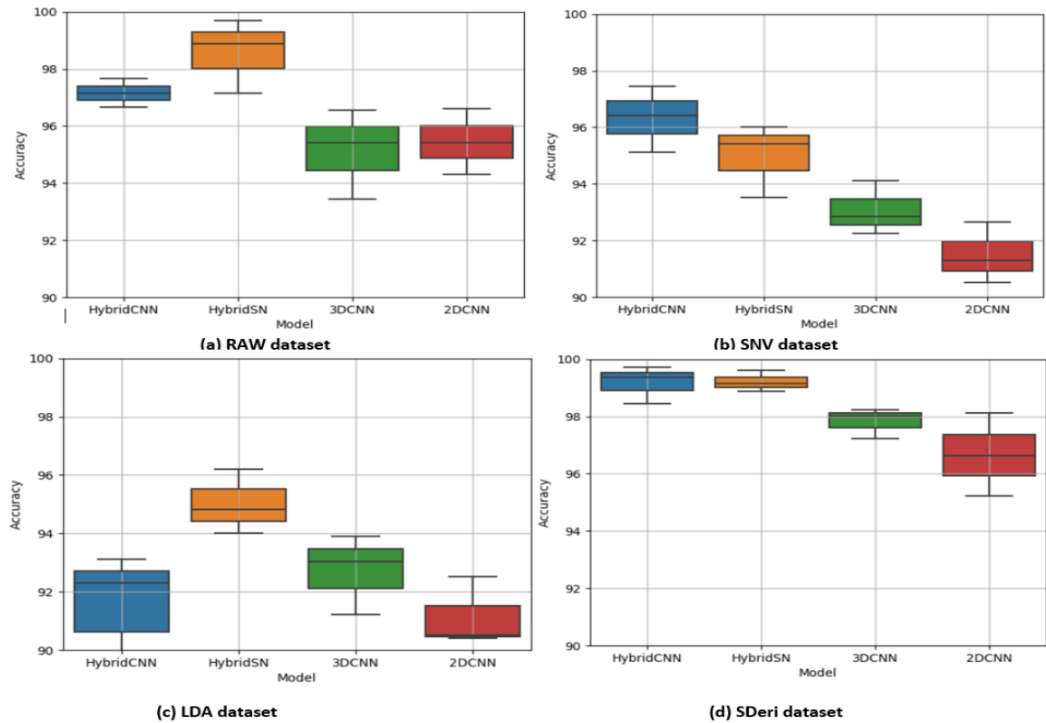


**Figure 4-6** Box and whisker plot of classification accuracy scores for quinoa for the four algorithms: Hybrid-CNN, 3DCNN, 2D CNN and HybridSN. (A) Raw dataset, (B) SNV dataset, (C) LDA dataset, (D) Hybrid CNN dataset

| Models     | Quinoa                   |               |               |                |              |
|------------|--------------------------|---------------|---------------|----------------|--------------|
|            | Pre- processing datasets |               |               |                |              |
|            | Metrics                  | RAW           | SDeri         | SNV            | LDA          |
| Hybrid-CNN | OA                       | 97.15±0.51    | 99.18±0.66    | 96.33±1.15     | 91.45±2.22   |
|            | F-Score                  | 96.85±0.26    | 98.85±0.26    | 95.79±2.17     | 91.46±1.44   |
|            | Kappa                    | 96.49±0.44    | 98.76±0.65    | 95.88±1.28     | 90.68±2.13   |
|            | Training Time(s)         | 1021.06±75.44 | 1017.93±29.98 | 1033.03±35.25  | 252.43±12.24 |
| 3DCNN      | OA                       | 95.14±1.58    | 97.82±0.531   | 93.07±0.96     | 92.71±1.36   |
|            | F-Score                  | 95.36±0.49    | 95.36±0.49    | 92.51±1.82     | 91.90±1.29   |
|            | Kappa                    | 94.57±2.04    | 97.05±1.70    | 91.603±1.68    | 91.33±2.11   |
|            | Training Time(s)         | 2274.73±14.88 | 2220.80±51.00 | 23104.96±42.84 | 983.16±15.34 |
| 2DCNN      | OA                       | 95.44± 0.56   | 96.657±1.445  | 91.49±1.08     | 91.15±1.18   |
|            | F-Score                  | 94.27±1.811   | 94.27±1.811   | 90.84±1.34     | 90.59±1.00   |
|            | Kappa                    | 93.68±1.636   | 95.23±0.94    | 90.43±1.03     | 90.38±2.60   |
|            | Training Time(s)         | 498.51±18.06  | 456.30±75.32  | 672.96±55.09   | 119.06±61.99 |
| HybridSN   | OA                       | 98.57±1.30    | 99.01±0.38    | 94.98±1.31     | 95.01±1.12   |
|            | F-Score                  | 98.43±0.79    | 98.47±0.79    | 94.46±1.91     | 94.86±1.34   |
|            | Kappa                    | 97.76±0.85    | 98.96±0.22    | 94.05±1.50     | 94.87±2.19   |
|            | Training Time(s)         | 1845.40±49.75 | 1874.86±25.31 | 1841.50±18.75  | 523.00±11.13 |

**Table 4-3**

Classification  
accuracy assessment  
and computation cost  
with different models  
and pre-processing  
methods for quinoa

| Model      | Cowpea                 |               |               |               |               |
|------------|------------------------|---------------|---------------|---------------|---------------|
|            | Pre-processing Methods |               |               |               |               |
|            | Metrics                | RAW           | SDeri         | SNV           | LDA           |
| Hybrid-CNN | OA                     | 98.39±0.59    | 99.24±0.29    | 95.81±1.38    | 97.70±1.38    |
|            | F-Score                | 98.37±0.26    | 99.19±0.33    | 94.46±1.49    | 97.32±0.20    |
|            | Kappa                  | 97.81±1.10    | 98.64±0.11    | 96.73±1.82    | 96.58±1.16    |
|            | Training Time(s)       | 971.37±8.01   | 950.60 ±4.23  | 964.01±1.85   | 403.97±16.72  |
| 3DCNN      | OA                     | 97.65±1.22    | 97.76±0.69    | 93.40±1.09    | 97.96±0.41    |
|            | F-Score                | 98.44±0.27    | 97.92±0.29    | 94.19±1.81    | 99.21±0.32    |
|            | Kappa                  | 98.10±0.56    | 98.63±0.41    | 94.02±1.59    | 99.36±0.07    |
|            | Training Time(s)       | 2141.63±14.32 | 2190.13±27.17 | 2158.40±73.89 | 1017.50±79.53 |
| 2DCNN      | OA                     | 92.77±1.267   | 95.82±0.74    | 91.38±0.481   | 98.45±0.18    |
|            | F-Score                | 93.653±1.580  | 94.59±0.73    | 91.48±0.63    | 97.75±0.48    |
|            | Kappa                  | 92.393±0.890  | 99.32±0.36    | 91.75±2.44    | 96.93±0.29    |
|            | Training Time(s)       | 774.00±1.85   | 851.03±32.63  | 872.30±15.99  | 398.57±41.27  |
| HybridSN   | OA                     | 96.16±0.435   | 98.19±0.85    | 96.09±0.47    | 97.73±0.85    |
|            | F-Score                | 95.99±0.15    | 98.76±0.55    | 95.81±1.83    | 98.65±0.52    |
|            | Kappa                  | 96.04±0.78    | 98.48±0.76    | 94.75±2.73    | 98.42±0.12    |
|            | Training Time(s)       | 1080.00±23.51 | 1049.53±18.32 | 1038.53±55.28 | 622.67± 23.05 |

**Table 4-4** Classification accuracy assessment and computation cost with different models and pre-processing methods for cowpea

### 4.3.3 Effective wavelength selection

Although the proposed model with the SDeri dataset produced the best performance in classifying plant nutrient status, the high dimensionality of the hypercube posed computational challenges. Hence, effective wavelengths that accurately distinguish between the N and P levels were selected and used to retrain the models. From Figure 4.7, the selected wavelengths for cowpea were localized in the blue, red-edge, near-infrared and short-wave infrared regions (having the highest number of selected wavelengths). The second derivative reflectance values exhibit minimal variations among treatments in the visible spectrum (390 nm- 700 nm), indicating comparable amounts of chlorophyll and other pigments across nutritional circumstances. Additionally, in the infrared regions (700 nm -1300 nm), a sharp peak is observed around the 1000 nm, with some variations among the treatments, particularly between LNLP and the other groups. Moreover, the HNHP treatment exhibits slightly higher reflectance as compared to the other treatments. The shortwave infrared regions (1300 nm- 2500 nm) exhibited increased peaks and troughs, especially around the 1400 nm, 1900 nm, and 2200 nm. The HNHP and HNLP treatments generally showed a high reflectance values whilst the LNLP showed lower reflectance values. Sixteen (16) wavebands were selected as the most sensitive wavelengths, with three selected from the blue regions (411, 431 and 455 nm), two from the red-edge regions (683 and 691 nm), four from the near-infrared regions (791, 871, 919 and 923 nm) and seven from the SWIR regions (1438, 1450, 1461, 1485, 1508, 1602 and 1660 nm).

In quinoa, the wavelengths were selected within the blue, near infrared and SWIR regions. From Figure 4.8, twelve (12) wavebands were selected. These include four from the blue region (571, 575, 583 and 599 nm), one from the red spectrum (607 nm), four from the near-infrared regions (723, 731, 924 and 967 nm) and three from the SWIR regions (1680 nm, 1475 nm and 1752 nm). The differences in the selected wavebands could be due to the variability in the plant structure and the changes in the nutrient composition of the plants.

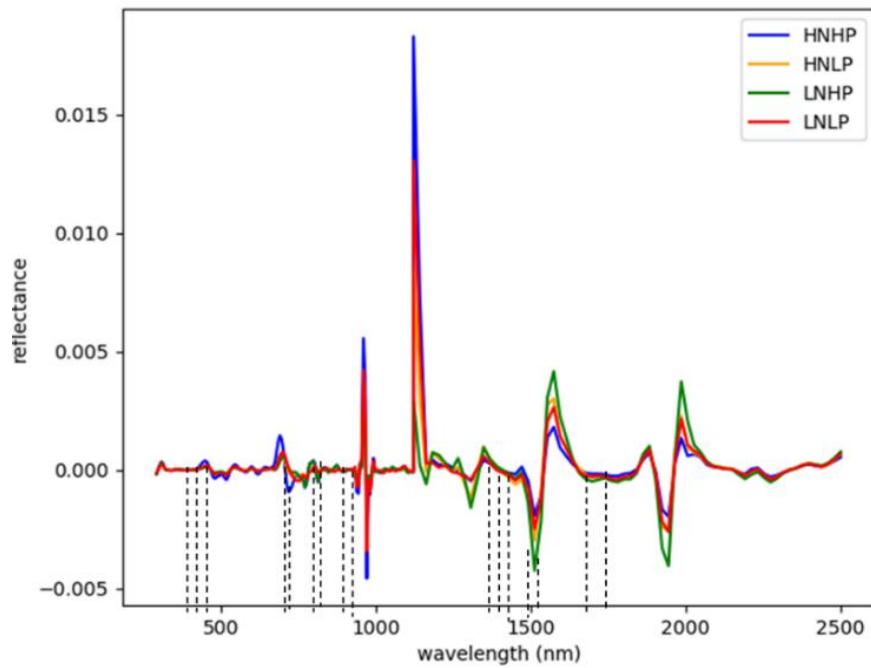


**Figure 4-7** Plot displaying the second derivate reflectance values over a wavelength range from approximately 390 nm to 2500 nm. The vertical black dashed lines are marked at specific wavelength positions, indicating selected wavelengths for cowpea species. The wavelength selection was done using CFS feature selection method.

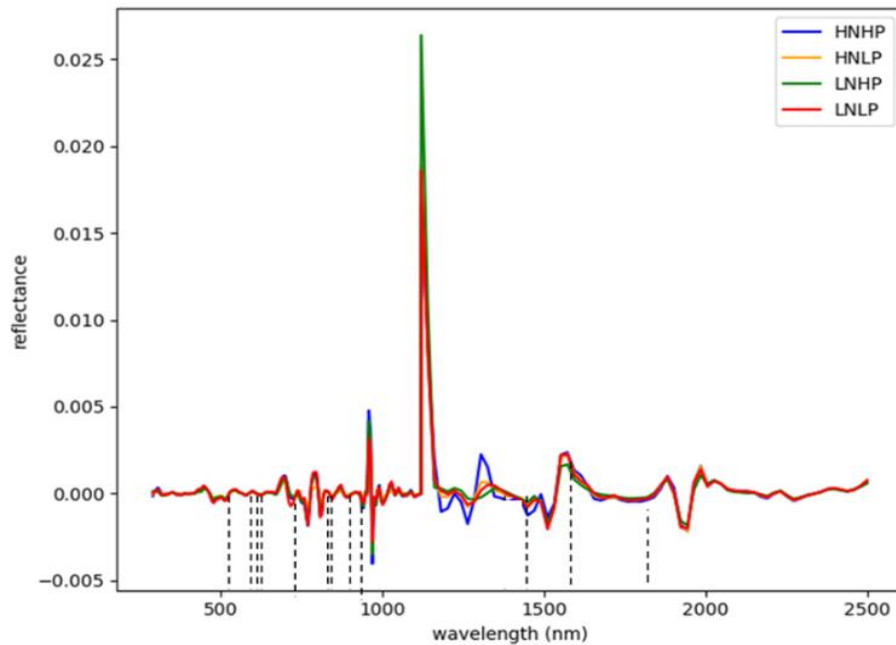


**Figure 4-8** Plot displaying the second derivate reflectance values over a wavelength range from approximately 390 nm to 2500 nm. The vertical black dashed lines are marked at specific wavelength positions, indicating selected wavelengths for quinoa species. The wavelength selection was done using CFS feature selection method..



#### **4.3.4 Results of discrimination analysis at different growth stages**

After the waveband's selection, the full plant (with full spatial resolution and selected spectral wavelength) in the second derivative format was used as a dataset to retrain the different models to predict the plants' N and P status at different growth stages. Datasets were generated using the same approach as that in subsection 4.2.4.2. The training datasets were augmented to artificially enlarge the number of training images using rotation and flipping methods.

Tables 4.5 and 4.6 show the performance of the models when trained with the selected wavelengths for cowpea and quinoa, respectively. All the models performed well (with 79-100% accuracy on the test data) irrespective of the growth stage under consideration. This performance was better than that found in Siedliska et al (2021), where a 40-100% classification accuracy rate was achieved for predicting different P levels of three plant species at five growth stages. Comparing the four models, the proposed hybrid model had the best performance, attaining above 94% and 97% test OA scores across the selected growth stages for cowpea and quinoa, respectively. The 2DCNN model had the highest misclassification, with 15.92% in cowpea at growth stage IV and 20.77% in quinoa at growth stage II in the test dataset. For each species, there were high classification errors at the early growth stages, especially in the LNHP and HNLP plants. However, as the plants developed, there were clear distinctions between the spectral signatures of the different treatments, which resulted in an increase in the classification accuracy.

|              | Plant      | Cowpea               |            |             |                  |            |              |
|--------------|------------|----------------------|------------|-------------|------------------|------------|--------------|
|              |            | Training dataset (%) |            |             | Test dataset (%) |            |              |
| Growth stage | Model      | Accuracy             | F-Score    | Kappa-Score | Accuracy         | F-Score    | Kappa-Score  |
| I            | Hybrid CNN | 99.12±0.10           | 99.61±0.08 | 99.35±0.27  | 98.51±0.25       | 98.88±0.25 | 98.44±0.31   |
|              | 3D-CNN     | 96.21±0.41           | 96.08±0.21 | 96.02±0.13  | 93.27±0.66       | 93.12±0.76 | 93.04±0.0.44 |
|              | 2D-CNN     | 93.25±0.14           | 93.21±0.56 | 93.17±0.25  | 89.21±0.17       | 88.71±0.17 | 89.21±0.37   |
|              | HybridSN   | 98.43±0.21           | 98.02±0.35 | 98.44±0.13  | 95.27±0.89       | 96.05±0.73 | 96.68±0.11   |
| II           | Hybrid CNN | 99.26±0.50           | 99.05±0.25 | 99.08±0.21  | 99.48±0.56       | 99.25±0.33 | 99.51±0.62   |
|              | 3D-CNN     | 97.15±0.10           | 97.22±0.16 | 97.21±0.53  | 94.11±0.06       | 94.41±0.25 | 94.05±0.81   |
|              | 2D-CNN     | 95.00±0.13           | 95.16±0.12 | 95.33±0.26  | 91.13±0.26       | 91.22±0.17 | 91.05±0.11   |
|              | HybridSN   | 99.15±0.33           | 99.34±0.49 | 99.04±0.54  | 98.22±0.71       | 97.15±0.05 | 97.01±0.24   |
| III          | Hybrid CNN | 99.54±0.24           | 99.42±0.11 | 99.63±0.12  | 98.68±0.22       | 98.41±0.34 | 98.22±0.21   |
|              | 3D-CNN     | 97.41±0.15           | 97.51±0.24 | 97.25±0.84  | 93.74±0.41       | 93.51±0.25 | 93.41±0.45   |
|              | 2D-CNN     | 96.28±0.64           | 96.12±0.85 | 99.18±0.15  | 92.19±0.37       | 91.87±0.14 | 91.59±0.26   |
|              | HybridSN   | 99.89±0.27           | 99.22±0.13 | 99.45±0.13  | 97.44±0.71       | 97.42±0.44 | 97.29±0.44   |
| IV           | Hybrid CNN | 99.73±0.40           | 99.64±0.27 | 99.11±0.05  | 98.92±0.45       | 98.59±0.24 | 98.64±0.23   |
|              | 3D-CNN     | 98.24±0.19           | 99.26±0.16 | 99.82±0.14  | 95.23±0.67       | 92.25±0.53 | 95.22±0.31   |
|              | 2D-CNN     | 98.18±0.25           | 98.13±0.29 | 98.32±0.22  | 84.08±0.81       | 84.22±0.58 | 84.55±0.67   |
|              | HybridSN   | 98.51±0.22           | 98.17±0.31 | 98.25±0.35  | 86.44±0.7        | 86.02±0.51 | 86.58±0.74   |
| V            | Hybrid CNN | 99.03±0.12           | 98.14±0.32 | 98.07±0.45  | 98.35±0.22       | 97.18±0.22 | 97.34±0.25   |
|              | 3D-CNN     | 95.95±0.29           | 98.58±0.44 | 98.61±0.52  | 88.42±0.21       | 88.64±0.25 | 88.21±0.67   |
|              | 2D-CNN     | 90.28±0.41           | 90.44±0.25 | 90.11±0.23  | 85.22±0.19       | 85.23±0.58 | 85.44±0.10   |
|              | HybridSN   | 98.66±0.44           | 96.35±0.28 | 96.46±0.15  | 92.42±0.16       | 92.25±0.44 | 92.41±0.31   |

**Table 4-5**

Model performance on selected wavebands for classification of cowpea N and P levels at five growth stages

| Plant        |            | Quinoa               |            |             |                  |            |             |
|--------------|------------|----------------------|------------|-------------|------------------|------------|-------------|
|              |            | Training dataset (%) |            |             | Test dataset (%) |            |             |
| Growth stage | Model      | Accuracy             | F-Score    | Kappa-Score | Accuracy         | F-Score    | Kappa-Score |
| I            | Hybrid CNN | 99.12±0.26           | 99.03±0.51 | 98.97±0.54  | 95.34±0.21       | 94.21±0.44 | 94.51±0.52  |
|              | 3D-CNN     | 96.55±0.11           | 96.65±0.15 | 96.02±0.45  | 90.31±0.12       | 90.32±0.22 | 90.45±0.31  |
|              | 2D-CNN     | 92.38±0.50           | 91.43±0.55 | 91.25±0.43  | 85.81±0.65       | 86.24±0.24 | 84.21±0.56  |
|              | HybridSN   | 97.21±0.25           | 97.02±0.35 | 97.42±0.11  | 91.65±0.33       | 90.15±0.40 | 90.16±0.25  |
| II           | Hybrid CNN | 99.08±0.22           | 99.32±0.20 | 99.01±0.21  | 97.21±0.32       | 97.44±0.14 | 96.15±0.22  |
|              | 3D-CNN     | 94.22±0.35           | 94.39±0.37 | 98.44±0.33  | 85.44±0.15       | 85.48±0.34 | 85.07±0.45  |
|              | 2D-CNN     | 97.01±0.18           | 96.56±0.25 | 96.89±0.61  | 79.23±0.26       | 78.14±0.29 | 79.11±0.37  |
|              | HybridSN   | 96.55±0.16           | 96.23±0.32 | 96.34±0.20  | 85.15±0.22       | 85.33±0.32 | 85.01±0.43  |
| III          | Hybrid CNN | 99.58±0.28           | 99.04±0.05 | 99.63±0.05  | 98.55±0.44       | 98.13±0.22 | 97.31±0.32  |
|              | 3D-CNN     | 98.11±0.31           | 97.89±0.33 | 98.03±0.23  | 92.14±0.42       | 92.27±0.41 | 92.18±0.37  |
|              | 2D-CNN     | 97.25±0.28           | 97.12±0.47 | 97.18±0.28  | 86.23±0.34       | 86.27±0.37 | 85.89±0.2   |
|              | HybridSN   | 96.89±0.12           | 96.63±0.45 | 96.55±0.21  | 85.41±0.29       | 85.25±0.22 | 85.09±0.29  |
| IV           | Hybrid CNN | 99.29±0.44           | 99.18±0.51 | 99.23±0.25  | 98.29±0.29       | 98.80±0.27 | 97.74±0.18  |
|              | 3D-CNN     | 98.34±0.35           | 98.22±0.53 | 98.42±0.31  | 90.43±0.11       | 90.55±0.23 | 90.3±0.36   |
|              | 2D-CNN     | 96.81±0.36           | 96.43±0.25 | 96.32±0.14  | 84.18±0.45       | 82.33±0.15 | 82.45±0.22  |
|              | HybridSN   | 96.45±0.42           | 96.17±0.28 | 96.44±0.28  | 80.34±0.31       | 80.22±0.44 | 80.45±0.43  |
| V            | Hybrid CNN | 97.72±0.37           | 97.54±0.14 | 97.28±0.25  | 95.18±0.26       | 95.23±0.58 | 95.27±0.12  |
|              | 3D-CNN     | 96.05±0.22           | 95.78±0.33 | 96.11±0.37  | 89.32±0.12       | 89.53±0.24 | 89.11±0.31  |
|              | 2D-CNN     | 95.44±0.12           | 95.14±0.16 | 9.31±0.42   | 82.15±0.41       | 82.43±0.15 | 82.18±0.36  |
|              | HybridSN   | 96.11±0.58           | 96.52±0.49 | 96.34±0.44  | 92.77±0.55       | 80.05±0.33 | 81.26±0.23  |

**Table 4-6** Model performance on selected wavebands for classification of quinoa N and P levels at five growth stages

Table 4-7 is the summary of the confusion matrix for the proposed model, which explains the accuracy and misclassification of the individual classes for cowpea and quinoa. For both species, the model accurately classified HNHP treatment across all the five growth stages. The model had difficulties in classifying some variants of the treatments, especially those with low phosphorus. In cowpea, HNLP (5% error) and LNHP (> 10% error) were misclassified at growth stage I. Moving to growth stage II, although a similar trend occurred, there was an improvement in the classification accuracy of LNHP (3.64% increment). In the subsequent growth stages, the model improved its performance, with > 95% at growth stages III, IV and V. In quinoa, there was a high accuracy in the HNHP and HNLP treatments. However, the model had lower performance in correctly classifying the LNHP treatment 92.82%. Moving across the growth spectrum, the model performance increased, achieving above 98% classification for HNHP and HNLP at growth stages IV and V.

**Table 4-7** Summary of confusion matrices created for the proposed model for nutrient status identification at five stages of plant growth.

| Plant  | Treatment | I     | II    | III   | IV    | V     |
|--------|-----------|-------|-------|-------|-------|-------|
| Cowpea | HNHP      | 98.32 | 99.57 | 99.73 | 99.41 | 98.92 |
|        | HNLP      | 95.00 | 94.00 | 97.32 | 98.52 | 96.06 |
|        | LNHP      | 86.84 | 90.50 | 96.30 | 98.50 | 95.31 |
|        | LNLP      | 96.21 | 98.00 | 99.40 | 99.15 | 98.23 |
| Quinoa | HNHP      | 97.15 | 99.41 | 99.23 | 99.10 | 98.65 |
|        | HNLP      | 90.22 | 95.12 | 98.01 | 99.27 | 96.14 |
|        | LNHP      | 92.82 | 97.21 | 95.13 | 99.27 | 98.88 |
|        | LNLP      | 98.52 | 96.75 | 98.31 | 98.58 | 98.25 |

#### 4.4 Discussion

This study aimed to develop a deep learning based HSI pipeline for identifying the N and P status in cowpea and quinoa plant species at selected growth stages. This was achieved by developing a hybrid 3D-2D CNN model to automatically learn and evaluate the spectral and spatial characteristics of the canopy components of the species. Previous

studies utilized 1D-, 2D- and 3D-CNN models for hyperspectral image analysis. However, these models have limitations in terms of their feature extraction methods, which impact their performance. While 1D-CNN extracts only spectral information and 2D-CNN deals with spatial information, 3D-CNN extracts both spectral and spatial information simultaneously. Nonetheless, the complexity introduced by 3D CNN modeling could adversely affect its performance and output. Therefore, it is prudent to develop a model that combines 3D- and 2D-CNN architectures to learn features and patterns from spectral hypercube data while decreasing the computational complexities and processing time.

In this study, patches of hypercube (15 X 15 X 223) were extracted from the plant canopy with different N and P levels and used as an input to train the proposed model. Four different spectral transformation techniques were experimented to select the one with the most discriminative features for classifying plant nutrient levels. The performance of the models were compared with standard 3D-CNN, 2D-CNN and HybridSN models (Section 4.3.2). As shown in Table 3 (section 4.3.2), the proposed method achieved an accuracy of > 94% irrespective of the pre-processing technique used. However, the models trained with the second derivative spectra (SDeri) outperformed the others. The derivative processing reduced the background signal and image artifacts, which subsequently improved the discriminating power of the dataset. The high performance of the SDeri-based models agrees with the findings of Siedliska et al. (2021) and Luo et al. (2019), who achieved similar results using derivative spectra. Additionally, although LDA achieved comparably low classification accuracy, it improved the interpretability of the spectra information by replacing the original variables with a group of discriminants while preserving their original information.

A greedy stepwise CFS technique was applied to select the most informative set of wavelengths across the various spectral bands using the second derivative information of the spectral data. The second derivative spectral data improves the feature selection process by providing a clearer and more robust view of meaning spectral variations, reducing noise and baseline effects and enabling the identification of subtle but significant changes. The variations in the treatments in the infrared regions could be attributed to the sensitivity of this region to leaf cell structure and water content, suggesting that nutrient availability of plants may have influenced these traits. The differences in reflectance

around the SWIR regions suggest that nutrient treatments may influence water content in plant tissues. The little variation in second derivative reflectance values among treatments in the visible regions (400 nm -700 nm) and the selection of a few spectral bands for both plants indicates similar levels of chlorophyll and other pigments across nutrient that affects the nutrient composition of the plants. This agrees with Siedliska et al., (2021b) who argues that phosphorus nutrient supply does not significantly affect the reflectance of some sections of the visible spectrum (500 -600 nm). The selection of features around the infrared and SWIR could have been influenced by the phosphorus dynamics in both plants. Zhang et al., (2023) developed an optimal prediction model for estimating plant phosphorus concentration using machine learning. Their results showed that P deficiency increased leaf reflectance in the visible region while decreasing the reflectance in the infrared regions. Siedliska et al., (2021b) also identified features in the SWIR regions as sensitive to phosphorus while the infrared regions are noted for nitrogen level variations. In a similar work, Omid et al., (2020) developed an optimized hyperspectral band selection pipeline for nitrogen status detection in grape leaves. They identified 825 (near infrared), 1386 nm and 1619 nm (SWIR regions) as very sensitive to nitrogen dynamics. Marang et al., (2021) worked on utilizing machine learning to identify combinations of hyperspectral features to detect nitrogen (N) concentration in cotton canopies. Their feature selection method identified the red-edge inflection point around 705 nm to be critical for differentiating between N levels of the cotton plants. The above works agree with this study suggesting that the red-edge, infrared, and SWIR regions are highly sensitive to nitrogen and phosphorus nutrition, respectively.

The selected wavebands were used to retrain the models to identify the plant nutrient status at different growth stages. The proposed method achieved the highest performance, with over 94% accuracy for both species. Although the HybridSN and 3D-CNN models performed well comparably, the high training time limited their practical application. Practically, the selected wavebands can be used to develop a multispectral imaging system to predict plant nutrient status. It should also be noted that all the models experienced a decrease in accuracy when classifying nutrient status at the early growth stages, particularly for the LNHP and HNLP treatments. The decrease in accuracy could be attributed to the fact that the differences in canopy spectral properties were considerably minimal between the four treatments at the early growth stage. As a result,

it became difficult for the models to accurately differentiate the nutrient statuses based on spectral information alone during these early growth stages. This is challenging because, practically, farmers and plant breeders are more interested in identifying plant nutrient status at the early growth stage for proper nutrient management. Hence, further studies are required on the dynamic response of crop canopy to subtle changes in N and P concentration at the early growth stages using hyperspectral data.

## **4.5 Conclusion**

Hyperspectral data hold significant potential for monitoring the N and P nutrition of quinoa and cowpea, enabling the provision of optimal conditions for development and growth. This study proposed the use of hyperspectral imaging in tandem with a hybrid 3D-2D CNN model to identify the nutrient status of cowpea and quinoa at selected growth stages. The experiment results presented demonstrate the capability of the proposed model to accurately distinguish plant nitrogen and phosphorus levels, based on selected wavebands from the second-order derivative of the reflectance spectra. The use of separable convolutions in the 2D-CNN block of the proposed model reduced the model's complexity by reducing the number of trainable parameters. This subsequently reduced the processing time while enhancing learning efficiency, which is advantageous. These findings suggest that the proposed model could be integrated into a system for the non-invasive detection of nitrogen and phosphorus deficiencies in crops. Moreover, the success of the waveband selection process shows the potential of developing a multispectral sensor system equipped with the selected wavebands as a viable alternative to hyperspectral imaging for nutrient stress detection. These findings highlight the potential of the proposed model for the early detection and precise management of nutrient stress in cowpea and quinoa plants.

## **Acknowledgements**

This research was supported by the OCP S.A. under the University of Mohammed VI Polytechnic, Rothamsted Research and Cranfield University project (FPO4). Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) through the Delivering Sustainable Wheat program (BB/X011003/1)

## 4.6 References

- Amigo, J. M. (2010). Practical issues of hyperspectral imaging analysis of solid dosage forms. *Analytical and Bioanalytical Chemistry*, 398(1), 93–109. <https://doi.org/10.1007/s00216-010-3828-z>
- Barnes, R. J., M. S. Dhanoa, and S. J. L. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5), 772–777. <https://doi.org/doi.org/10.1366/00037028942022>
- Clevers, J. G. P. W., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, 23(1), 344–351. <https://doi.org/10.1016/j.jag.2012.10.008>
- Cloutis, E. A., Connery, D. R., Dover, F. J., & Major, D. J. (1996). Airborne multi-spectral monitoring of agricultural crop status: Effect of time of year, crop type and crop condition parameter. *International Journal of Remote Sensing*, 17(13), 2579–2601. <https://doi.org/10.1080/01431169608949094>
- Dyrmann, M., Karstoft, H., & Midtby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*, 151(2005), 72–80. <https://doi.org/10.1016/j.biosystemseng.2016.08.024>
- Firat, H., Asker, M. E., & Hanbay, D. (2022). Classification of hyperspectral remote sensing images using different dimension reduction methods with 3D/2D CNN. *Remote Sensing Applications: Society and Environment*, 25(December 2021), 100694. <https://doi.org/10.1016/j.rsase.2022.100694>
- Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors (Switzerland)*, 17(9). <https://doi.org/10.3390/s17092022>
- Gao, Y., Song, X., Liu, K., Li, T., Zheng, W., Wang, Y., Liu, Z., Zhang, M., Chen, Q., Li, Z., Li, R., Zheng, L., Liu, W., & Miao, T. (2021). Mixture of controlled-release and conventional urea fertilizer application changed soil aggregate stability, humic acid molecular composition, and maize nitrogen uptake. *Science of the Total Environment*, 789, 147778. <https://doi.org/10.1016/j.scitotenv.2021.147778>
- Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. *Journal of Machine Learning Research*, 15, 315–323.
- Hairuddin, M. A., Tahir, N. M., & Shah Baki, S. R. (2011). Representation of Elaeis Guineensis nutrition deficiency based on image processing approach. *ICCAIE 2011 - 2011 IEEE Conference on Computer Applications and Industrial Electronics, Iccae*, 607–611. <https://doi.org/10.1109/ICCAIE.2011.6162206>



- Hassanijalilian, O., Igathinathane, C., Doetkott, C., Bajwa, S., Nowatzki, J., & Haji Esmaeili, S. A. (2020). Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Computers and Electronics in Agriculture*, 174(April), 105433. <https://doi.org/10.1016/j.compag.2020.105433>
- Hennessy, A., Clarke, K., & Lewis, M. (2020). Hyperspectral classification of plants: A review of waveband selection generalisability. *Remote Sensing*, 12(1). <https://doi.org/10.3390/RS12010113>
- Hong, Z., & He, Y. (2020). Rapid and nondestructive discrimination of geographical origins of longjing tea using hyperspectral imaging at two spectral ranges coupled with machine learning methods. *Applied Sciences (Switzerland)*, 10(3), 1–12. <https://doi.org/10.3390/app10031173>
- Huang, L., Luo, R., Liu, X., & Hao, X. (2022). Spectral imaging with deep learning. *Light: Science and Applications*, 11(1). <https://doi.org/10.1038/s41377-022-00743-6>
- Kumar, B., Dikshit, O., Gupta, A., & Singh, M. K. (2020). Feature extraction for hyperspectral image classification: a review review. *International Journal of Remote Sensing*, 41(16), 6248–6287. <https://doi.org/10.1080/01431161.2020.1736732>
- Luo, Z., Thorp, K. R., & Abdel-Haleem, H. (2019). A high-throughput quantification of resin and rubber contents in *Parthenium argentatum* using near-infrared (NIR) spectroscopy. *Plant Methods*, 15(1), 1–14. <https://doi.org/10.1186/s13007-019-0544-3>
- Ma, D., Maki, H., Neeno, S., Zhang, L., Wang, L., & Jin, J. (2020). Application of non-linear partial least squares analysis on prediction of biomass of maize plants using hyperspectral images. *Biosystems Engineering*, 200, 40–54. <https://doi.org/10.1016/j.biosystemseng.2020.09.002>
- Masters-Clark, E., Shone, E., Paradelo, M., Hirsch, P. R., Clark, I. M., Otten, W., Brennan, F., & Mauchline, T. H. (2020). Development of a defined compost system for the study of plant-microbe interactions. *Scientific Reports*, 10(1), 1–9. <https://doi.org/10.1038/s41598-020-64249-0>
- McCauley, A., Jones, C., & Jacobsen, J. (2011). Plant Nutrient Functions and Deficiency and Toxicity Symptoms. *Nutrient Management Module*, 9(9), 1–16.
- Meier, U., Bleiholder, H., Buhr, L., Feller, C., Hack, H., Heß, M., Lancashire, P., Schnock, U., Stauß, R., Van den Boom, T., Weber, E., & Zwerger, P. (2009). The BBCH system to coding the phenological growth stages of plants-history and publications. *Journal Für Kulturpflanzen*, 61(2), 41–52. <https://doi.org/10.5073/JfK.2009.02.01>
- Mirzaei, S., Van hamme, H., & Khosravani, S. (2019). Hyperspectral image classification using Non-negative Tensor Factorization and 3D Convolutional Neural Networks.

- Signal Processing: Image Communication*, 76(May), 178–185.  
<https://doi.org/10.1016/j.image.2019.05.004>
- Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., & Scheunders, P. (2017). Close range hyperspectral imaging of plants: A review. *Biosystems Engineering*, 164, 49–67.  
<https://doi.org/10.1016/j.biosystemseng.2017.09.009>
- Mohan, A., & Venkatesan, M. (2020a). Chapter 4. HybridCNN based hyperspectral image classification using multiscale spatio-spectral features. *Infrared Physics and Technology*, 108(December 2019). <https://doi.org/10.1016/j.infrared.2020.103326>
- Mohan, A., & Venkatesan, M. (2020b). HybridCNN based hyperspectral image classification using multiscale spatio-spectral features. *Infrared Physics and Technology*, 108(December 2019). <https://doi.org/10.1016/j.infrared.2020.103326>
- Mu, Q., Kang, Z., Guo, Y., Chen, L., Wang, S., & Zhao, Y. (2021a). Chapter 4. Hyperspectral image classification of wolfberry with different geographical origins based on three-dimensional convolutional neural network. *International Journal of Food Properties*, 24(1), 1705–1721.  
<https://doi.org/10.1080/10942912.2021.1987457>
- Mu, Q., Kang, Z., Guo, Y., Chen, L., Wang, S., & Zhao, Y. (2021b). Hyperspectral image classification of wolfberry with different geographical origins based on three-dimensional convolutional neural network. *International Journal of Food Properties*, 24(1), 1705–1721. <https://doi.org/10.1080/10942912.2021.1987457>
- Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., & Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods*, 15(1), 1–10.  
<https://doi.org/10.1186/s13007-019-0479-8>
- Nkemelu, D. K., Omeiza, D., & Lubalo, N. (2018). *Deep Convolutional Neural Network for Plant Seedlings Classification*. <http://arxiv.org/abs/1811.08404>
- Pandey, P., Ge, Y., Stoerger, V., & Schnable, J. C. (2017). High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging. *Frontiers in Plant Science*, 8(August), 1–12. <https://doi.org/10.3389/fpls.2017.01348>
- Paoletti, M. E., Haut, J. M., Plaza, J., & Plaza, A. (2019). ISPRS Journal of Photogrammetry and Remote Sensing Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158(November 2018), 279–317. <https://doi.org/10.1016/j.isprsjprs.2019.09.006>
- Perez-Sanz, F., Navarro, P. J., & Egea-Cortines, M. (2017). Plant phenomics: An overview of image acquisition technologies and image data analysis algorithms. *GigaScience*, 6(11), 1–18. <https://doi.org/10.1093/gigascience/gix092>
- Pi, W., Du, J., Bi, Y., Gao, X., & Zhu, X. (2021). 3D-CNN based UAV hyperspectral

- imagery for grassland degradation indicator ground object classification research. *Ecological Informatics*, 62(March), 101278. <https://doi.org/10.1016/j.ecoinf.2021.101278>
- Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Roy, S. K., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. (2020). HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2), 277–281. <https://doi.org/10.1109/LGRS.2019.2918719>
- Sadeghi-Tehran, P., Virlet, N., & Hawkesford, M. J. (2021). A neural network method for classification of sunlit and shaded components of wheat canopies in the field using high-resolution hyperspectral imagery. *Remote Sensing*, 13(5), 1–17. <https://doi.org/10.3390/rs13050898>
- Sadeghi-Tehran, P., Virlet, N., Sabermanesh, K., & Hawkesford, M. J. (2017). Chapter 2. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. *Plant Methods*, 13(1), 1–17. <https://doi.org/10.1186/s13007-017-0253-8>
- Siedliska, A., Baranowski, P., Pastuszka-Woźniak, J., Zubik, M., & Krzyszczak, J. (2021). Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. *BMC Plant Biology*, 21(1), 1–17. <https://doi.org/10.1186/s12870-020-02807-4>
- Siedliska, A., Baranowski, P., Zubik, M., Mazurek, W., & Sosnowska, B. (2018). Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biology and Technology*, 139(February), 115–126. <https://doi.org/10.1016/j.postharvbio.2018.01.018>
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, 21(2), 110–124. <https://doi.org/10.1016/j.tplants.2015.10.015>
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Computational Intelligence and Neuroscience*, 2016. <https://doi.org/10.1155/2016/3289801>
- Uddling, J., Gelang-Alfredsson, J., Piikki, K., & Pleijel, H. (2007). Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings. *Photosynthesis Research*, 91(1), 37–46. <https://doi.org/10.1007/s11120-006-9077-5>
- Williams, D., Britten, A., McCallum, S., Jones, H., Aitkenhead, M., Karley, A., Loades, K., Prashar, A., & Graham, J. (2017). A method for automatic segmentation and

- splitting of hyperspectral images of raspberry plants collected in field conditions. *Plant Methods*, 13(1), 1–12. <https://doi.org/10.1186/s13007-017-0226-y>
- Xia, C., Yang, S., Huang, M., Zhu, Q., Guo, Y., & Qin, J. (2019). Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis. *Infrared Physics and Technology*, 103(July), 103077. <https://doi.org/10.1016/j.infrared.2019.103077>
- Yamashita, H., Sonobe, R., Hirono, Y., Morita, A., & Ikka, T. (2020). CHAPTER 4: Dissection of hyperspectral reflectance to estimate nitrogen and chlorophyll contents in tea leaves based on machine learning algorithms. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-73745-2>
- Ye, X., Abe, S., & Zhang, S. (2020). Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging. *Precision Agriculture*, 21(1), 198–225. <https://doi.org/10.1007/s11119-019-09661-x>
- Zhai, Y., Cui, L., Zhou, X., Gao, Y., Fei, T., & Gao, W. (2013). Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: Comparison of partial least-square regression and support vector machine regression met. *International Journal of Remote Sensing*, 34(7), 2502–2518. <https://doi.org/10.1080/01431161.2012.746484>
- Zhang, J., Cheng, T., Guo, W., Xu, X., Qiao, H., Xie, Y., & Ma, X. (2021). Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods. *Plant Methods*, 17(1), 1–14. <https://doi.org/10.1186/s13007-021-00750-5>
- Zhu, F., Zhang, D., He, Y., Liu, F., & Sun, D. W. (2013). Application of Visible and Near Infrared Hyperspectral Imaging to Differentiate Between Fresh and Frozen-Thawed Fish Fillets. *Food and Bioprocess Technology*, 6(10), 2931–2937. <https://doi.org/10.1007/s11947-012-0825-6>

## **5 Proximal Hyperspectral Imaging for Phenotyping Plant Drought Stress Using Univariate and Multivariate Modelling Techniques**

Frank Gyan Okyere <sup>1,2</sup>, Daniel Cudjoe <sup>1,2</sup>, Pouria Sadeghi-Tehran<sup>1</sup>, Nicolas Virlet<sup>1</sup>, March Castle <sup>1</sup>, Andrew B. Riche<sup>1</sup>, Latifa Greche<sup>1</sup>, Fady Mohareb <sup>2</sup>, Daniel Simms <sup>2</sup>, Manal Mhada <sup>3</sup> and Malcolm John Hawkesford <sup>1, \*</sup>

<sup>1</sup>Sustainable Soils and Crops, Rothamsted Research, Harpenden, United Kingdom, 2 School of Water.

<sup>2</sup>Energy and Environment, Cranfield University, Cranfield, United Kingdom, 3 AgroBioSciences.

<sup>3</sup>Department, University of Mohammed VI Polytechnic, Ben Guerir, Morocco

Based on: Okyere FG, Cudjoe D, Virlet N, Riche AB, Castle M, Greche L, Simms D, Mhada M, Mohareb F and Hawkesford MJ (2024 Proximal hyperspectral imaging for phenotyping plant drought stress using univariate and multivariate modelling techniques. Submitted to the International Journal for Remote Sensing to consider for publication

### **5.1 Abstract**

Accurate detection of drought stress in plants is essential for water use efficiency and agricultural output. Hyperspectral imaging (HSI) provides a non-invasive method in plant phenotyping, allowing the long-term monitoring of plant health due to sensitivity to subtle changes in leaf constituents. The broad spectral range of HSI enables the development of different vegetation indices (VIs) to analyze plant trait responses to multiple stresses, such as the combination of nutrient and drought stresses. However, known VIs may underperform when subjected to multiple stresses. This study presents new VIs in tandem with machine learning models to identify drought stress in wheat plants under varying nitrogen (N) levels. A pot wheat experiment was set up in the glasshouse with four treatments: well-watered high-N (WWHN), well-watered low-N (WWLN), drought-stress high-N (DSHN) and drought-stress low-N (DSLN). In addition to ensuring that

plants were watered according to the experiment design, photosynthetic rate ( $P_n$ ) and stomatal conductance ( $g_s$ ) (which are used to assess plant drought stress) were taken regularly, serving as the ground truth data for this study. The proposed VIs, together with known VIs, were used to train three classification models: support vector machines (SVM), random forest (RF), and deep neural networks (DNN) to classify plants based on their drought status. The proposed VIs achieved more than 0.94 accuracy across all models, and their performance further increased when combined with known VIs. The combined VIs were used to train three regression models to predict the stomatal conductance and photosynthetic rates of plants. The random forest regression model performed best, suggesting that it could be used as a stand-alone tool to forecast  $g_s$  and  $P_n$  and track drought stress in wheat. This study shows that combining hyperspectral data with machine learning can effectively monitor and predict drought stress in crops, especially in varying nitrogen conditions.

**Keywords:** drought stress; gas exchange measurements; hyperspectral imaging; machine learning; vegetation indices

## 5.2 Introduction

Despite advancements in agronomic management and breeding procedures, crop productivity remains susceptible to abiotic stresses such as drought. During drought stress, plants close their stomata to conserve water, decreasing the absorption of carbon dioxide required for photosynthesis (Pirasteh-Anosheh et al., 2016). Furthermore, drought stress can affect nutrient absorption and cause nutrient imbalances, impacting plant metabolic processes. Hence, developing tools to detect plant drought stress is essential for prompt intervention and management to minimize crop losses.

Conventionally, agronomists and breeders evaluate drought stress by visual grading, for example using the stay-green morphological features of plants (Duan et al., 2018). However, this method is subjective and prone to errors since stresses like iron and nitrogen deficiency could exhibit similar stress symptoms. Multiple approaches, including gas exchange measurements—stomatal conductance ( $g_s$ ) and photosynthetic rate ( $P_n$ ) (L. Xu & Baldocchi, 2003), soil moisture monitoring (Leone, 2022), leaf temperature measurement (Grant et al., 2016), and water potential assessment (Y. Zhang

et al., 2022), have been deployed to detect drought stress in plants. While these methods provide detailed and precise information, they have low throughput and may have limited applicability to field conditions. Furthermore, some of these methods are invasive, limited to real-time monitoring, and subject to the spatial variability of field crops.

With advancements in computer vision, remote sensing data derived from different sensors, such as thermal, visible, and hyperspectral imaging (HSI), can effectively detect and monitor the temporal and spatial impact of drought conditions (Mertens et al., 2021)(Proctor et al., 2021). Recently, there has been a surge of interest in HSI applications for abiotic stress assessment, especially nutrient and drought stress analysis. Drought stress leads to subtle modifications in the chemical, physiological, and structural components of plants, which can be detected within the short-wave infrared region (1300–1500 nm) of the spectrum (Peñuelas & Filella, 1998). With prolonged drought stress, there are changes in leaf pigments, which can also be measured through the spectral variations in the visible region (Satterwhite & Henley, 1990). The complex dynamics of drought stress initiation and progression suggest that a single spectral band or VIs could be insufficient in detecting and monitoring drought stress over time.

One standard method for analyzing HSI data in plant phenotyping is to extract vegetation indices (VIs), which are mathematical combinations of spectral reflectance characteristics of vegetation at different wavelengths (Asaari et al., 2019b). VIs have the benefit of reducing the effect of scale factors, including lighting conditions and slope effects (Jay et al., 2017). VIs such as the renormalized difference vegetation index (RDVI), normalized photochemical reflectance index (PRI<sub>norm</sub>), photochemical reflectance index (PRI570), normalized difference vegetation index (NDVI), water index (WI), and normalized water index (NWI) have been used to detect drought stress and monitor its progression. Ihuoma and Madramootoo (Ihuoma & Madramootoo, 2017) presented a review of different VIs for monitoring plant drought stress for irrigation management.

To develop VIs for stress monitoring, spectral averaging, which calculates the average spectrum over the pixel domain, is performed after pre-processing. After spectral averaging, the resulting data are still huge and multicollinear. The high-dimensional

features can cause algorithmic instability, which can affect the accuracy of data analysis (Sun et al., 2010). Moreover, from the hundreds of wavelengths scanned, only a small subset may be associated with the desired trait; the remainder is usually redundant or irrelevant, which may increase the computational processing and overfitting. The high-dimensional data may also be susceptible to noise and non-uniformity, which may affect the interpretation and accuracy of the analysis. This may be mitigated by employing dimensionality reduction techniques, either feature extraction or selection. In feature extraction, the data are transformed from a high to a low-dimensional space. In contrast, a subset of relevant features from the original hyperspectral datasets is selected for feature selection, discarding redundant or irrelevant ones. Linear discriminant analysis (LDA) and principal component analysis (PCA) are examples of feature extraction methods. Filter (e.g., ReliefF and correlation-based feature selection (Chutia et al., 2017)), wrapper (sequential feature selection, recursive feature elimination (Nagasubramanian et al., 2018; Rady et al., 2017)), and embedding techniques (e.g., random forest and LASSO (Yang et al., 2013)) are examples of feature selection methods. Remeseiro and Bolon-Canedo (Remeseiro & Bolon-Canedo, 2019) documented the detailed operation of different feature selection methods for HSI analysis. Determining a standard method to select spectral features in HSI analysis is challenging despite the availability of several feature selection methods (Damodaran et al., 2017). Each feature selection strategy has its benefits and setbacks. Ensemble learning methods that combine the predictions of multiple selection methods to improve the overall accuracy and robustness of models are often employed to minimize the effect of using a single feature selection model.

Machine learning (ML) models have been widely used for hyperspectral data analysis because they handle complex data patterns and relationships. ML methods such as support vector machines (SVM) and random forest (RF), among others, have been used to evaluate the predictive power of multiple spectral indices for plant drought (Mertens et al., 2021) and nutrient stress (Pandey et al., 2017) analysis. Pairing whole spectra, extracted features, or selected features with different ML methods can improve the accuracy of plant drought stress classification. VIs are usually developed for specific stresses and may fail in plants with multiple stresses.



The relationship between drought stress and plant N content is complex and has a substantial effect on plant growth and development. Nitrogen contributes to the production of stress-related compounds such as osmoprotectants and secondary metabolites, which enable plants to cope with drought stresses (Anas et al., 2020). However, high N levels promote vegetative growth, which subsequently leads to increased water requirements of the plants. This can exacerbate the effects of drought stress in plants (Gastal et al., 2015). Furthermore, low N levels lead to a decrease in photosynthetic efficiency, which is already compromised during drought, leading to huge reductions in plant photosynthetic rates (Lawlor, 2002). Since N is a critical component of proteins affecting plant stomatal behavior and water-use efficiency, understanding the dynamics of plant spectral characteristics in response to drought stress under different N contents is essential. From previous studies (Gastal & Lemaire, 2002)(Seleiman et al., 2021), adequate N content promotes the growth and yield of plants. However, excess N levels may increase plant water demand, potentially worsening drought stress. There is little knowledge on the non-invasive assessment of excess or inadequate N levels on plants with variable water content.

This study aims to understand plant canopy spectral interactions under different N and water levels. First, we examined the effectiveness of using various known VIs to assess plant drought stress under different N regimes. An ensemble machine learning method was developed to select sensitive features from which new VIs corresponding to drought stress under low and high N content were proposed. Finally, a deep neural network was proposed and trained with the known and proposed VIs to predict the stomatal conductance and photosynthetic rate of plants. Since gas exchange measurements are effective in assessing drought stress, the proposed and known VIs were used as input to train conventional ML models to predict  $P_n$  and  $g_s$  in wheat.

### **5.3 Materials and Methods**

Figure 5.1 shows the workflow of the methods used in this experiment. The general steps include HSI acquisition, data pre-processing, selection of known VIs, sensitive waveband selection, development of proposed VIs, development of ML models, identification of drought stress, and prediction of  $P_n$  and  $g_s$ .

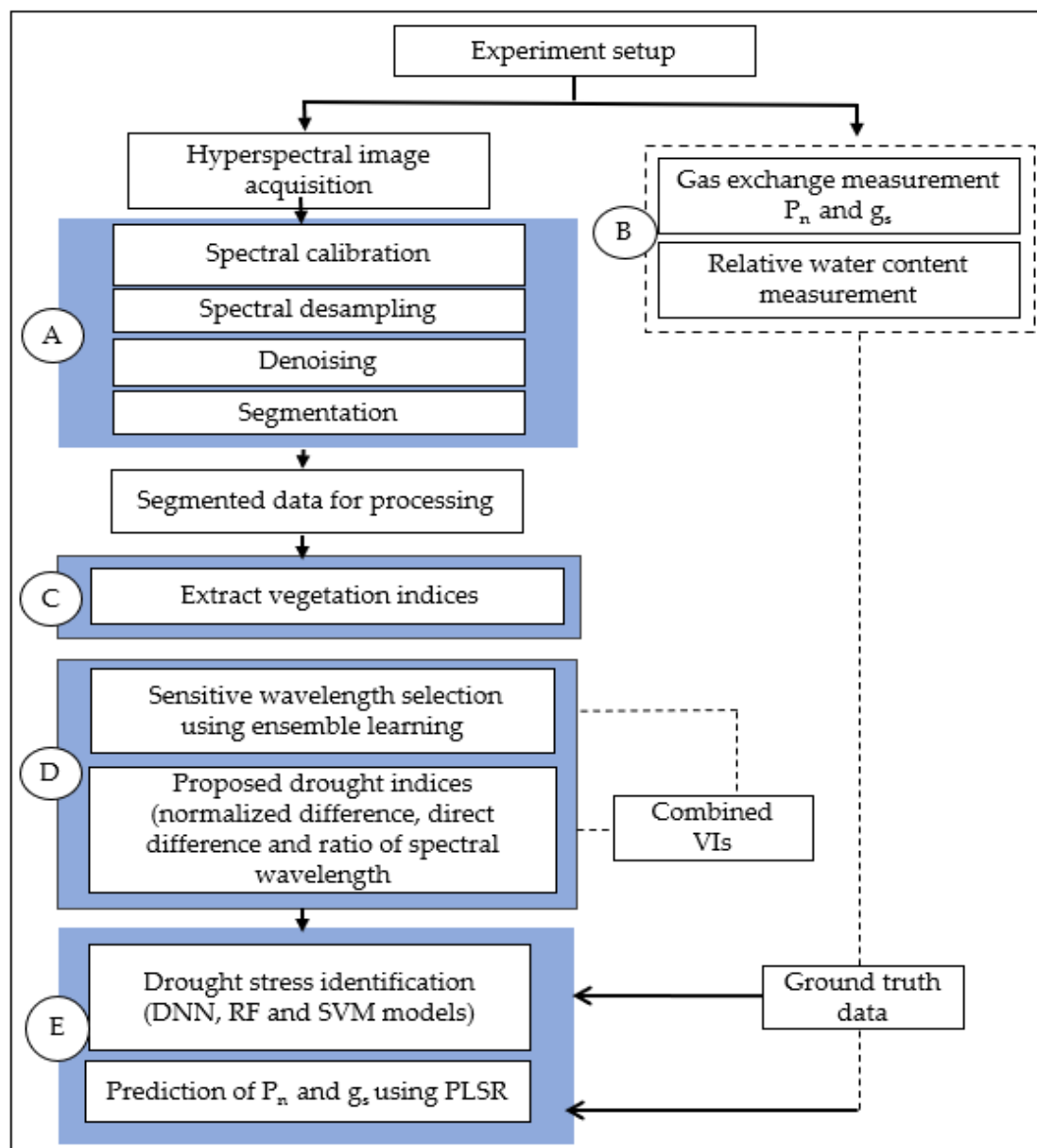


Figure 5-1 The schematic diagram of the methodology for analysing spectral images for drought stress identification (A) is the preprocessing step (involving data calibration, denoising, desampling and segmentation), (B) drought stress-related physiological measurements ( $g_s$  and  $P_n$ ), (C) the extraction of known VIs (D) ensemble learning model for selecting sensitive spectral wavelengths (E) the development of classification and regression models for identification of drought stress and prediction of gas exchange traits ( $P_n$  and  $g_s$ ) respectively.

### 5.3.1 Experiment setup

A drought experiment was set up in a glasshouse facility (<https://www.cranfield.ac.uk/facilities/plant-growth-facility>, accessed on 10 November 2022) at Cranfield University. This facility has a state-of-the-art phenotyping platform (Lemnatec Scanalyzer system) for high-throughput data acquisition. This experiment is the same as the experiment in Chapter 4 of this thesis. However, a hyperspectral dataset was acquired for this chapter using the VNIR hyperspectral sensor. In this experiment, a wheat cultivar, Cadenza, was planted in 8 cm by 6 cm pots filled with low-N peat soil (Levington Advance M3, CTS, BHGS Ltd, Worcestershire, , UK). Plants were grown under natural light (with an average light intensity of 450–600  $\mu\text{mol}/\text{m}^2/\text{s}$  PAR) with a 20 °C to 23 °C day temperature range, while the optimum night temperature was between 18 °C and 20 °C. Plants (48 in number) consisting of four treatments and 12 replicates were arranged in a randomized complete block pattern. The treatments comprised plants with two N and water content levels: WWHN, WWLN, DSHN, and DSLN (WW = well-watered; DS = drought stress; HN = high N; and LN = low N). The plants were fertilized with two N levels, high N and low N, made of 42.5 mM and 4.25 mM concentrations, respectively, at 30, 37, and 44 DAS (days after sowing). They were prepared from a modified Letcombe nutrient solution (Masters Clark et al., 2020). For the drought stress treatment, all the plants were first watered with equal amounts of deionized water until the tillering stage (44 DAS), when the drought stress was imposed (0 DADS—day after drought stress). All the well-watered treatments (WWHN and WWLN) were kept at 80% field capacity (FC) (% FC is the proportion of soil moisture content at field capacity), while the watering was stopped for the drought stress treatments (DSHN and DSLN). The well-watered plants were watered slowly and thoroughly with deionized water each day until the soil reached the desired moisture level (80% FC). To ensure the WW plants were at 80% FC or above, a portable soil moisture meter was used to measure the volumetric soil water content daily. In addition, the pots were weighed daily to determine the rate of water decrease in the DS-based treatments. The soil surface was covered with white pebbles to minimize water loss through evaporation. The WW plants were watered daily until the end of the experiment, which lasted for 15 days.

### 5.3.2 Physiological measurements of drought stress

Drought stress can significantly affect the gas exchange rate of plants, reflecting plant physiological responses to low water content. Gas exchange measurements such as stomatal conductance ( $g_s$ ), transpiration rate, and photosynthetic rate ( $P_n$ ) are often used to track changes in the physiological properties of plants before noticeable symptoms such as yellowing or wilting become apparent (Thameur et al., 2012). This study used the  $g_s$  ( $\text{mmol H}_2\text{O m}^{-2} \text{ s}^{-1}$ ) and  $P_n$  ( $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ) of plants as ground truth measurements to track changes in plant physiological traits resulting from drought stress and N deficiencies. Stomatal conductance measures the rate at which  $\text{CO}_2$  or water vapor enters and exits through plant stomatal pores. It is a critical physiological indicator for screening drought-tolerance genotypes of plants (Keshtiban et al., 2015). Additionally, the photosynthetic rate represents the capacity of plants to convert carbon dioxide into organic compounds (N. T. Nguyen et al., 2003).

A LI-6400XT portable photosynthesis system (LI-COR Biosciences Inc., Lincoln, NE, USA) was used to measure the gas exchange measurements of the plants a day before the water stress was induced and continued every three days until the end of the experiment. Stomatal conductance measurements were made between midday and 1700 h on the fully expanded leaf of the main stem that had been marked and tagged. The  $\text{CO}_2$  concentration in the leaf chamber was fixed at  $400 \mu\text{mol CO}_2 \text{ mol}^{-1}$ . To achieve the maximum photosynthetic capacity, the leaf flow rate, temperature, relative humidity, and photosynthetically active radiation were set at  $200 \mu\text{mol s}^{-1}$ ,  $20^\circ\text{C}$ , 50–65%, and  $1800 \mu\text{mol m}^{-2} \text{ s}^{-1}$ , respectively. The stomatal conductance values were recorded once the measurement was stabilized. The leaf area was corrected during data processing when the leaf was smaller than the cuvette of the chamber.

To analyze the effects of each treatment on the dynamics of the gas exchange measurements, the average  $g_s$  and  $P_n$  of each treatment were subjected to a one-way ANOVA test followed by Tukey's honest significant difference (HSD), with \*  $p < 0.05$  considered statistically significant. The ANOVA test compares the means across multiple groups to determine if there exist any statistically significant differences between them.

The statistical analysis was performed in RStudio (Ver. 1.1.414, RStudio, Boston, MA, USA).

### 5.3.3 Hyperspectral Data Acquisition

The Lemnatec Scanalyzer housing a hyperspectral camera (hyperspec® inspector™ Headwall Photonic) in the glasshouse was used to acquire spectral images 3 m above the ground. The camera is a push broom type that covers the visible-and-near-infrared (VNIR) regions with wavelengths between 390 nm and 1015 nm. The sensor uses an FWHM (full width at half maximum) image slit of 2.5 nm to gather data at 0.7 nm (in the VNIR area) and 6 nm (within the SWIR region). As a result, a hypercube with a dynamic range of 16 bits containing 1015 spectral bands is produced. Data were collected between 10 am and 12 pm on each imaging day. Imaging began on the day the drought was induced and continued every three days until the end of the experiment.

### 5.3.4 Hyperspectral data pre-processing

Obtaining useful information from hyperspectral images requires pre-processing to normalize the spectral data from ambient illumination, reduce noise and other artifacts, and improve the data quality for further processing. The pre-processing steps include radiometric calibration, spectral down-sampling, and noise removal. Radiometric calibration standardizes and adjusts the radiometric data recorded by converting the raw sensor measurements to meaningful physical units such as radiance or reflectance. This is important in reducing the variable illumination and the dark current effect on the spectra. In this study, a white panel (Zenith Lite™ Ultralight Targets 95%R, Sphereoptics®) was imaged as the white reference data ( $Ref_{white}$ ) while the camera lens was covered with an opaque cap to collect the dark reference data ( $Ref_{dark}$ ) (Zhu et al., 2013). The image reflectance was obtained by using Equation 5.1.

$$reflectance_{image} = \frac{Raw_{image} - Ref_{dark}}{Ref_{white} - Ref_{dark}} \quad \text{Equation 5.1}$$

Hyperspectral imaging captures high-resolution images across a wide range of narrow and continuous spectral bands. The massive amount of information presents computational challenges that require down-sampling of the data for effective processing. Down-sampling involves reducing the number of spectral bands or channels in the data,

subsequently reducing the spectral resolution. According to Sadeghi-Tehran et al. (Sadeghi-Tehran et al., 2021), down-sampling helps reduce the computational complexities and noise generated during hyperspectral data acquisition. This study used a band-averaging technique of 2 nm spectral width to down-sample the data. The process involves grouping two adjacent spectral bands and finding the average values to generate a new set of bands. As a result, the down-sampled data had fewer wavebands overall but still had a representative spectral characteristics.

Spectral smoothing and denoising methods were applied to the raw spectra to reduce noise levels and improve the signal-to-noise ratio, revealing underlying spectral patterns. Spectral smoothing and denoising include using filters on the spectral bands to remove spikes and smooth spectral curves. This process also isolates sensitive features that could be masked by noise at various wavelengths. Here, a Savitzky–Golay filter (a commonly used low-pass filter) to smooth and denoise the spectral bands was used to fit a polynomial function to the data inside a moving window as a form of polynomial smoothing. The filter chooses an odd-sized window of spectral points for each data point in the spectrum and then fits the least square using the high-order polynomial. During this process, the relevant data points are eventually swapped out for the matching values of the fitted polynomial. A window of size 13 and a second-degree polynomial were used as the parameters. It should be noted that a small window size generates significant artifacts, while a large window size can be more effective in reducing large-scale noise and smoothing out noise of variable frequencies. However, large window sizes tend to blur fine spectral details, which could distort the originality of the spectral signature (Rinnan et al., 2009).

### **5.3.5 Segmenting the hyperspectral data**

After pre-processing, the HSI data were segmented using a selected spectral ratio and Otsu thresholding. A method that exploits the difference in the shape of the infrared red regions of the spectrum in the plant and background was developed for the spectral ratio. In this case, a normalized difference ratio between 910 and 950 nm wavelengths was extracted to create the spectral ratio. The combination of spectral ratio and the Otsu thresholding resulted in a binary image where the vegetation pixels were labeled as one and the non-vegetation pixels as zeros. Supplementary information Figure E-1 shows the original image and its corresponding image of a sample of the hyperspectral data.

### **5.3.6 Extracting Known Vegetation Indices**

Vegetation indices (VIs) of plants derived from reflectance values in specific bands are indicative of responses to different stresses. Combining VIs or customizing them based on unique vegetation characteristics and environmental conditions is a common practice to monitor plant stress (Koh et al., 2022). This study extracted twenty-five known VIs sensitive to nitrogen variations and drought stress (Table 5-1). The top ten VIs that correlated well with the gas exchange measurements were evaluated to understand their sensitivities to subtle changes in the nitrogen and water content of plants. The selected VIs were used subsequently for the subsequent analysis.

**Table 5-1** List of selected VIs to monitor plant drought stress.

| Vegetation Indices                                | Formula                           | Reference                |
|---------------------------------------------------|-----------------------------------|--------------------------|
| Normalized difference vegetation index (NDVI)     | $(R800 - R680)/(R800 + R680)$     | (Tucker, 1979)           |
| Chlorophyll index green (CI-green)                | NIR/Green -1                      | (Gitelson et al., 2003)  |
| Renormalized difference vegetation index (ReNDVI) | $R800 - R670/(R800 + R670)^{1/2}$ | (Sims & Gamon, 2002)     |
| MERIS terrestrial chlorophyll index (MTCI)        | $(R753 - R708)/(R708 - R681)$     | (Haboudane et al., 2004) |
| Red edge NDVI (RENDVI)                            | $(R705 - R740)/(R705 + R740)$     | (Tucker, 1979)           |
| Normalized difference vegetation index (NDVI750)  | $(R750 - R680)/(R750 + R680)$     | (Tucker, 1979)           |
| Modified red edge simple ratio index (mRESR)      | $(R750 - R445)/(R750 + R445)$     | (Sims & Gamon, 2002)     |
| Photochemical reflectance index (PRI710)          | $(R531 - R710)/(R531 + R710)$     | (Sims & Gamon, 2002)     |
| Photochemical Reflectance Index (PRI720)          | $(R531 - R720)/(R531 + R720)$     | (Gamon, J.A, 1992)       |
| Structure insensitive pigment index (SIPI)        | $(R800 - R455)/(R800 + R705)$     | (Penuelas et al., 1995)  |
| Pigment specific simple ratio (PSSRa)             | $R800/R680$                       | (Blackburn, 1998)        |
| Reflectance difference (RD)                       | $R800 - R680$                     | (Blackburn, 1998)        |
| Chlorophyll index red edge (CI-red edge)          | $(R750 - R700)/(R700)$            | (Podani & Czárán, 1997)  |



|                                                                 |                                                                                     |                                |
|-----------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------|
| Water band index (WBI)                                          | $(R950/R900)$                                                                       | (H. R. Xu et al., 2007)        |
| Transformed chlorophyll absorption in reflectance index (TCARI) | $3X [(R705 - 665) - 0.2X (R705 - R560) X (R705/R665)]$                              | (Haboudane et al., 2004)       |
| Optimized soil-adjusted vegetation index (OSAVI)                | $((1 + 0.16) X (R865 - R665)/(R865 - R665 + 0.16))$                                 | (Rondeaux et al., 1996)        |
| Enhanced vegetation index (EVI)                                 | $2.5 \times [(R800 - R680)/(R800 + 6 \times R680 - 7.5 \times R450 + 1)]$           | (Buschmann & Nagel, 1993)      |
| Soil adjusted vegetation index (SAVI)                           | $((1 + 0.5) X (R801 - R670)/(R801 + R670 + 0.5))$                                   | (Qi et al., 1994)              |
| Optimized soil adjusted vegetation index (OSAVI800)             | $(1 + 0.16) (R800 + R670)/(R800 + R670 + 0.61)$                                     | (Rondeaux et al., 1996)        |
| Red edge vegetation index (RSVI)                                | $(NIR/Red)-1$                                                                       | (Qi et al., 1994)              |
| Improved SAVI with self-adjustment factor L (MSAVI)             | $0.5 \times \{2 \times R800 + 1 - (2 \times R800 + 1)^2 - 8 \times (R800 - R670)\}$ | (Qi et al., 1994)              |
| Normalized difference infrared index (NDII)                     | $(R780 - R710)/(R780 - R680)$                                                       | (Datt, 1999)                   |
| Normalized difference water index (NDWI)                        | $(R560 - R830)/(R560 + R830)$                                                       | (McFeeters, 1996)              |
| Difference vegetation index (DVI)                               | $R800 - R670$                                                                       | (Perry & Lautenschlager, 1984) |
| Vegetation stress ratio (VRS)                                   | $R725/R702$                                                                         | (White et al., 2008)           |

### **5.3.7 Wavelength selection and new drought stress indices**

#### **5.3.7.1 Wavelength Selection Using Ensemble Learning**

Hyperspectral images have a high spectral resolution comprising hundreds of narrow bands. However, a significant portion of the spectral bands may be strongly correlated (multicollinear). To reduce this multicollinearity, sensitive features related to the phenotypic traits of interest are extracted using machine learning and statistical methods (Okere et al., 2023). This study implemented an ensemble learning method to select the most sensitive features (Figure 5.2). Three feature selection models, correlation feature selection (CFS), chi-square (CS), and ReliefF (RFF) were developed on 70% of training datasets and tested on 30% of test datasets. Each model was trained using a k-fold cross-validation technique where the training dataset was divided into K (5) subsets to train and validate the model multiple times. The features selected were ranked in order of importance, and the top ten features from each model were selected. Because each model had its drawbacks, the feature subset that was ultimately selected might not be the best in the feature space. The features selected were combined to obtain 30 features. A further selection was made on the combined feature subset using a Boruta SHAP algorithm (Manikandan et al., 2024), which was ranked using a recursive feature selection method.

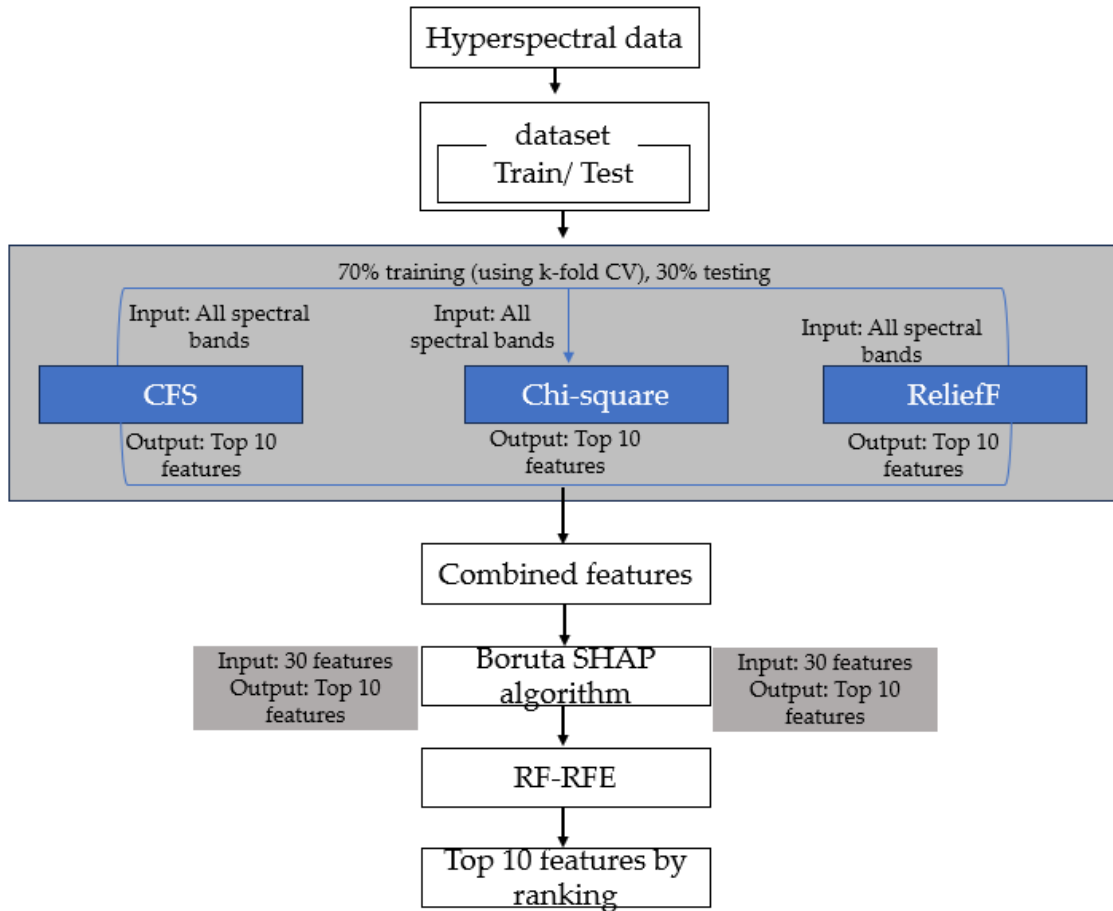


**Figure 5-2** Flowchart outlining the feature selection process applied to hyperspectral data. This workflow is based on an ensemble feature selection technique comprising three base models: correlation feature selection (CFS), chi-square and ReliefF methods. While each based model produces top ten selected features, the Boruta SHAP algorithm selects the final 10 wavelengths.

### 5.2.7.2 Proposed Drought Stress Indices

The selected wavelengths were used to develop new indices: drought-N ratio index (RDI), normalized difference drought-N index (NDDI), and drought difference index (DDI) using Equations 5.2–5.4. These equations were selected because they help minimize the effect of varying light conditions, including sunlight intensity, angle of sunlight, etc., on the plant reflectance measurements. Additionally, the indices from these equations reduce the effects of atmospheric conditions such as haze, aerosols, and scattering on plant reflectance (Jackson & Huete, 1991)

Equation 5.2

$$DRI = \frac{RD_{\lambda 1}}{RD_{\lambda 2}}$$

$$NDDI = \frac{RD_{\lambda 1} - RD_{\lambda 2}}{RD_{\lambda 1} + RD_{\lambda 2}}$$

Equation 5.3

$$DDI = RD_{\lambda 1} - RD_{\lambda 2}$$

Equation 5.4

Where  $RD_{\lambda 1}$  and  $RD_{\lambda 2}$  are any two selected wavelengths.

Using a custom-developed algorithm, the three proposed indices were calculated for every possible pair combination of the selected wavelengths. The relationship between the generated indices and the gas exchange measurements was ascertained using correlation analysis. A matrix plot displaying a distinct pattern with multiple hotspots with somewhat varied coefficient of determination ( $R^2$ ) values were produced by plotting all the squares of correlation coefficient  $r$  values, which reflect the coefficient of determination. The optimal wavelength combinations with the highest  $R^2$  values were chosen as the proposed indices.

### 5.3.8 Machine learning models for drought stress identification

Three machine learning algorithms, SVM, RF, and DNN, were trained using the selected known VIs, proposed VIs, and combined VIs (combination of known and proposed VIs) to identify drought stress in wheat. All three models are supervised learning models for both classification and regression tasks. SVM finds the optimal hyperplane that maximizes the margin between classes in the feature space, while RF is an ensemble learning model (Rodriguez-Galiano et al., 2012) that uses bagging techniques where multiple decision tree models are trained on various subsets of data independently (Breiman, 2020). DNN consists of multiple layers of neurons (input, hidden, and output layers), with each layer containing numerous neurons connected through weighted to learn a hierarchical representation of data through a backpropagation algorithm (Rumelhart et al., 1986). Supplementary information Table E-1 shows a comprehensive

characteristic of each model and the reasons for choosing the model for the identification of drought stress in wheat.

### **5.3.9 Multivariate analysis for stomatal conductance and photosynthetic rate predictions**

Based on prior research, gas exchange measures can evaluate and track the dynamics of drought stress since minute changes in these physiological traits indicate responses of plants to water availability. However, the tools for measuring gas exchange are costly, low throughput, and sometimes destructive. Hence, regression models were developed to predict the plant  $g_s$  and  $P_n$  using the plant VIs, which could be used as a proxy tool to monitor drought stress. Four models, a polynomial regression (PR), random forest regression (RFR), support vector regression (SVR), and partial least squares regression (PLSR), were trained to evaluate their abilities to predict  $g_s$  and  $P_n$ . The RFR, SVR, and PR were each trained with the VIs as the independent factors, while the  $g_s$  or  $P_n$  was the dependent variable. The PLSR model was computed considering the whole spectral reflectance as the independent variable and  $g_s$  or  $P_n$  as the dependent variable. Here, the PLSR model was trained by concurrently finding the principal components that account for the highest variance and low multicollinearity in the dependent and independent variables (Helmholz et al., 2014). This results in fewer uncorrelated components (variables) from the large spectrum with little loss of information.

### **5.3.10 Model Training and Testing**

All All the models were trained with 70% of the dataset and tested on the remaining 30%. In ML training, optimization of parameters by cross-validation (CV) search over various parameter settings is required to improve the accuracy of predictions and classification and minimize errors. Popular optimization techniques include gridsearchCV (Buitinck et al., 2013), randomized search (Bergstra & Bengio, 2012), and Bayesian optimization (Brochu et al., 2010). Since random search is computationally less expensive and consumes less time for processing, the randomsearchCV tool in the Scikit learn package (Python 3) was used to fine-tune the parameters of the ML models. Table 5-2 includes the list of parameters tuned and the range of values considered. Following the model

parameter tuning and training, the best-performing set of parameters was used for model fitting and classification or regression.

The metrics root mean square error (RMSE), coefficient of determination ( $R^2$ ), and mean absolute error (MAE), as shown mathematically from Equations 5.5 to 5.7, were used to assess the validity of the regression models in predicting the  $P_n$  and  $g_s$ . The  $R^2$ , ranging from 0 to 1, measures the proportion of the variance in the dependent variable that is predictable from the independent variables.  $R^2$  values close to 1 indicate a good fit, which shows a good performance of the model, and vice versa. The RMSE measures the square root of the average squared differences between the predicted and observed variables, which gives a sense of the magnitude of the errors made by the models. A low RMSE suggests a good performance of the model. Generally, a high  $R^2$  and low RMSE indicate good performance of the model. The MAE is the average of the absolute differences between predicted and observed variables. A lower MAE shows that the model's predictions are close to the ground truth values. To objectively assess the performance of the classification models, we used three widely used classification metrics: average accuracy (AA), F-score, and Cohen's kappa score, as shown in Equations 5.8-5.10.

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{Equation 5.5}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad \text{Equation 5.6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad \text{Equation 5.7}$$

Where  $n$  is the number of data,  $x_i$  is the observed values,  $y_i$  is the predicted values and the bar denotes the mean of the variable. Statistical calculations were performed using the statistical package in RStudio.

**Equation 5.8**

$$AA = \frac{TP+TN}{TP + TN + FP + FN} \quad \text{Equation 5.9}$$

$$F\text{-Score} = 2 \frac{P \times R}{P + R} \quad \text{Where } P = \frac{TP}{TP + FP} \quad \text{and} \quad R = \frac{TP}{TP + FN}$$

$$K = \frac{P_o - P_c}{1 - P_c} \quad \text{Equation 5.10}$$

Where TP: true positive, TN: True negative, FP: false positive, FN: false negative, Po is the probability of observed agreements and Pc is the expected agreement.

**Table 5-2** Model parameters for tuning and training using random search CV

| Model | Parameters                           | Range                                              |
|-------|--------------------------------------|----------------------------------------------------|
| DNN   | Hidden layers                        | 1,2,3,4,5                                          |
|       | Number of neurons                    | 50, 100, 150, 200, 300                             |
|       | Activation function                  | identity, logistics, tanh, ReLU                    |
|       | Weight optimization                  | lbfgs, sgd, adam                                   |
|       | Regularization penalty ( $\alpha$ )  | 0.00001, 0.0001, 0.001, 0.01                       |
|       | Learning rate                        | constant, adaptive, in scaling                     |
|       | Batch size                           | 200, 300, 400, 500, 600, 700                       |
|       | Momentum for gradient descent update | 0.9.                                               |
|       | Exponential decay rate ( $\beta$ )   | 0.9.                                               |
| SVM   | Kernel type                          | rbf, poly, linear                                  |
|       | Degree of the polynomial kernel      | 1, 2, 3                                            |
|       | Regularization parameter (C)         | 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10,000 |
|       | Kernel coefficient (gamma)           | 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10,000 |
| RF    | Number of trees                      | 10, 30, 50, 70, 90, 110, 130, 150, 170, 190, 210   |
|       | Maximum depth of the tree            | 10, 20, 30, 40, 50, 60, 70, 80, 90, 100            |

|                                       |                                |
|---------------------------------------|--------------------------------|
| Number of features for the best split | sqrt (1 8 1), log2(1 8 1), 181 |
| Minimum samples for splitting         | 2, 5, 10                       |
| Bootstrap samples for building tree   | True, False                    |

## 5.4 Results

Before analyzing the results of the interaction between the proposed and known VIs and how they are affected by drought stress and variable N levels, it is important to understand how the multiple stresses affected the gas exchange measurements (which served as ground truth measurements for this study) and the whole spectra of each treatment. Section 3.2 summarizes the results of the plants' gas exchange measurements under drought and various N levels, while Section 3.3 examines the spectrum characteristics of each treatment at three different growth stages. The following sections further detail the extraction and proposal of new VIs and how they respond to drought stress when combined with the known Vis.

### 5.4.1 Reference data of gas exchange measurements

Figure 5.3 is a box and whisker plot of the  $P_n$  measurements of the plants from the 0–15 DADS. Supplementary Table E-2 is a summary of a mixed linear model regression (MixedLM) output for the  $P_n$  of plants showing the random and fixed effects of the timepoints and the different treatments. It was generally observed that most of the fixed effects (treatments, timepoints and their interactions) had high p-values indicating that only a few showed statistically significant differences (WWLN treatment at timepoint: 9 DADS). This suggests the model generally does not detect strong evidence that the treatments or timepoint influences the outcome variable ( $P_n$ ) in this dataset. Although, holistically, there model showed no statistically significant differences at different time point, a Tukey's test was conducted per each timepoint to access the interactions of the treatments at each point. The WWHN treatment produced the highest  $P_n$  throughout the drought stages, while the DS treatments (DSHN, DSLN) steadily declined with time. It was observed that the DS treatments were not significantly different at the beginning of the drought (0 DADS), despite the variations in their N levels. However, the DSHN treatment was statistically different from the WWLN treatment. Plants with varying N levels exhibited statistical differences three days after drought stress initiation. That is,



the WWLN and WWHN were statistically different ( $p < 0.049$ ), while the DSLN plants showed differences with the WWHN ( $p < 0.045$ ). The WWLN and DSLN, however, showed no differences at this stage. The stress in the DS treatments became more noticeable on the 6 DADS, where the means of both treatments (DSHN and DSLN) were statistically different from the WWHN treatment ( $p < 0.026$  and  $p < 0.015$  for DSHN and DSLN, respectively). In contrast, the WWLN did not differ much from the DSLN ( $p < 0.97$ ) treatment; however, it was statistically different from the WWHN ( $p = 0.036$ ). At the end of the experiment, there were clear distinctions between the drought-stressed (DSHN and DSLN) and well-watered (WWHN and WWLN) treatments.

Figure 5.4 shows the time series analysis of the four treatments at 6 different time points for the  $g_s$  of the plants. Supplementary Table E-3 is a summary of the the output of a mixed linear model which incorporates both fixed and random effects to the model. The intercept (0.257, with  $p < 0.001$ ) which represents the estimated value when all other variables are at their reference is statistically significant. The output of the model suggests that timepoints have significant effects on the dependent variable, with most of them having negative effects relative to the baseline. The interaction between the groups and timepoints demonstrates that different groups experience the timepoints effects differently. For example, groups like WWHN and WWLN show positive interaction effects at specific timepoints, whereas other groups like (DSLN) show less noticeable impacts. Furthermore, the random effect for group variance is minimal, showing that the model's fixed effects capture the majority of the variability.

A posthoc analysis using Tukey's test on the  $g_s$  measurement of the plants revealed a pattern comparable to the  $P_n$  measurement, as shown in Figure 5.4. Throughout the drought stages, the WWHN treatment had the highest  $g_s$ , with the maximum at six DADS ( $0.37 \text{ mol H}_2\text{Os}^{-1}$ ). Despite having varying N levels, the DS treatments showed no significant differences at 0 DADS. At 0 DADS, the means of the DS treatments (DSHN and DSLN) were statistically different from the WWLN treatment. After three days, the WWHN treatment had the highest  $g_s$  ( $0.341 \text{ mol H}_2\text{Os}^{-1}$ ), while the DSLN obtained the lowest  $g_s$  ( $< 0.252 \text{ mol H}_2\text{Os}^{-1}$ ). There were observable differences in the ranges and mean values of the WWHN treatments and the other treatments, with  $p$ -values = 0.025, 0.011, and 0.012 for the WWLN, DSLN, and DSHN treatments, respectively. However, the DS treatments were not significantly different from the WWLN. On the six DADS, the WW

treatments (WWHN and WWLN) had higher mean  $g_s$  than the DS treatments. While the DS treatments were statistically different from the WWHN treatment, the DSLN and WWLN treatments were statistically not different. On the 12 DADS, there were 80.3% and 60.2% decreases in  $g_s$  for the DSHN and DSLN, respectively (indicating a fast decline of  $g_s$  for the DS treatments). On the final day (15 DADS), a significant difference was observed in drought-stressed and well-watered plants, with the WWHN having the highest and the DSHN having the lowest  $g_s$  measurements.

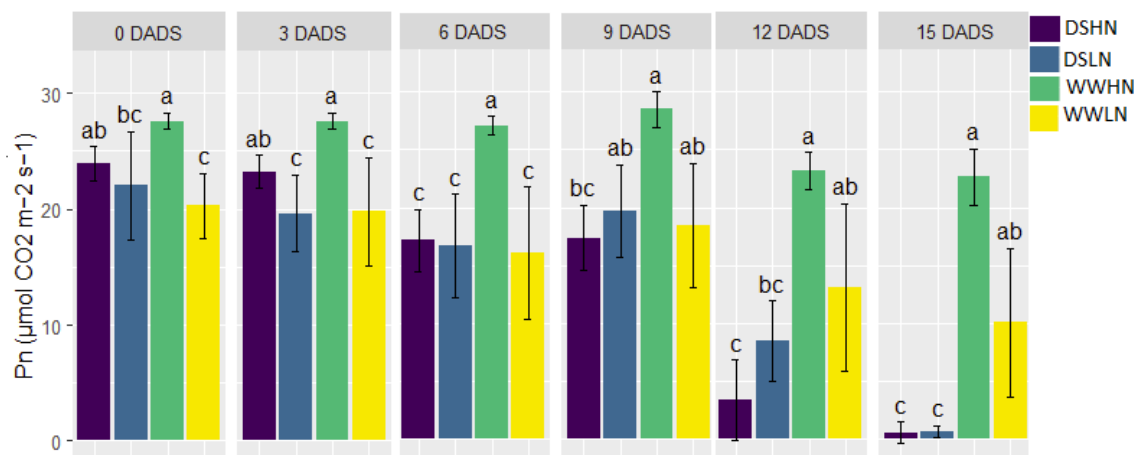


**Figure 5-3** Photosynthetic rates from 0 to 15 DADS for four different treatments: WWHN, WWLN, DSHN, and DSLN. The WWHN and WWLN are the water well-watered plants with high and low nitrogen, respectively, while the DSHN and DSLN are the drought-stressed plants with high and low N levels, respectively. The results presented are mean and standard deviations from the original data; the dissimilar lower-case group (a, b, and c) represents a significant difference with  $p < 0.05$ .

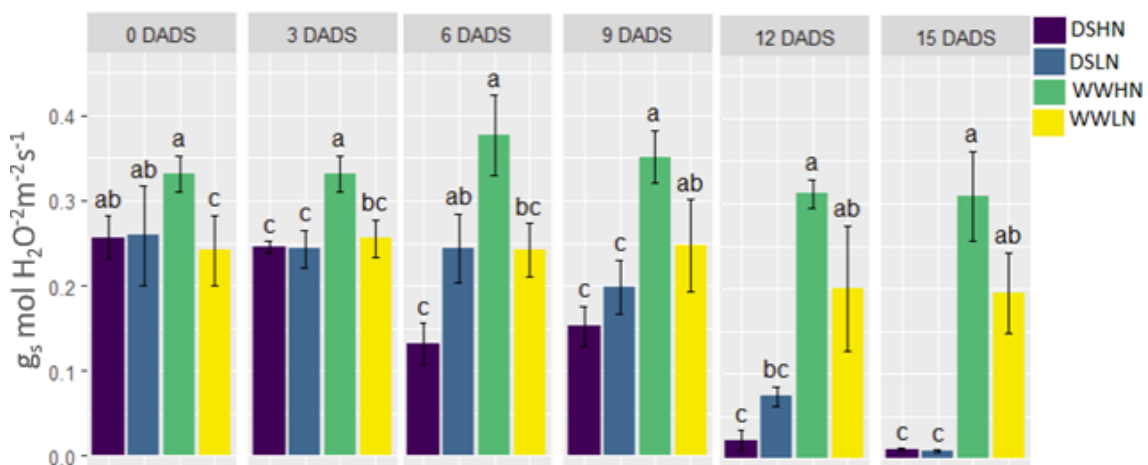
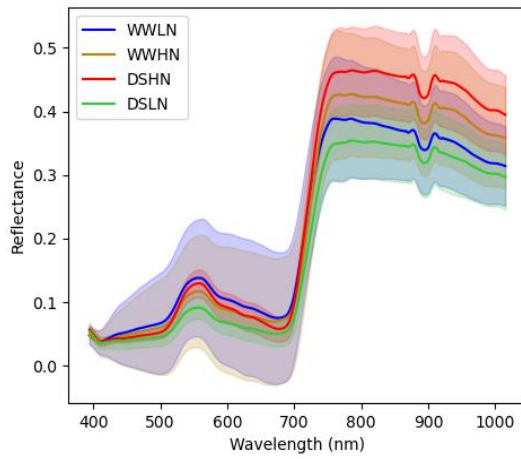


**Figure 5-4** Stomatal conductance for the four treatments: WWHN, WWLN, DSHN, and DSLN from 0 to 15 DADS. The WWHN and WWLN are the water well-watered plants with high and

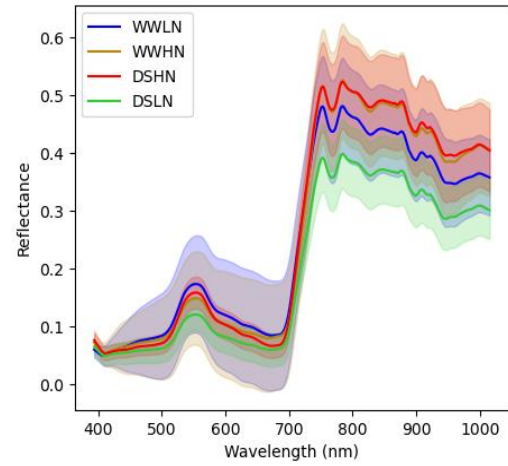
low nitrogen, respectively, while the DSHN and DSLN are the drought-stressed plants with high and low N levels, respectively. The results presented are mean and standard deviations from the original data; the dissimilar lower-case group (a, b, and c) represents a significant difference with  $p < 0.05$

### 5.3.2 Spectral Reflectance Analysis

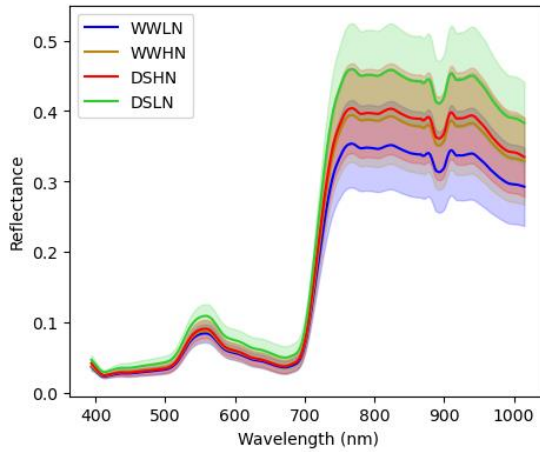
Mean spectral curves for the different treatments at 0, 6, and 15 DADS are shown in Figure 5.5. The spectral curves for all the treatments showed peaks at 570 nm and troughs at 680 nm. The spectral reflectance of all the treatments exhibited comparable patterns at 0 DADS, showing relatively low reflectance, particularly in the blue spectrum (around 450 nm) and red regions (around 670 nm). However, the DSLN treatment had slightly higher reflectance comparably. The red edge region (around 700 nm) for all the treatments showed a sharp progression from the red to the near-infrared regions, with the high N treatments (DSHN and WWHN) showing a relatively higher reflectance in this region as compared to the low N treatments, which showed low reflectance. Conversely, on the 6 DADS, the general trend in the visible spectrum was similar for all the treatments with low reflectance in the blue (~450 nm) and red (~670 nm) regions. In the red edge region, the WWHN had the highest reflectance (0.57 at around 730 nm), while the DSHN had the lowest reflectance (0.39 around 720 nm). All the treatments had a slightly downward trend in the reflectance value in the near-infrared regions. Furthermore, reflectance's for the well-watered plants (WWLN and WWHN) were close in proximity, with WWHN showing slightly higher reflectance in the near-infrared regions. On the final day of the experiment (15 DADS), all the treatments showed relatively low reflectance with minor variations, with the DSHN treatment showing a relatively high reflectance (0.11) in the visible region. All the treatments showed a sharp increase in reflectance in the red edge region starting around 680 nm and transitioning to the near-infrared regions. However, the DSLN had a slightly less sharp and shifted shape in this region. Interestingly, the DSLN had the highest reflectance in the near-infrared region



(a) Spectral reflectance at 0 DADS



(b) Spectral reflectance at 6 DAD



(c) Spectral reflectance at 15 DADS

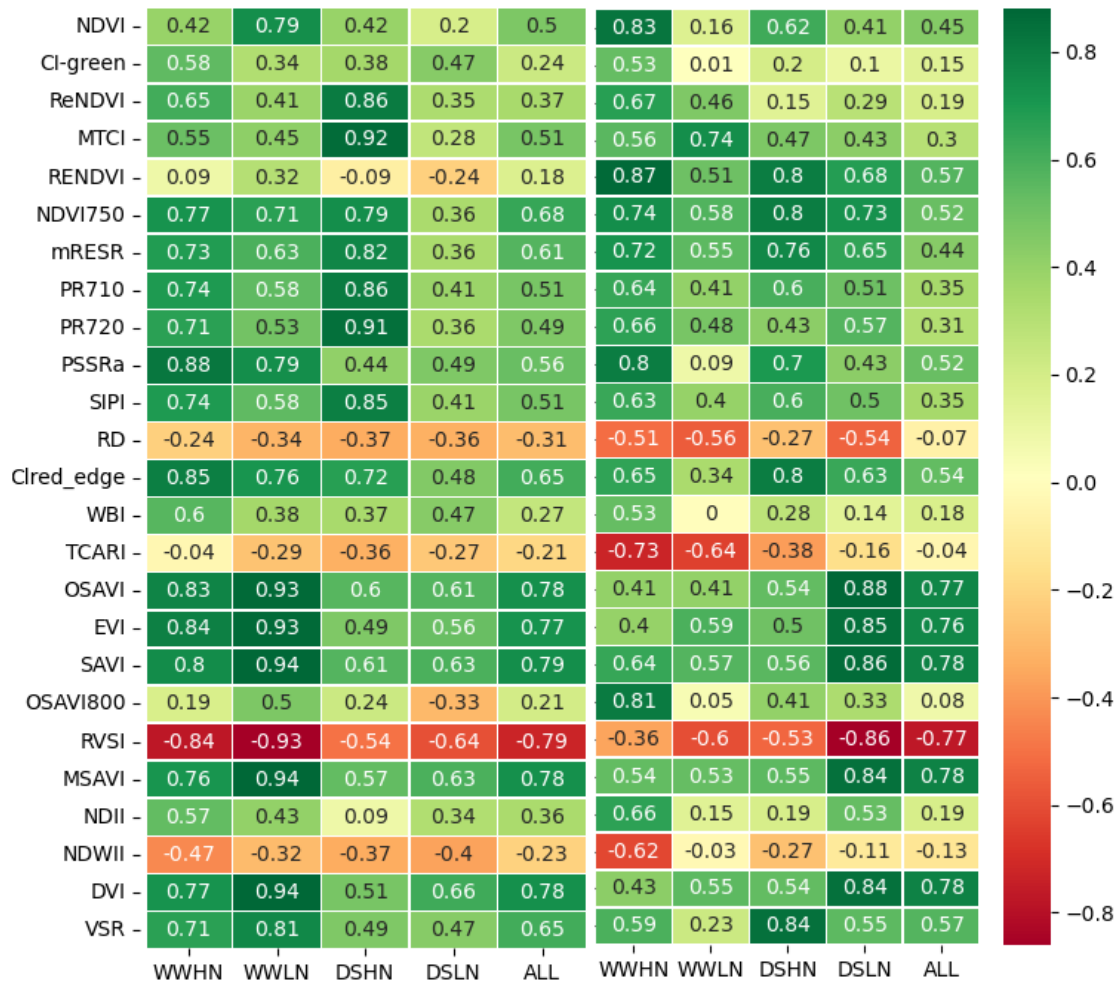
**Figure 5-5** Spectral reflectance of the averaged DSHN, DSLN, WWHN, and WWLN treatments for 0 DADS (a), 6 DADS (b), and 15 DADS (c). Spectral values are shown as mean  $\pm$  standard deviation. The WWHN and WWLN are the water well-watered plants with high and low nitrogen, respectively, while the DSHN and DSLN are the drought-stressed plants with high and low N levels, respectively

#### 5.4.2 Correlation between the known VIs and gas exchange measurements ( $P_n$ and $g_s$ )

Figure 5.6 displays heat maps of the Pearson correlation coefficient ( $r$ ) between the known VIs and the two gas exchange variables ( $P_n$  and  $g_s$ ). The correlation analysis was performed for each treatment and the combined treatments (ALL). Figure 5.6a shows good relationships between the VIs and  $g_s$  in the WWHN treatment. PSSRa had the highest correlation coefficient (0.88), while TCARI obtained the lowest ( $-0.04$ ). Similarly, most VIs had strong correlations in the WWLN treatment, with DVI, SAVI,

and MSAVI obtaining the highest correlation coefficient (0.94) and TCARI with the lowest (−0.29). However, a contrasting trend was observed in the DS-based treatment. With the DSHN, while the MTCI with  $r = 0.92$  had the best correlation with the  $g_s$ , a third of the extracted VIs had a poor relationship with the  $g_s$  ( $r < 0.4$ ). The DSLN treatment reported the lowest correlation with the  $g_s$  measurements. SAVI (0.79) and RSVI (−0.79) were observed to have the highest correlation, while NDVI (0.20) had the lowest correlation with the  $g_s$ . When the treatments were combined, most of the known VIs showed a positive correlation with the  $g_s$  except NDWII, RSVI, TCARI, and RD indices. The RSVI and SAVI had the highest correlation coefficients (−0.79 and 0.79, respectively), and the NDVI800 had the lowest (0.18).

From Figure 5.6b, a positive correlation was observed between the known VIs and the  $P_n$  measurements in the WWHN treatment, except for a few indices (NDWII, RSVI, TCARI, and RD), which were negatively correlated. Most indices had a high correlation coefficient, with RENDVI (0.87). The results obtained for the WWLN treatment show that compared to the WWHN, there was a relatively low correlation between the indices and the  $P_n$  measurements. MTCI had the best correlation (0.74) while the NDWII (−0.03) had the lowest. In the DSHN treatment, the VSR index obtained the highest correlation coefficient (0.84) and the NDII (0.19) had the lowest. Unlike the  $g_s$ , the indices in the DSLN treatment had relatively good correlations with the  $P_n$ . The highest correlation was observed in the OSAVI index (0.88), while the lowest was recorded in the CI\_green index (0.10). With the combined treatments, the MSAVI, SAVI, and DVI reported the highest correlation with the  $P_n$  (0.78), while the RD had the lowest correlation (−0.07).



(a) Stomatal Conductance (b) Net photosynthetic rate

**Figure 5-6** Pearson correlation between the extracted features and the gas exchange measurements ( $P_n$  and  $g_s$ ). VIs with a correlation of more than 0.5 were selected for further analysis. The heatmap highlights how different vegetation indices correlate with the four treatments (WWHN, WWLN, DSHN, DSLN). Strongly positive correlations (green cells) suggest that these indices are good predictors or indicators for the treatments, whereas negative correlations (red cells) indicate the opposite.

### 5.4.3 Waveband selection and proposed indices

#### 5.4.3.1 Spectral band pair correlation

Spectral band pair correlation is the correlation analysis between two specific spectral wavelengths. It aids in understanding the interactions of different spectral bands, providing insights into the characteristics of the observed traits. In HSI data, a significant portion of the spectrum suffers from multicollinearity, which refers to the presence of

strong correlations or dependencies between spectral wavelengths in a hyperspectral dataset. Figure 5.7 is a colormap showing the correlation between all pairs of wavebands of the data, providing a visual representation of how spectral features across a wide range of wavelengths are interrelated, indicating potential areas of redundancy or key spectral. The lighter color bands across the colormap suggest that certain wavelength ranges have high correlations with each other, while darker regions indicate lower correlations, suggesting that spectral features in these areas do not share much information or are not related linearly. The patterns off the diagonal (away from the central diagonal line) represent cross-wavelength correlations. Large blocks of uniform color off the diagonal indicate that features at one wavelength range are consistently correlated with features at another range. It is observed that the band pair correlation within the NIR region (700–1015 nm) follows a general pattern: adjacent bands had a strong tendency to correlate with one another. For example, around 394–450 nm, 740–790 nm, and 920–1015 nm, there are regions where spectral features are strongly correlated. However, among the highly correlated spectral bands in this region, a few bands had low correlations with their adjacent bands (940–953 nm). Spectral bands within the visible range (394–650 nm) had a low correlation between adjacent wavebands. However, bands between 511 nm and 576 nm were highly correlated to each other and to bands between 702 nm and 746 nm, showing the redundancy of some of these spectral features.

.

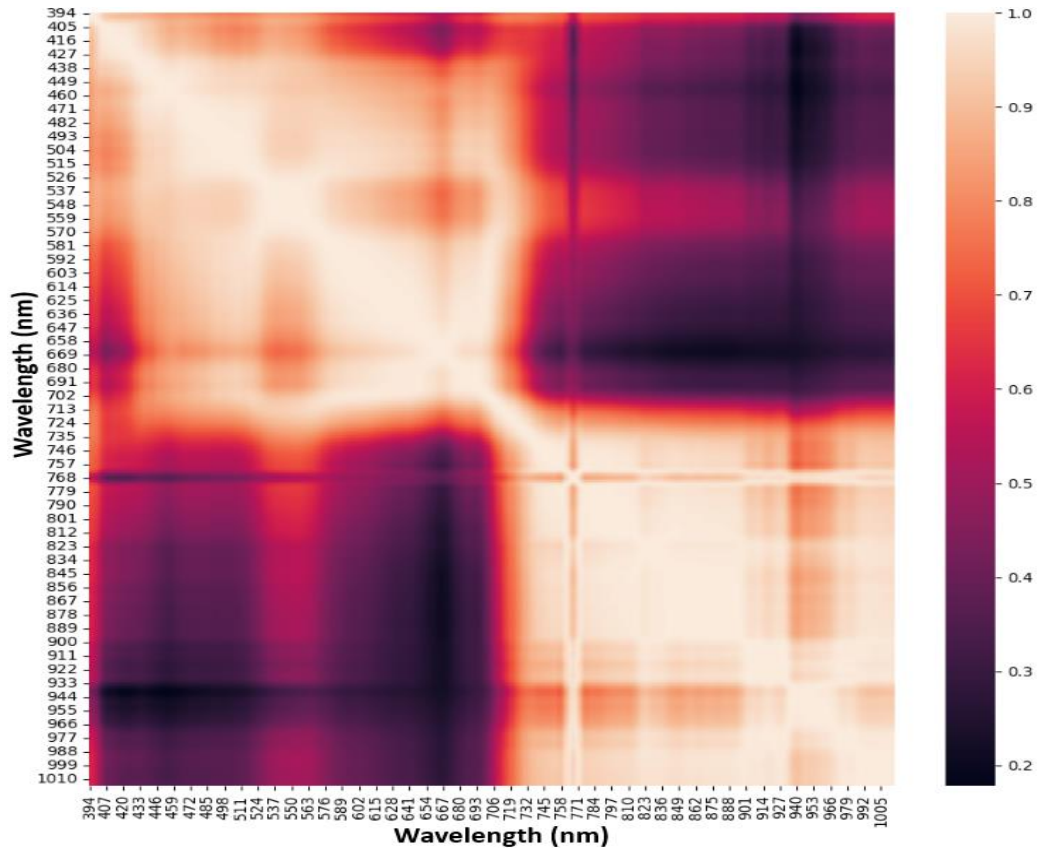


Figure 5-7 Spectral correlation matrix of hyperspectral wavelengths represented as a colourmap image of the correlation between all pairs of spectral features from 394 nm to 1015 nm. The colour scale on the right ranges from 0.1 (black) to 1.0 (light peach), with 1.0 indicating a very strong positive correlation (meaning the two wavelengths have very similar reflectance patterns across the data). The Lower values (dark purple/black) represent weaker correlations, (meaning the reflectance values at these wavelengths differ significantly)

#### 5.4.3.2 Output of the ensemble model waveband selection

Feature selection is critical in hyperspectral imaging analysis to enhance model performance by reducing overfitting, interpretability, and computational complexity (Koh et al., 2022). The high correlation between the different spectral wavebands shows the significance of feature selection in eliminating redundant information while keeping the relevant data. Each model in this study selected the sensitive spectral features based on the order of importance. Table 5-3 reports the top ten spectral wavelengths selected by the ensemble model. The chi-square method selected most of the top features within the green spectrum from 555–570 nm and the rest within the red regions (670–675 nm). The top ten features using the ReliefF method were within the red (674–690 nm), red edge (722 nm), and infrared (949–957 nm) regions. The CFS selection method identified the top ten features within the green regions (542–547 nm), the red region (669–671 nm), and



the infrared regions (939–957 nm). Generally, all three selection methods identified wavelength 674 nm as one of the most informative features. Also, the ReliefF and CFS identified 669, 949, 940, and 957 nm as the most sensitive features.

The output of the individual models was integrated with the Boruta SHAP model to improve the selection process and reduce the dimensions of the selected features. The output of the Boruta model was ranked in order of importance using the RF-RFE model. The ReliefF model ranked wavelengths in the green region as the top two (553 nm and 557 nm), followed by the red wavelengths (669 nm and 674 nm). Wavelength in the near-infrared region followed as the fifth most important feature. The 542 nm wavelength was selected as the tenth most important feature.

**Table 5-3** The top 10 wavelengths selected by the Chi-square, SFM-RF and Boruta feature selection methods and the ensemble model for the selection and ranking of features

| <b>Selected Wavelengths (nm)</b> |                   |                |            |            |
|----------------------------------|-------------------|----------------|------------|------------|
| <b>Rank</b>                      | <b>Chi-Square</b> | <b>ReliefF</b> | <b>CFS</b> | <b>RFE</b> |
| 1                                | 555               | 680            | 669        | <b>553</b> |
| 2                                | 554               | 689            | 674        | <b>557</b> |
| 3                                | 556               | 949            | 939        | <b>669</b> |
| 4                                | 553               | 722            | 936        | <b>674</b> |
| 5                                | 557               | 683            | 957        | <b>722</b> |
| 6                                | 552               | 674            | 949        | <b>940</b> |
| 7                                | 636               | 940            | 671        | <b>957</b> |
| 8                                | 673               | 670            | 547        | <b>636</b> |
| 9                                | 674               | 669            | 546        | <b>683</b> |
| 10                               | 672               | 957            | 542        | <b>542</b> |

### 5.4.3.3 Proposed Drought Stress Indices

The selected wavelengths were combined in different forms using the RDI, NDDI, and DDI equations to obtain the proposed drought stress indices. To analyze the relationship between the drought stress and the proposed indices, the indices with high  $R^2$  values were selected. From supplementary information (Table E-4), the NDDI and RDI produced the best indices that correlated highly with the gas exchange measurements ( $P_n$  and  $g_s$ ). Figure 5.8 is a correlation heatmap showing the relationship between the top ten proposed indices and the gas exchange measurements ( $P_n$  and  $g_s$ ). Based on the  $R^2$  values, the NDDI with wavelengths ( $\lambda_{940}$  and  $\lambda_{957}$ ) had the best correlation for  $P_n$  ( $r = -0.78$ ), whereas RDI with wavelengths ( $\lambda_{669}$  and  $\lambda_{636}$ ), ( $\lambda_{636}$  and  $\lambda_{542}$ ), and NDDI with wavelengths ( $\lambda_{940}$  and  $\lambda_{557}$ ) produced the best correlation ( $r = 0.67$ ) with  $g_s$ .

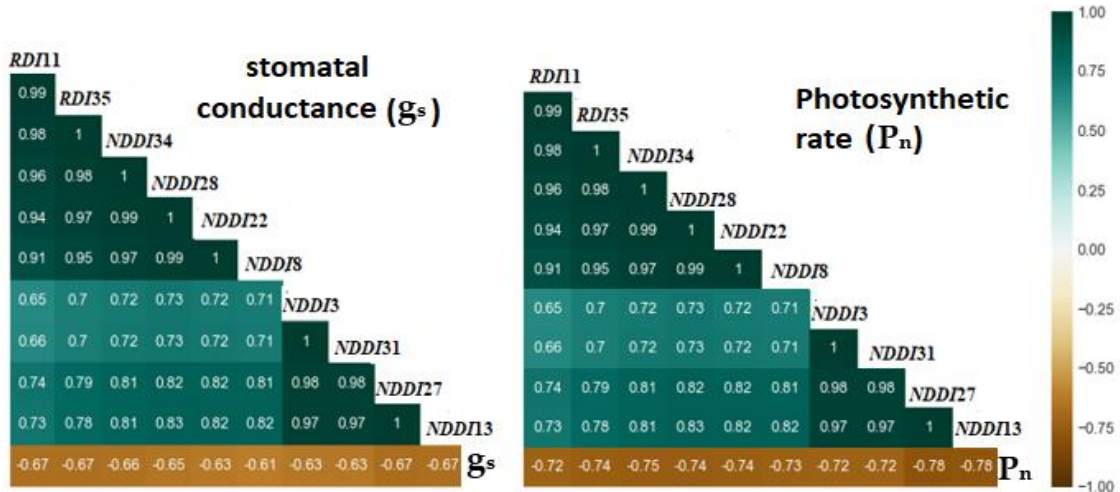


**Figure 5-8** Correlations heatmap showing the relationships between the proposed indices (RDI, DDI and NDDI) and the two gas exchange measurements ( $P_n$  and  $g_s$ ). The correlation coefficients are between 1 to -1 with the colour gradient ranging from dark teal (high positive correlation) to brown/orange (high negative correlation).

### 5.4.4 Machine learning-based drought detection

This section evaluated the performance of the three traditional classification models (RF, SVM, and DNN) on identifying plants with drought stress with different nitrogen levels. Four different training features were used: known VIs, proposed VIs, combined VIs (proposed and known VIs), and PCA-transformed features. The performances of the models were evaluated based on the type of training features used.

#### 5.4.4.1 Drought stress identification using machine learning models

Table 5-4 shows the performance of the models when trained with the known, proposed, combined VIs and PCA extracted features. Figure 5.9 depicts the confusion matrices of the three models on the test dataset. This depicts the performance of the models in classifying the treatments: WWHN, WWLN, DSHN, and DHLN, which correspond to the different nitrogen and drought stress levels. The confusion matrices are displayed as heatmaps, with the classification accuracy increasing with the depth of color. The number of predictions made by the model for each class is shown by the cells in the heatmap. The *x*-axis shows the actual class, while the *y*-axis displays the predicted class.

Figure 5.9a shows the performance of the various models trained with the known Vis, with the DNN model showing the best results. All the models had high accuracies in identifying the high N-based treatments (DSHN and WWHN), with the RF model achieving a perfect accuracy score for the DSHN treatment identification. However, the SVM model had low accuracy for WWHN treatment. For drought-stressed plants in low N-based treatments, particularly DSLN, increased misclassification of treatments was observed. All the models had high F1-scores and above 0.80 Cohen kappa scores, which shows good agreement between predicted and actual class labels. The study demonstrates a performance improvement in the SVM and DNN models when trained with the proposed models, with the DNN achieving the best performance (Figure 5.9b). The SVM and DNN had a perfect accuracy score for WWHN and SVM while having low misclassification for other treatments. A high F1-score (0.911) and relatively low Kappa score (0.881) for the RF model shows that the model has resulted in false positives in contrast to false negatives for classifying the treatments. From Figure 5.9c, all the models performed well when trained with the combined Vis, as evidenced by the high percentage of correctly predicted cases in the diagonal elements of the confusion matrices. The RF model performed best, followed by the SVM and DNN models. However, a one-way ANOVA test on the F1-scores shows no significant variation in the performance of the models (Supplementary Information, Table E-3). However, all the models showed comparably high misclassification for drought-based treatments. From the confusion matrices in Figure 5.9d, the three models had above 87% accuracy when trained with the PCA-transformed features. The RF model outperformed the others, producing above 0.90

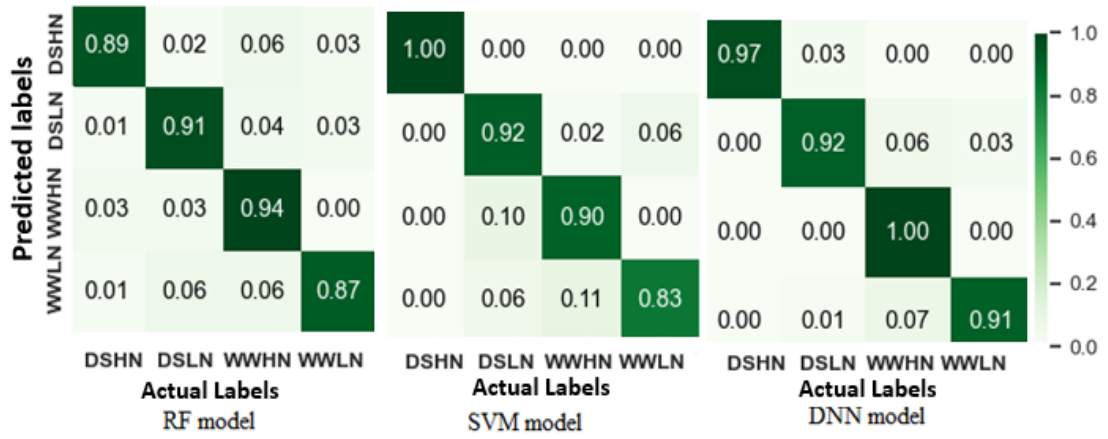
accuracy scores in all four classes. However, it struggled to identify DSHN, as 6% of the treatment was classified as WWLN. The SVM model followed the RF model, recording the best performance in DSLN classification with a 4% misclassification error.

**Table 5-4** Performance metrics for RF, SVM and DNN models for identification of drought stress plants (metrics include Average Accuracy (AA), F1-score, Kappa score)

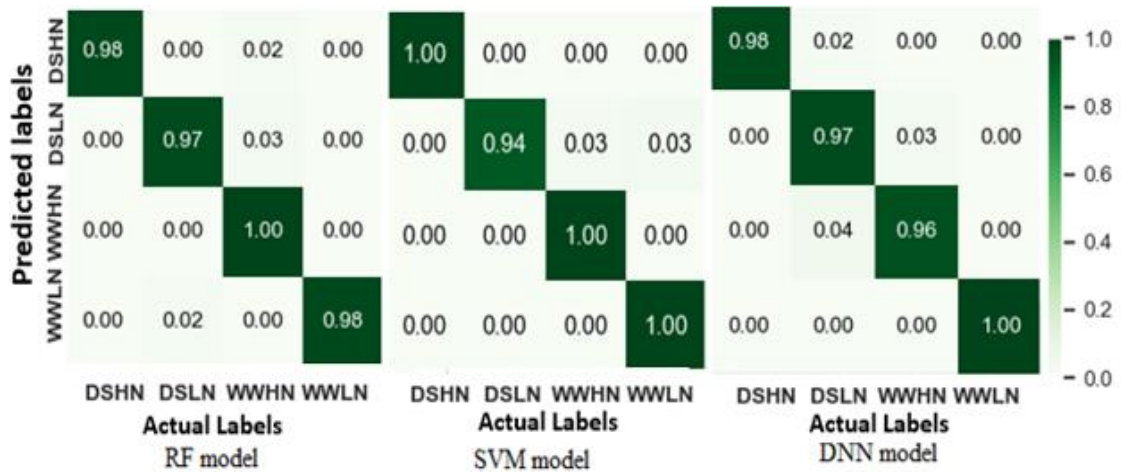
|                 | <b>Metrics</b> |           |                |              |
|-----------------|----------------|-----------|----------------|--------------|
| <b>Features</b> | <b>Model</b>   | <b>AA</b> | <b>F-Score</b> | <b>Kappa</b> |
| Known VIs       | RF             | 0.921     | 0.925          | 0.893        |
|                 | SVM            | 0.887     | 0.881          | 0.882        |
|                 | DNN            | 0.938     | 0.935          | 0.914        |
| Proposed VIs    | RF             | 0.914     | 0.911          | 0.881        |
|                 | SVM            | 0.924     | 0.930          | 0.919        |
|                 | DNN            | 0.948     | 0.949          | 0.933        |
| Combined VIs    | RF             | 0.983     | 0.984          | 0.965        |
|                 | SVM            | 0.981     | 0.982          | 0.975        |
|                 | DNN            | 0.977     | 0.979          | 0.969        |
| PCA Features    | RF             | 0.961     | 0.962          | 0.960        |
|                 | SVM            | 0.941     | 0.940          | 0.921        |
|                 | DNN            | 0.901     | 0.900          | 0.868        |



(a) Model trained with known Vis



(b) Model trained with proposed VIs



(c) Model trained with combined (known and proposed) VIs



(d) Model trained with PCA-transformed features

**Figure 5-9** Confusion matrices depicting the performance of the SVM, RF and DNN classifiers trained with (a) known VIs, (b) proposed VIs, (c) combined VIs and (d) PCA transformed features

#### 5.4.3.2 Multivariate model analysis for stomatal conductance and photosynthetic rate predictions

Figures 5.10 and 5.11 are scatter plots illustrating the relationship between the actual and predicted  $g_s$  and  $P_n$ . Table 5-5 also summarizes  $R^2$ , RMSE, and MAE, depicting the performance of the regression models. It was observed in Figure 10 that most of the models achieved high prediction accuracy, with the random forest regression (RFR) outperforming the others with  $R^2 = 0.87$ , RMSE = 0.035, and MAE = 0.015. The high  $R^2$  and low RMSE suggest that the RFR model is robust and insensitive to noise. In contrast, the PR model with an average  $R^2$  of 0.53 and RMSE of 0.52 was the least accurate in predicting the  $g_s$  values. During the PLSR modeling using the whole spectrum, optimal latent features were selected to train the model to avoid the curse of dimensionality. In this case, fewer latent variables with maximum  $R^2$  and minimum RMSE were selected. For the  $g_s$ -based PLSR models, 20 latent variables were selected. The PLSR model exhibited a high prediction score with a mean  $R^2$  of  $0.842 \pm 0.02$  for the test scores. Although the RFR and SVR were trained with a limited number of features (10 sensitive spectral features), their performances were as good as the PLSR model.

Figure 5.11 shows that all the models achieved considerably high performance in predicting the  $P_n$  values except the PR model, which had  $R^2 = 0.74$ . The RFR model exhibited the highest score with  $R^2 = 0.940 \pm 0.05$ , RMSE = 0.015, and MAE = 0.004.

Also, the PR model underperformed when trained with the combined VIs, achieving  $0.740 \pm 0.01$  ( $R^2$ ),  $0.144 \pm 0.281$  (RMSE), and  $0.127 \pm 0.04$  (MAE). Similarly, to the PLSR model for predicting  $g_s$ , 25 sensitive latent variables were selected for training the  $P_n$ -based PLSR model, which performed well with  $R^2 = 0.910 \pm 0.04$ , RMSE = 0.015, and MAE = 0.004.

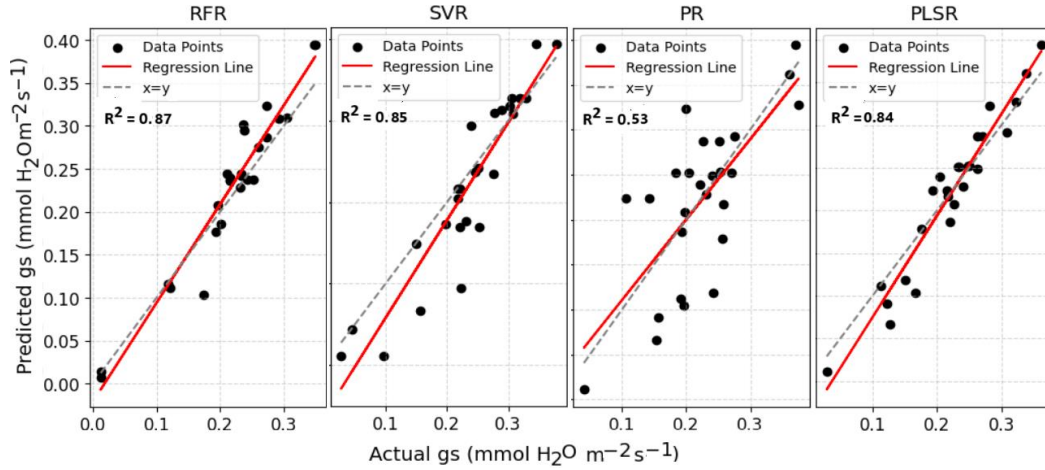


**Figure 5-10.** Prediction of plant  $g_s$  using four models (RF, SVR, PLSR, and PR). All the models were trained with the combined VIs except the PLSR, which were trained with the whole spectrum.

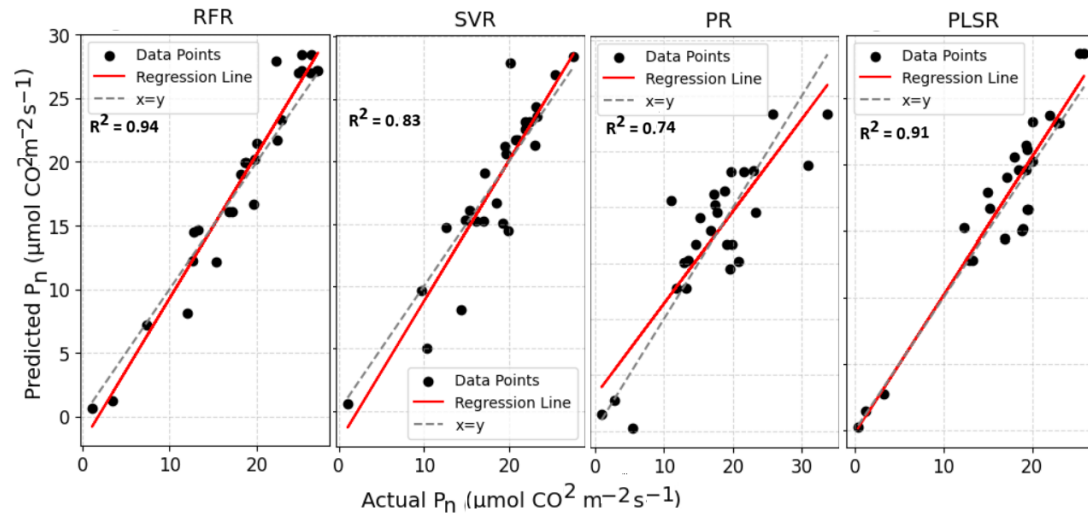


**Figure 5-11** Prediction of plant  $g_s$  using four models (Random Forest regression (RF), Support vector regression (SVR), Partial least square regression (PLSR) and Polynomial regression (PR)). All the models were trained with the combined VIs except the PLSR, which was trained with the whole spectra.

**Table 5-5** Performance of regression models for  $g_s$  and  $P_n$ 

| <b>Stomatal Conductance (<math>g_s</math>)</b> |            |            |           |             |
|------------------------------------------------|------------|------------|-----------|-------------|
| <b>Metrics</b>                                 | <b>RFR</b> | <b>SVR</b> | <b>PR</b> | <b>PLSR</b> |
| $R^2$                                          | 0.871      | 0.845      | 0.534     | 0.842       |
| RMSE                                           | 0.035      | 0.038      | 0.221     | 0.031       |
| MAE                                            | 0.015      | 0.011      | 0.142     | 0.017       |
| <b>Photosynthetic Rate (<math>P_n</math>)</b>  |            |            |           |             |
| <b>Metrics</b>                                 | <b>RFR</b> | <b>SVR</b> | <b>PR</b> | <b>PLSR</b> |
| $R^2$                                          | 0.940      | 0.830      | 0.740     | 0.910       |
| RMSE                                           | 0.015      | 0.063      | 0.144     | 0.018       |
| MAE                                            | 0.004      | 0.013      | 0.127     | 0.007       |

## 5.5 Discussion

This study utilized spectral information from hyperspectral imaging combined with different machine-learning models to identify drought stress in wheat species supplied with varying N levels. Since gas exchange measurements such as  $g_s$  and  $p_n$  are good indicators for drought stress monitoring, they were measured and analyzed in this study. From the results, the  $g_s$  detected drought stress on the three DADS since the control group (WWHN) and the DS treatments significantly differed on this day. Drought stress was detectable on the six DADS using the  $P_n$  measurements, indicating the viability of using gas exchange measurements to track drought dynamics in wheat species. This supports the findings of (Flexas et al., 2009), who analyzed the dynamics of plant water stress and recovery using the photosynthetic parameters, as  $g_s$  and mesophyll conductance  $CO_2$  ( $g_m$ ), where the  $g_s$  declined to 0.1 and less than  $0.05 \text{ mol } CO_2 \text{ m}^{-2} \text{ s}^{-1}$  in moderately and severely water-stressed plants, respectively.

Analysis of the spectral curves of the hyperspectral data reveals differences in treatments in the visible and near-infrared regions. The low reflectance of all the treatments in the visible regions was expected due to the strong chlorophyll absorption of the plants. This agrees with multiple studies that stipulate that healthy plants have low reflectance in the visible region, while high reflectance is normally observed in chlorophyll-depleted plants



(Hansen & Schjoerring, 2003)(Ayala-Silva & Beyl, 2005). Although the low N treatments had a relatively low N content, they could not cause high reflectance, probably because the plants had a high concentration of chlorophyll at the start of the experiments. Nitrogen affects the overall plant health and leaf cellular structure of plants, where adequate N content builds strong cellular structure while degraded structure is observed in the plants with low N. The treatments with high N (WWHN and DSHN) with a strong cellular structure scattered light more effectively, which was evidenced in high reflectance in the red edge regions. The red edge region for these treatments is because of the leaf internal structure, which caused a rapid transition from strong absorption in the visible region to strong reflectance in the NIR region (Knippling, 1970). The high reflectance of the DSHN and WWLN treatments in the visible regions on the six DADS could be attributed to the decreased leaf pigments caused by the drought stress and low N level, respectively, which affected the leaf chlorophyll content (Debnath et al., 2021). The DSLN showing a relatively low reflectance in the NIR region shows that plants under drought stress with low nitrogen experience significant stress, which leads to reduced NIR reflectance. The drought and low N levels causes a cellular structural degradation, leading to a low reflectance in the NIR region.

Twenty-five known VIs were extracted, with some VIs (RVSI, SAVI, NDVI705, NDVI750, and EVI) revealing high correlations with the two gas exchange measurements. From previous studies (B. Gao, 1996)(Haboudane et al., 2002), some of these VIs are associated with plant N, which shows that these indices may not be effective for drought stress analysis. Hence, further analysis proposing drought spectral indices was performed. Due to the high dimensionality of HSI, spectral averaging and sensitivity analysis were performed to select the wavebands responsive to drought stress and nitrogen deficiency. Over 600 features were discarded during the feature selection process, revealing that a small subset of spectral features could capture a significant amount of the most valuable information. In contrast, most of the remaining features were typically redundant or contributed to noise (Moghimini et al., 2019). The selection of sensitive spectral features in the red edge and green spectral regions shows the responsiveness of the wavelengths in these regions to both N variations and drought stress. Additionally, the selection and ranking of the 553 nm wavelength by the ensemble

model as the most sensitive shows that it could decipher a physical meaning hidden in the high-dimensional spectral data of drought-stressed plants irrespective of the nitrogen level present.

The evaluation of the proposed VIs using a spectral combination of the sensitive features selected revealed that the NDDI and RDI-based indices had strong relationships with drought stress, with NDDI30 and NDI20 having the strongest  $R^2$  values of 0.78 and 0.69, respectively. The proposed indices were primarily derived from wavelengths in the blue (500, 550, and 580 nm) and near-infrared (710, 760, 770, and 783 nm) regions, which are reported as wavebands commonly used to measure drought stress and plant N status (Thenkabail et al., 1995). This finding is confirmed by the work of Colovic et al. (Colovic et al., 2022), who identified the double difference index (DDI) produced from the near-infrared regions (749, 720, and 701 nm) as the best-performing index in explaining the variation in plant water levels. This shows the relatively important function of the near-infrared spectrum in drought stress analysis and that a single band might not be practical to evaluate plant health (drought stress) due to plant fluctuating nitrogen status and the dynamic nature of the drought stress.

Three models (SVM, RF, and DNN) were developed to identify stressed plants using the known VIs, the proposed Vis, and the combined VIs and PCA-transformed dataset. Generally, all three classifiers had outstanding performances. This shows that traditional machine learning models such as SVM and RF perform well in detecting drought stress in wheat species with the right feature selection. Although the models performed well when trained with the known or proposed VIs, their performance significantly improved when the two sets of VIs were combined. This could be because the combined VIs features cover the spectral areas strongly linked to both nitrogen and drought stress. For instance, indices such as OSAVI, RVSI, and NDDI30 are derived from the green and near-infrared regions, which strongly correlate to nitrogen concentration and drought stress. The PCA method reduced the dimensionality of the hyperspectral dataset (622 wavebands) to two feature sets representing the principal components. While PCA produces new feature components that maximize data variance, ignoring the lower-order PC components may have resulted in information loss, which affected the performance of the models (Cheriyadat, 2003).

Finally, while previous studies show that gas exchange measurements can monitor drought stress dynamics in plants, the process is costly and has low throughput. This study demonstrates the capability of training regression models using the combined VIs to accurately predict  $g_s$  and  $P_n$ , as shown in Section 3.5.2; however, it is essential to note some limitations. The combined VIs were formulated from different spectral data, resulting in non-linear and complex data. Due to the non-linear and non-addictive nature of the data, the PR model was unable to learn the non-linear trend in the data.

This study contributes to the advancement of precision agriculture demonstrating the effectiveness of combined VIs with machine learning models for drought stress monitoring and gas exchange measurement prediction. The high accuracy in identifying stress despite the variable N levels shows the potential for precise and reliable stress monitoring. This study offers insights into the interactions between variable N levels and drought stress dynamics. The methodology has strong potential for large-scale application in agriculture, becoming a baseline for large-scale drought stress monitoring and management. This can enhance high crop yield by optimizing irrigation and nutrient application. Although the study was conducted in a glasshouse, the scalability of hyperspectral imaging combined with machine learning modeling means that this method can be deployed in large-scale agricultural monitoring systems for high-throughput phenotyping.

## **5.6 Conclusion**

In the field, crops can suffer multiple stresses, such as drought and nutrient deficiencies, affecting their yield and overall production cost. While these stresses have separate effects on plants, they also interact and have some common responses. Hence, identifying key traits to monitor drought-stressed plants at variable N status is crucial to improving crop yield. This study utilized spectral information from HSI combined with machine learning to identify drought stress and predict gas exchange measurements ( $g_s$  and  $P_n$ ) in wheat species. The experimental results demonstrated the capability of our proposed ensemble model in selecting spectral features that are highly responsive to drought dynamics in plants. A combination of the selected features resulted in proposed VIs, which achieved

high accuracy in identifying the drought-stressed plants compared to using known VIs when used to train traditional machine learning models. However, the performance of the models improved significantly when they were trained with a combination of proposed and known VIs, where DNN, RF, and SVM improved by 2.5%, 1.2%, and 1.8%, respectively, compared to the models trained with the known VIs. In addition to identifying drought stress by classification, the combined VIs were also used to train three regression models to predict  $g_s$  and  $P_n$  measurements. The RF regression model outperformed the others in accurately predicting  $g_s$  and  $P_n$  with error margins of 0.8 and 0.4, respectively. Although excellent results were obtained, more research is required to validate the conclusions on a larger spatial scale, exploring the potential application of the models for other plant species.

## Acknowledgements

This research was supported by the OCP S.A. under the University of Mohammed VI Polytechnic, Rothamsted Research and Cranfield University project (FP04). Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) through the Delivering Sustainable Wheat program (BB/X011003/1).

## 5.7 Reference

- Anas, M., Liao, F., Verma, K. K., Sarwar, M. A., Mahmood, A., Chen, Z. L., Li, Q., Zeng, X. P., Liu, Y., & Li, Y. R. (2020). Fate of nitrogen in agriculture and environment: agronomic, eco-physiological and molecular approaches to improve nitrogen use efficiency. *Biological Research*, 53(1), 1–20. <https://doi.org/10.1186/s40659-020-00312-4>
- Asaari, M. S. M., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., & Scheunders, P. (2019). Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Computers and Electronics in Agriculture*, 162(May), 749–758. <https://doi.org/10.1016/j.compag.2019.05.018>
- Ayala-Silva, T., & Beyl, C. A. (2005). Changes in spectral reflectance of wheat leaves in response to specific macronutrient deficiency. *Advances in Space Research*, 35(2), 305–317. <https://doi.org/10.1016/j.asr.2004.09.008>
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization.

- Blackburn, G. A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing*, 19(4), 657–675. <https://doi.org/10.1080/014311698215919>
- Breiman, L. (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12343 LNCS, 503–515. [https://doi.org/10.1007/978-3-030-62008-0\\_35](https://doi.org/10.1007/978-3-030-62008-0_35)
- Brochu, E., Cora, V. M., & de Freitas, N. (2010). *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning*. <http://arxiv.org/abs/1012.2599>
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). *API design for machine learning software: experiences from the scikit-learn project*. 1–15. <http://arxiv.org/abs/1309.0238>
- Buschmann, C., & Nagel, E. (1993). In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*, 14(4), 711–722. <https://doi.org/10.1080/01431169308904370>
- Cheriyadat, A. (2003). *Limitations of principal component analysis for dimensionality-reduction for classification of hyperspectral data*. December. <http://en.scientificcommons.org/49172179>
- Chutia, D., Bhattacharyya, D. K., Sarma, J., & Raju, P. N. L. (2017). An effective ensemble classification framework using random forests and a correlation based feature selection technique. *Transactions in GIS*, 21(6), 1165–1178. <https://doi.org/10.1111/tgis.12268>
- Colovic, M., Yu, K., Todorovic, M., Cantore, V., Hamze, M., Albrizio, R., & Stellacci, A. M. (2022). Hyperspectral Vegetation Indices to Assess Water and Nitrogen Status of Sweet Maize Crop. *Agronomy*, 12(9), 1–17. <https://doi.org/10.3390/agronomy12092181>
- Damodaran, B. B., Courty, N., & Lefevre, S. (2017). Sparse Hilbert Schmidt Independence Criterion and Surrogate-Kernel-Based Feature Selection for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(4), 2385–2398. <https://doi.org/10.1109/TGRS.2016.2642479>
- Datt, B. (1999). A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using Eucalyptus leaves. *Journal of Plant Physiology*, 154(1), 30–36. [https://doi.org/10.1016/S0176-1617\(99\)80314-9](https://doi.org/10.1016/S0176-1617(99)80314-9)

- Debnath, S., Paul, M., Motiur Rahaman, D. M., Debnath, T., Zheng, L., Baby, T., Schmidtke, L. M., & Rogiers, S. Y. (2021). Identifying individual nutrient deficiencies of grapevine leaves using hyperspectral imaging. *Remote Sensing*, 13(16), 1–21. <https://doi.org/10.3390/rs13163317>
- Duan, L., Han, J., Guo, Z., Tu, H., Yang, P., Zhang, D., Fan, Y., Chen, G., Xiong, L., Dai, M., Williams, K., Corke, F., Doonan, J. H., & Yang, W. (2018). Chapter 3. Novel digital features discriminate between drought resistant and drought sensitive rice under controlled and field conditions. *Frontiers in Plant Science*, 9(April). <https://doi.org/10.3389/fpls.2018.00492>
- Flexas, J., Barón, M., Bota, J., Ducruet, J. M., Gallé, A., Galmés, J., Jiménez, M., Pou, A., Ribas-Carbó, M., Sajnani, C., Tomàs, M., & Medrano, H. (2009). Photosynthesis limitations during water stress acclimation and recovery in the drought-adapted Vitis hybrid Richter-110 (V. berlandieri×V. rupestris). *Journal of Experimental Botany*, 60(8), 2361–2377. <https://doi.org/10.1093/jxb/erp069>
- Gamon, J.A, P. and C. B. F. (1992). A Narrow-Waveband Spectral Index That Tracks Diurnal Changes in Photosynthetic Efficiency. *REMOTE SENS. ENVIRON*, 6(1), 22–42.
- Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. [https://doi.org/https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gastal, F., & Lemaire, G. (2002). N uptake and distribution in crops: An agronomical and ecophysiological perspective. *Journal of Experimental Botany*, 53(370), 789–799. <https://doi.org/10.1093/jexbot/53.370.789>
- Gastal, F., Lemaire, G., Durand, J.-L., & Louarn, G. (2015). Chapter 8 - Quantifying crop responses to nitrogen and avenues to improve nitrogen-use efficiency (V. O. Sadras & D. F. B. T.-C. P. (Second E. Calderini (eds.); pp. 161–206). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-417104-6.00008-X>
- Gitelson, A. A., Gritz †, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, 160(3), 271–282. <https://doi.org/https://doi.org/10.1078/0176-1617-00887>
- Grant, O. M., Ochagavía, H., Baluja, J., Diago, M. P., & Tardáguila, J. (2016). Thermal imaging to detect spatial and temporal variation in the water status of grapevine (Vitis vinifera L.). *The Journal of Horticultural Science and Biotechnology*, 91(1), 43–54. <https://doi.org/10.1080/14620316.2015.1110991>
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of

- crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3), 337–352. <https://doi.org/10.1016/j.rse.2003.12.013>
- Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81(2–3), 416–426. [https://doi.org/10.1016/S0034-4257\(02\)00018-4](https://doi.org/10.1016/S0034-4257(02)00018-4)
- Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86(4), 542–553. [https://doi.org/10.1016/S0034-4257\(03\)00131-7](https://doi.org/10.1016/S0034-4257(03)00131-7)
- Helmholz, P., Rottensteiner, F., & Heipke, C. (2014). Semi-automatic verification of cropland and grassland using very high resolution mono-temporal satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 97, 204–218. <https://doi.org/10.1016/j.isprsjprs.2014.09.008>
- Ihuoma, S. O., & Madramootoo, C. A. (2017). Recent advances in crop water stress detection. *Computers and Electronics in Agriculture*, 141, 267–275. <https://doi.org/10.1016/j.compag.2017.07.026>
- Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, 11(3), 185–200. [https://doi.org/10.1016/S0167-5877\(05\)80004-2](https://doi.org/10.1016/S0167-5877(05)80004-2)
- Jay, S., Gorretta, N., Morel, J., Maupas, F., Bendoula, R., Rabatel, G., Dutartre, D., Comar, A., & Baret, F. (2017). Estimating leaf chlorophyll content in sugar beet canopies using millimeter- to centimeter-scale reflectance imagery. *Remote Sensing of Environment*, 198, 173–186. <https://doi.org/10.1016/j.rse.2017.06.008>
- Keshtiban, R. K., Carvani, V., & Imandar, M. (2015). Effects of salinity stress and drought due to different concentrations of sodium chloride and polyethylene glycol 6000 on germination and seedling growth characteristics of pinto bean (*Phaseolus vulgaris* L.). *Advances in Environmental Biology*, 237+.
- Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, 1(3), 155–159. [https://doi.org/10.1016/S0034-4257\(70\)80021-9](https://doi.org/10.1016/S0034-4257(70)80021-9)
- Koh, J. C. O., Banerjee, B. P., Spangenberg, G., & Kant, S. (2022). Automated hyperspectral vegetation index derivation using a hyperparameter optimisation framework for high-throughput plant phenotyping. *New Phytologist*, 233(6), 2659–2670. <https://doi.org/10.1111/nph.17947>

- Lawlor, D. W. (2002). Limitation to photosynthesis in water-stressed leaves: Stomata vs. Metabolism and the role of ATP. *Annals of Botany*, 89(SPEC. ISS.), 871–885. <https://doi.org/10.1093/aob/mcf110>
- Leone, M. (2022). Advances in fiber optic sensors for soil moisture monitoring: A review. *Results in Optics*, 7(December 2021), 100213. <https://doi.org/10.1016/j.rio.2022.100213>
- Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A. L., Manikandan, R., & Gandomi, A. H. (2024). Classification models combined with Boruta feature selection for heart disease prediction. *Informatics in Medicine Unlocked*, 44(January), 101442. <https://doi.org/10.1016/j.imu.2023.101442>
- McFeeters, S. K. (1996). NDWI BY McFEETERS. *Remote Sensing of Environment*, 25(3), 687–711.
- Mertens, S., Verbraeken, L., Sprenger, H., Demuynck, K., Maleux, K., Cannoot, B., De Block, J., Maere, S., Nelissen, H., Bonaventure, G., Crafts-Brandner, S. J., Vogel, J. T., Bruce, W., Inzé, D., & Wuyts, N. (2021). Proximal Hyperspectral Imaging Detects Diurnal and Drought-Induced Changes in Maize Physiology. *Frontiers in Plant Science*, 12(February), 1–18. <https://doi.org/10.3389/fpls.2021.640914>
- Moghim, A., Yang, C., & Marchetto, P. M. (2019). *Thesis- Integrating Hyperspectral Imaging and Artificial Intelligence to Develop Automated Frameworks for High-throughput Phenotyping in Wheat*. February.
- Nagasubramanian, K., Jones, S., Singh, A. K., Singh, A., Ganapathysubramanian, B., & Sarkar, S. (2018). *Explaining hyperspectral imaging based plant disease identification: 3D CNN and saliency maps*. *Nips 2017*, 3–10. <http://arxiv.org/abs/1804.08831>
- Nguyen, N. T., Mohapatra, P. K., Fujita, K., Nakabayashi, K., & Thompson, J. (2003). Effect of nitrogen deficiency on biomass production, photosynthesis, carbon partitioning, and nitrogen nutrition status of Melaleuca and Eucalyptus species. *Soil Science and Plant Nutrition*, 49(1), 99–109. <https://doi.org/10.1080/00380768.2003.10409985>
- Okyere, F. G., Cudjoe, D., Sadeghi-Tehran, P., Virlet, N., Riche, A. B., Castle, M., Greche, L., Simms, D., Mhada, M., Mohareb, F., & Hawkesford, M. J. (2023). Modeling the spatial-spectral characteristics of plants for nutrient status identification using hyperspectral data and deep learning methods. *Frontiers in Plant Science*, 14(October), 1–17. <https://doi.org/10.3389/fpls.2023.1209500>
- Pandey, P., Ge, Y., Stoerger, V., & Schnable, J. C. (2017). High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging. *Frontiers in Plant Science*, 8(August), 1–12. <https://doi.org/10.3389/fpls.2017.01348>



- Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31(2), 221–230.
- Peñuelas, J., & Filella, L. (1998). Technical focus: Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, 3(4), 151–156. [https://doi.org/10.1016/S1360-1385\(98\)01213-8](https://doi.org/10.1016/S1360-1385(98)01213-8)
- Perry, C. R., & Lautenschlager, L. F. (1984). Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment*, 14(1–3), 169–182. [https://doi.org/10.1016/0034-4257\(84\)90013-0](https://doi.org/10.1016/0034-4257(84)90013-0)
- Pirasteh-Anosheh, H., Saed-Moucheshi, A., Pakniyat, H., & Pessarakli, M. (2016). Stomatal responses to drought stress. *Water Stress and Crop Plants: A Sustainable Approach*, 1–2, 24–40. <https://doi.org/10.1002/9781119054450.ch3>
- Podani, J., & Czárán, T. (1997). Individual-centered analysis of mapped point patterns representing multi-species assemblages. *Journal of Vegetation Science*, 8(2), 259–270. <https://doi.org/10.2307/3237355>
- Proctor, C., Dao, P. D., & He, Y. (2021). Close-range, heavy-duty hyperspectral imaging for tracking drought impacts using the PROCOSINE model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 263, 107528. <https://doi.org/10.1016/j.jqsrt.2021.107528>
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119–126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)
- Rady, A., Ekramirad, N., Adedeji, A. A., Li, M., & Alimardani, R. (2017). Hyperspectral imaging for detection of codling moth infestation in GoldRush apples. *Postharvest Biology and Technology*, 129, 37–44. <https://doi.org/https://doi.org/10.1016/j.postharvbio.2017.03.007>
- Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, 112(February), 103375. <https://doi.org/10.1016/j.compbimed.2019.103375>
- Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>

- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2), 95–107. [https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7)
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Sadeghi-Tehran, P., Virlet, N., & Hawkesford, M. J. (2021). A neural network method for classification of sunlit and shaded components of wheat canopies in the field using high-resolution hyperspectral imagery. *Remote Sensing*, 13(5), 1–17. <https://doi.org/10.3390/rs13050898>
- Satterwhite, M., & Henley, J. (1990). *Hyperspectral signatures (400 to 2500 nm) of vegetation, minerals, soils, rocks, and cultural features: Laboratory and field measurements*. 478. <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA239496>
- Seleiman, M. F., Al-Suhaibani, N., Ali, N., Akmal, M., Alotaibi, M., Refay, Y., Dindaroglu, T., Abdul-Wajid, H. H., & Battaglia, M. L. (2021). Drought stress impacts on plants and different approaches to alleviate its adverse effects. *Plants*, 10(2), 1–25. <https://doi.org/10.3390/plants10020259>
- Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2), 337–354. [https://doi.org/https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/https://doi.org/10.1016/S0034-4257(02)00010-X)
- Sun, Y., Todorovic, S., & Goodison, S. (2010). Local-learning-based feature selection for high-dimensional data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 1610–1626. <https://doi.org/10.1109/TPAMI.2009.190>
- Thameur, A., Lachiheb, B., & Ferchichi, A. (2012). Drought effect on growth, gas exchange and yield, in two strains of local barley Ardhaoui, under water deficit conditions in southern Tunisia. *Journal of Environmental Management*, 113, 495–500. <https://doi.org/https://doi.org/10.1016/j.jenvman.2012.05.026>
- Thenkabail, P. S., Smith, R. B., & De Pauw, E. (1995). Wiegand and Richardson, † International Center for Agricultural Research in the Dry Areas 1990), natural vegetation (Friedl et al., 1994), and in (ICARDA). *Environ*, 71(99), 158–182.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)

- White, D. C., Williams, M., & Barr, S. L. (2008). Detecting sub-surface soil disturbance using hyperspectral first derivative band ratios of associated vegetation stress. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 37(Figure 1).
- Xu, H. R., Ying, Y. B., Fu, X. P., & Zhu, S. P. (2007). Near-infrared Spectroscopy in detecting Leaf Miner Damage on Tomato Leaf. *Biosystems Engineering*, 96(4), 447–454. <https://doi.org/10.1016/j.biosystemseng.2007.01.008>
- Xu, L., & Baldocchi, D. D. (2003). Seasonal trends in photosynthetic parameters and stomatal conductance of blue oak (*Quercus douglasii*) under prolonged summer drought and high temperature. *Tree Physiology*, 23(13), 865–877. <https://doi.org/10.1093/treephys/23.13.865>
- Yang, P., Liu, W., Zhou, B. B., Chawla, S., & Zomaya, A. Y. (2013). Ensemble-based wrapper methods for feature selection and class imbalance learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7818 LNAI(PART 1), 544–555. [https://doi.org/10.1007/978-3-642-37453-1\\_45](https://doi.org/10.1007/978-3-642-37453-1_45)
- Zhang, Y., Zha, Y., Jin, X., Wang, Y., & Qiao, H. (2022). Changes in Vertical Phenotypic Traits of Rice (*Oryza sativa* L.) Response to Water Stress. *Frontiers in Plant Science*, 13(July), 1–19. <https://doi.org/10.3389/fpls.2022.942110>
- Zhu, F., Zhang, D., He, Y., Liu, F., & Sun, D. W. (2013). Application of Visible and Near Infrared Hyperspectral Imaging to Differentiate Between Fresh and Frozen-Thawed Fish Fillets. *Food and Bioprocess Technology*, 6(10), 2931–2937. <https://doi.org/10.1007/s11947-012-0825-6>

## 6 General Discussions

This research aimed to develop and implement a robust image-based plant phenotyping system capable of detecting and analysing plant phenotypic traits for nutrient and drought stress identification and prediction. These traits include morphological (height, aerial density, biomass), colour, texture and spectral features; however, the selection depends on the objective of the studies. Several workflows and algorithms were developed to extract features, process, and analyse them to understand their response to plant nutrients and drought stress. To optimise the workflows developed for this research, finding the right balance between the sensor technologies for data collection and processing and selecting the appropriate plant phenological growth stage is required. For this reason, two primary imaging sensors were employed in this research: digital RGB and hyperspectral cameras. The pipelines were tested and validated on cowpea, quinoa and wheat species at selected growth stages. Here, the Zadok's scale (Zadoks & Board, 1999) was used to recognize the different phenological stages by growth codes from GS 00 – GS 99 (for quinoa and cowpea) and GS 00 - GS 56 (for wheat). This research mainly assessed two stresses: nutrition (nitrogen and phosphorus) and drought.

Four new methods were proposed and developed for data analysis, contributing to a better understanding of plant nutrient and drought stress dynamics and how they affect growth and yield. The first method (Chapter 2) is a proposed image processing pipeline with a novel segmentation algorithm to segment plants from complex backgrounds affected by illumination in controlled and field environments. Furthermore, nitrogen- and phosphorus-sensitive plant traits were extracted, and a statistical comparison was made of the performance of five machine learning models to predict nitrogen and phosphorus levels. The second method (Chapter 3) involves the use of statistical and machine learning algorithms to select sensitive plant traits that are responsive to plant drought stress under varying nitrogen levels. In the third method (Chapter 4), a pre-and post-processing pipeline was developed to characterize the dynamics of plant spatial and spectral features for identifying nitrogen and phosphorus status at five growth stages. The final method (Chapter 5) proposes new vegetation indices to identify plant drought stress subjected to different nitrogen levels.

## **6.1 Plant stress detection using image processing systems.**

### **6.1.1 Digital imaging for stress analysis**

#### **6.1.1.1 Nutrient stress detection**

This study (Chapters 2 and 3) employed a digital camera for plant nutrient status and drought stress analysis. This sensor outputs a high spatial resolution image with three broad bands of red, blue and green colour components. It was adopted due to its high resolution, low cost, and portable size.

To process digital images, segmentation plays a vital role in extracting quality phenotypic data. A segmentation algorithm using a discrimination analysis based on selected colour features of plants combined with machine learning methods was developed in Chapter 2. The training features were extracted from multiple colour spaces (rgb, hsv, Lab, ybr, hls, yuv, luv, and xyz) that has high discriminative power and provide ideal ways to minimize illumination and shadow effect. In similar studies, Khan & AlGhamdi (2023) emphasize the significance of Lab and HSV colour spaces in segmenting plant vegetation under varied conditions. These features are device-independent and proven to perform well in segmenting vegetation under variable illumination conditions (Coy et al., 2016). Although individual colour spaces could be used as a single input to train classification models for segmentation, using multiple colour spaces is more robust to background noise and illumination variations. Similar to this work, Sadeghi-Tehran et al. (2017) developed a 21-colour feature vector from multiple colour spaces to segment wheat species in the field. Their high-dimensional training features can be computationally challenging, especially for large datasets requiring high-powered computer memory. In the proposed algorithm, the high-dimensional feature vector was reduced using a recursive feature elimination technique, which selects the top 12 features using importance score criteria. The selected features were non-redundant, had high contrast, and were illumination invariant.

This study shows that training a feed-forward neural network (MLP) yields a robust tool for vegetation segmentation. It was validated on cowpea, quinoa, and wheat species with different structural and morphological architectures, indicating a high segmentation performance on these plant species. Although the proposed segmentation method is

effective, it fails in specific scenarios, such as segmenting plants with brownish to black blotches resulting from prolonged stress (drought or nitrogen deficits). Because the blotches have similar colour characteristics as the soil (background), the proposed tool, in this instance, identifies them as background and eliminates them during segmentation. Moreover, extreme sunlight reflections affected the performance of this tool. Although the training features comprised image patches from variable light conditions, extreme light reflection caused strong heterogeneous flare artefacts, which affected the segmentation quality. Further work was done to improve the performance of this tool. The improved algorithm was based on colour and texture (Gabor, local binary patterns and entropy) features, unlike the previous one, which was trained on only colour features. In addition, it was ensured that the extracted patches were well distributed across the selected images. The improved version is robust and had comparable performance as the state-of-the-art UNet model. It effectively segmented plants with different visual colours and nutrient stress. This work demonstrates that the proposed segmentation approach can be a vital tool for data analysis pipeline development for high-throughput plant phenotyping. The pipeline developed based on this sensor could be deployed in edge devices for high-throughput phenotyping of glasshouse and field-based crops.

Several studies (Adams et al., 2020) have shown that with quality segmentation, colour features have close relationships with plant nutrients. For plant nutrient analysis, the findings from this study show that features such as DGCI, RGI GBI, and BBI were highly correlated to plants' nitrogen and phosphorus contents. Ideally, the DGCI feature contributed over 30% of the information required to train the ML models using a feature importance technique. This agrees with Shi et al. (2021), who suggested that DGCI could be used to monitor plant nitrogen deficiencies. While colour features from plant images can be valuable for detecting plant nutrient stress, this assertion is plant growth stage-dependent. For instance Wang et al. (2014) reported a high  $R^2$  for RGB features in estimating rice leaf N concentration.

#### **6.1.1.2 Drought stress detection**

The study in Chapter 2 shows that colour features provide important information for nutrient stress analysis; nevertheless, it is challenging to differentiate multiple stresses (such as nitrogen deficiency and drought stress) on plants solely based on colour because

they have similar colour variations. In Chapter 3, morphological, texture features and image descriptors were combined with selected colour features to analyze their sensitivities to drought stress and nitrogen variation. A processing and analysis pipeline comprising statistical (Morris method) and machine learning methods (One-at-a-time sensitivity method) was developed to identify the most responsive features to drought and nitrogen deficiency. Both sensitivity methods identified colour features as the most sensitive to nitrogen deficiencies, which agrees with the findings of Xiong et al. (2019), who identified 16 colour features derived from the indices of green colour components as responsive to nitrogen dynamics in plants. Interestingly, the ASM texture feature was identified as the most sensitive to drought stress. Other studies identified colour (greenness dominant trait) and morphological features as sensitive to drought stress. Zhang et al. (2022) proposed two new morphological descriptors; ratio of plant height to the centroid (RHC) and leaf angle, for drought stress analysis in rice species. Their findings demonstrated that, in the presence of water stress, the RHC and the leaf angle distribution parameters exhibited diverging trends, corresponding to the drought characteristics of rice at various growth stages. Duan et al. (2018) also presented four image-derived features: greenness plant area ratio (GPAR), ratio of total plant area and bounding rectangle (TBR), perimeter area ratio (PAR) and ratio of total plant area and convex hull (TCR) for discriminating drought-resistant and drought-sensitive accessions and dynamically quantifying the rice drought response under controlled conditions. From their findings, GPAR and PAR proved effective in quantifying drought stress response in the field. In most situations, multiple phenotypic traits are combined to analyze drought stress in plants due to the dynamic nature of this stress. In such situations, texture features are most present in the feature sets, as evident in the work of Zhuang et al. (2017). They identified texture features (energy features based on wavelet transform) and green channel-based colour features as important to detecting water stress in plants. These were used to build a two-stage ML model that could be used for the early detection of water stress in maize. This suggests that texture features are invaluable in assessing drought stress in plants.

Furthermore, the pipeline identified two main traits: angular second moment (ASM) and histogram of oriented gradients (HOG) features as the most sensitive to drought-stressed plants with variable nitrogen concentration levels. This means that drought stress and

nitrogen deficiency are more pronounced in the texture properties of wheat species. The drought-stressed plants in low-nitrogen soil alter the uniformity of the plant pixels at the canopy level, which are detectable using the ASM features since this feature measures the homogeneity or uniformity of pixel intensities of images. Also, the HOG features were used to compute the gradient information of images, capturing both local texture and shape information. During drought stress, plants undergo various physiological and morphological changes, including leaf wilting and curling, leaf size and area reduction and leaf thickness and stiffness ( Xu et al., 2010). These changes are related to plant shape and pixel intensities, so they are easily detectable using the HOG feature.

## **6.1.2 Hyperspectral imaging for stress analysis**

### **6.1.2.1 Nutrient stress analysis using hyperspectral imaging.**

Hyperspectral cameras capture images across a wide range of wavelengths, typically from the visible through the infrared to the shortwave infrared regions of the electromagnetic spectrum. In hyperspectral imaging (HSI), spectroscopy and imaging methods are combined to create a system that allows for the computation of the spatial map of spectral content variation. In this thesis, the hyperspectral sensor was applied at close range in the glasshouse to acquire imaging data for nutrient and drought analysis in two separate experiments. Although HSI acquires huge datasets giving detailed information about plants, the complex plant architecture and light interactions result in illumination variations that mask the actual spectral responses of plants (Mishra et al., 2017). In proximal HSI sensing of plants, illumination correction is the primary step before feature extraction. Although there is no clear solution to illumination, different chemometrics domains and machine learning methods were employed to manage illumination variability and enhance the quality of spectral data. These include standard normal variate (SNV), second-order derivate (SDeri), and discriminant analysis. With a validation accuracy of 99.18%, the SDeri dataset outperformed the others, indicating the effectiveness of the spectral derivative pre-processing technique in reducing the illumination effect while retaining the pertinent information of the plant spectra. This gives the derivative spectra an advantage over other chemometric methods like SNV and multiplicative scatter correction (MSC). This compares to the works of Yamashita et al. (2020), who suggest that derivative spectra effectively minimize the illumination effect



while retaining the informative parts of the spectra. However, the results of Mishra et al.(2021), who explored different chemometric pre-processing techniques such as SNV and SDeriv, show that the derivate spectra perform poorly in partial least square and DL modelling compared to the raw absorbance data. Moreover, according to Cloutis et al. (1996), the spectral derivative technique is sensitive to the signal-to-noise ratio of HSI data, suggesting that the higher the derivative order, the more susceptible the data is to noise. Nonetheless, this work reduced the spectral-to-noise ratio using the Savitzky-Golay filter to smoothen the noisy HSI data before applying the second derivative approaches. This approach improved the overall performance of the model. Moreover, implementing the spectral derivative analysis was made easier by using continuous spectral bands from wavelength and a high spectral resolution of 0.7 nm (Pu, 2017). Generally, the selection of a pre-processing technique depends on the specific requirement of the analysis, the characteristics of the spectral data and the aim of the study. To choose the method that produces the most accurate and dependable results for a particular application, investigating and contrasting the performance of different pre-processing techniques may be helpful. Also, although the dimensions of the dataset were reduced by applying a linear discriminant analysis (LDA), which improved the computational efficiency, the LDA dataset performed poorly comparably. This occurred probably because the LDA technique may have removed some spectral information sensitive to the traits under study.

While VIs are valuable tools for HSI analysis, their limitations affect their use. Several VIs depend on specific wavelength ranges, which do not capture subtle spectral features within other ranges. Moreover, VIs are derived from averaging spectra measured at specific wavelengths or bands, which affects the spatial variations of the object of interest. To consider the spatial and spectral characteristics of HSI in this study, a hybrid CNN model obtained from the fusion of 3D and 2D CNN blocks was proposed. Although several studies that employed CNN modelling in HSI were successful, there were challenges, including integration of spectral-spatial information, limited training data, the curse of dimensionality, and processing time. As discussed in Chapter 4, the proposed CNN model mitigated some of these challenges. Enough training data for the model was made available by developing small patches of hypercube from the plant dataset. These patches captured the spatial and spectral characteristics of the plants and were used to train the model. A similar approach was adopted by Mu et al. (2021) in classifying the

origin of four Chinese wolfberries. The results revealed that their proposed classifier, when trained with patches of the hypercube, outperformed the conventional SVM model that focuses only on spectral features.

Although hyperspectral data consists of hundreds of continuous images with high spectral resolution, previous studies show that many spectral bands are redundant and multicollinear, increasing the computational cost during processing. This study used a greedy stepwise correlation feature selection technique (CFS) to select the most sensitive wavelengths for each plant species (quinoa and cowpea) to lessen this difficulty and the multicollinearity effect during processing. The results showed that sixteen wavebands were sensitive to phosphorus and nitrogen stress in cowpea. On the other hand, the CFS method identified twelve wavebands in quinoa species that were sensitive to nitrogen and phosphorus variations. The selected wavebands were across the four main regions: visible, near-infrared, infrared and short-wave infrared regions. However, most wavebands were found in the SWIR regions. This region is known to possess relevant information on phosphorus in plants, as revealed in the work of Siedliska et al. (2018). While nitrogen is known to be highly correlated to wavelengths in the green and near-infrared bands (Velasco & Möllers, 2000), this study shows that the SWIR region has wavelengths sensitive to nitrogen. The works of Wijewardane et al. (2023) also showed that VIS-NIR-SWIR leaf spectral libraries could enable rapid and low-cost analysis of several important leaf physiological and biochemical traits (nitrogen, phosphorus, potassium, calcium, magnesium and sulphur), confirming the findings in this study. Additionally, the success of the waveband selection process using the CFS method indicates the potential of fusing the selected wavebands into a multispectral sensing system, which is a viable alternative to a hyperspectral imaging system for plant nutrient stress identification.

The selected wavebands were used to retrain the proposed hybrid CNN model, which performed well but with a slight reduction in accuracy compared to those trained with the full spectra. This could be due to the loss of some sensitive spectral features during the feature selection process. It should be noted that using a single feature selection method such as the CFS technique has merits and setbacks. An ensemble learning model that combines the predictions of multiple selection methods to improve the overall accuracy

and robustness should be used to maximize feature selection methods. The proposed model with the selected spectral features could identify stressed plants (at different growth stages) resulting from low nitrogen or phosphorus content. While the model performed poorly at the earlier growth stage, the performance improved with increasing growth stages. The low accuracy at the earlier growth stage could be due to the minimal differences in canopy spectral properties of the four treatments. At subsequent growth stages, however, the spectral characteristics depicting nitrogen and phosphorus stress were more prevalent, resulting in improved model performance. Also, at the final growth stage, the performance of the model slightly decreased for both species, which could be attributed to plant senescence. Plant senescence has similar spectral characteristics of nitrogen and phosphorus deficiency in plants, leading to many misclassifications of the model in this study.

#### **6.1.2.2 Using hyperspectral imaging for plant drought stress analysis**

Using hyperspectral imaging for non-destructive plant phenotyping is challenging due to the complicated imaging setup and light interactions with plant canopy, which results in noise addition and illumination variations. One common approach in reducing noise and illumination variations is spectral averaging, which calculates the average spectrum over the pixel domain (Mishra et al., 2020). This was employed in Chapter 5 to assess the sensitivities of different wavelengths for their responsiveness to plant drought stress under variable nitrogen levels. From the average spectra, known vegetation indices (VIs) were calculated and analyzed using machine-learning models. Additionally, an ensemble model feature selection method was employed to select sensitive features from which new vegetation indices were proposed. Ten of the 25 known VIs used in this study were highly correlated to the two gas exchange measurements. However, these VIs are known to be linked to nitrogen stress, which explains their correlation to the treatments with different nitrogen levels.

Unlike the feature selection process in Chapter 4, which utilized one model approach, an ensemble model method that aggregates the benefits of multiple models, which provides a robust feature selection pipeline, was proposed in Chapter 5. The ensemble model selected the top 10 spectral features, discarding the remaining. The effectiveness of the ensemble model indicates that a small section of spectral features can capture a significant

amount of the most important information, with the remaining features usually being redundant or adding noise. New VIs were proposed and developed from the selected VIs. The evaluation of the proposed VIs, the known VIs and the combined VIs (combining known and proposed VIs) using three machine learning models shows that the performance of the combined VIs trained models was higher than that of models trained with the known or proposed VIs. Most of the features of the combined VIs were localized in the green, red-edge and infrared regions, which have been reported as spectral bands sensitive to nitrogen deficiency and drought stress in plants (Rotbart et al., 2013), explaining the reason for the high performance of the models trained with these VIs. The effectiveness of the combined VIs was further shown in the regression models for predicting the gas exchange measurements of wheat species. All the models trained with combined features were as good as the partial least square regression models trained with the whole spectra. This suggests that rather than depending solely on the manual measurements of the gas exchange measurements, the combined VIs regression models can be utilised to predict gas exchange readings to monitor drought stress.

#### **6.1.2.3 The trade-off between sensors and phenotyping output**

For image-based non-invasive phenotyping, selecting an appropriate sensing system is essential to achieve the desired output. While the two sensors used in this study have shown promising results to be used for phenotyping the nutritional and drought status in plants, there are several trade-offs between the type of sensor to select. However, in this study, two key factors, (i) sensor and operational cost, (ii) data information, and (iii) data processing and application of sensing technology, were identified as factors affecting the selection of imaging sensors for phenotyping.

##### *(i) Sensor and operational cost*

Cost analysis is complicated and requires multiple variables to evaluate the affordability and efficiency of a digital RGB or hyperspectral phenotyping system. The sensor and processing cost were the two main variables considered in this research. Due to their low cost and portability, digital cameras are the most widely and commonly used sensors for plant phenotyping. Comparably, hyperspectral imaging systems have higher upfront costs due to the complexity of the HSI sensors and optics. Existing hyperspectral cameras can cost from £30,000, while some commercial systems can cost up to £150,000 (Stuart et

al., 2020). Meanwhile, commercial digital cameras for plant phenotyping can cost less than £2000. This means that while most stakeholders can adopt digital cameras in the phenotyping community, hyperspectral cameras are usually used by a section of the research community for research purposes. Finally, hyperspectral cameras are operated by and require specialist professionals from multiple fields, such as equipment setup, data gathering, processing and interpretation, whereas digital RGB cameras require minimal skill to operate. Since hyperspectral imaging collects data over a wide range of spectral bands, it produces large data volumes, necessitating high storage spaces. Comparatively, smaller data files are produced using digital RGB imaging.

(ii) *Data Information*

As explained in Chapters 2 to 5 of this thesis, the hyperspectral dataset comprises hundreds of narrow spectral bands providing comprehensive information about the reflection, absorption and scattering of light by plant tissues. Specific spectral regions are particularly informative for characterizing various aspects of plant physiology. For instance, the visible region (400 nm – 700 nm) is noted for chlorophyll and carotenoids absorbance. As indicated earlier, the hyperspectral imaging was used to identify plant nutrient status and drought stress. The nutrient stresses were concentrated around the green, NIR and SWIR regions, while the NIR region was more sensitive to the drought stress in wheat species. In addition to the nutrient and drought analysis, the hyperspectral dataset provides detailed information on leaf biochemistry, leaf water content, metabolic and photosynthetic processes. For example, the SWIR region (1250- 2500 nm) is sensitive to changes in plant water content, leaf internal structure and soil-plant interactions (Pandey et al., 2017). Hence, researchers rely on this information for the non-invasive analysis of water stress in crops. Despite reports that hyperspectral images have poor spatial resolution, the proximal sensing method used in this study (approximately 3 metres above the ground) ensured that the data had an excellent spatial resolution (1300 x 1000 per scan). On the other hand, RGB images have a high spatial resolution, providing detailed colour and texture information, making it possible to understand changes in plant morphology about stress. However, digital RGB imaging has limited spectral information (only capturing three spectral bands: red, green and blue), restricting its ability to distinguish subtle changes in plant response to stresses.

### *Data processing and applications*

The huge datasets generated by the hyperspectral camera require high-powered computers to process. This research used an Intel Core i7 processor high-powered computer with 32G RAM speed to process the spectral data. This computer is costly compared to the standard core i5, which has 4GB RAM for processing digital RGB images. This shows that digital cameras are advantageous to hyperspectral cameras in terms of both sensor and operational costs. Data processing for digital and hyperspectral imaging involves distinct approaches due to the differences in the nature of the data they capture. Both sensing methods are important for image-based plant phenotyping because they provide complementary tools for describing the characteristics of plants and how they react to their surroundings. However, each of them has its unique application. Digital imaging provides high-resolution data that can be used for time-series monitoring of plant growth, using parameters such as leaf area, canopy cover and plant height. Also, they can be used to analyze other physical properties of plants, including leaf shape, texture and colour, which enables the detection of symptoms associated with diseases, pest damage and nutrient deficiency. On the other hand, HSI is used to assess the physiological features of plants related to drought, nutrient deficits, and disease resistance. Genome-wide association studies can also use HSI to find genetic loci connected to particular spectral features. This facilitates the discovery of genetic markers for breeding programs.

Ultimately, this study shows that the choice of sensor for stress analysis is based on the use objectives, budget constraints, and available expertise.

## **6.2 Machine learning and statistical methods for plant phenotyping**

Machine learning played an important role in this research, enabling the analysis of large digital and hyperspectral imaging datasets. Traditional machine learning methods (support vector machines, random forest and decision trees) and deep learning methods (convolutional and deep neural networks) were employed for feature extraction/selection, trait classification and prediction for stress identification, pattern interpretability and visualization.

### 6.2.1 Feature Extraction and Selection

Different machine learning techniques were employed to extract or select features to enhance model performance. In Chapter 2, colour and texture features were extracted after image annotation and the most relevant features were selected using a random forest-based recursive feature elimination method (RF-RFE). The RF-RFE is a machine-learning technique that trains and iteratively removes the least important features based on a score metric. It effectively selected the most informative features while eliminating the redundant ones, ensuring a robust classification. That is, with a reduced number of training features, the processing time of the classification model was reduced by 62.8%, which is important in real-time processing, where high computational speed is required for the batch processing of images.

Furthermore, in Chapter 3, three supervised models were trained to recognise plant drought stress at various view orientations using eleven hand-crafted features. The models were trained with individual and combined features. The results show that ML models trained with combined features outperformed those trained with individual features. This means that although the individual features had high interpretability (easy to interpret and understand) since each feature represents a specific aspect of the data, the combined features helped capture complex interactions between variables that might not be apparent when considering them individually. Combining the features often involves feature engineering techniques, which can effectively encode domain information and enhance model performance. A linear discriminant analysis (LDA) tool was also employed to reduce the dimensions of the histogram of oriented gradient (HOG) and scale-invariant feature transform (SIFT) features only two discriminants (features). LDA aims to reduce the dimensions of a large dataset to a lower-dimensional space by finding a linear combination of input features to find the feature subspace that maximizes class separability. In both situations (HOG and SIFT features), the LDA inherently selected the features that most significantly contribute to the separation between the different treatments. Also, LDA was employed in Chapter 4 to reduce the spectral dimensions from 200 to 2 sensitive spectra. Comparing the results of the LDA-feature model to a principal component analysis (PCA) derived feature model, the LDA-features performed better in separating the four nutrient treatments under study. In all cases, LDA reduced redundancy and improved the model's generalization.

### **6.2.2 Trait analysis for stress classification and prediction**

This research developed multiple models to identify stress (by classification) or predict (by regression) plant stresses. In Chapter 2, three supervised classification models (support vector machines, random forest and multilayer perceptron) were trained with selected features to classify pixels into plants and non-plants for binary segmentation. The results demonstrated an improved segmentation quality using the ML models compared to colour-index-based segmentation techniques. This finding agrees with Adams et al. (2020), who observed an improvement in the robustness of their segmentation model when ML models were used to learn colour features from multiple colour orientations. Additionally, the strong performance of the MLP model, fast processing speed, and low training features requirements demonstrate how well it can be deployed on edge devices such as mobile phones, Raspberry Pi and other devices. To evaluate the performance of the proposed segmentation method in Chapter 2, an eXtreme gradient boosting (XGBoost) was implemented as a regression tool to predict SPAD readings of plants. The XGBoost regression models derived from the various segmentation techniques demonstrated that when images from the suggested segmentation method were used, models with comparable high predictive power and accuracy were obtained. The proposed technique can learn from different environmental conditions with a high level of robustness. XGBoost model was used because it has a high computing efficiency and works well with structured datasets for regression and classification predictive modelling (Chen & Guestrin, 2016).

Similarly, the three classification models (SVM, RF and DT) trained with the hand-crafted features in Chapter 3 for identifying drought stress under variable nitrogen levels achieved a high accuracy score irrespective of the angle of orientation of the plants. The results show that the models were robust to noise and offered regularization techniques to prevent overfitting while improving generalization performance. Moreover, from multiple studies, conventional ML models such as DT, RF and SVM offer flexibility in effectively handling nonlinear relationships and complex interactions with input features and output variables (Tan et al., 2005). This explains their use in Chapter 3 of this research.



The conventional ML models in this research performed well because they were used on limited training data. However, modelling the high-dimensional hyperspectral data requires deep learning models like convolutional neural networks (CNN) and RNN. Deep learning models are suitable for learning complex spatial and spectral representations from hyperspectral data, capturing detailed patterns and relationships. In Chapter 4, a hybrid CNN model was proposed to identify nitrogen and phosphorus stress in cowpea and quinoa plants at different growth stages. Although previous studies utilized other CNN architectures such as 1D-, 2D and 3D-CNN models for hyperspectral data analysis, their feature extraction methods have limitations, which can affect their performance. While 1D-CNN extracts only spectral information, 2D-CNN deals with spatial information without considering their spectral variabilities. Although 3D-CNN considers both the spectral and spatial information of hyperspectral data, their performance is affected by the high processing time. The proposed model combines 3D-CNN and separable 2D-CNN architectures, which learn spatial and spectral features and patterns from the hypercube data. Using separable convolutions in the 2D-CNN block reduces the model's complexity by reducing the number of trainable parameters, subsequently reducing the processing time.

### **6.2.3 Selecting ML and deep learning models for plant phenotyping**

Findings in this research show that both conventional ML and DL techniques have their strengths and weaknesses and can be effective for data analysis depending on the specific task, data patterns and characteristics and available resources. While conventional ML models were effective in feature extraction and sensitive feature selection, they were also better at interpreting the relationship between the input features and output predictions. For example, the RF and SVM models have high interpretability, offering an easy understanding of the relationship between input features and output results. Also, the ML models were used on limited training data with relatively low-dimensional feature spaces. This offers a computational advantage requiring less computation resources and processing time. On the other hand, deep learning architectures were used on data with complex patterns and relationships since they can learn intricate patterns from the different dimensional orientations of the data. They were mainly used in the processing and analysis of hyperspectral data. Deep learning eliminates the need for manual feature

engineering (extraction and selection) as it performs an end-to-end learning process, learning features and patterns while performing classification/regression in the network. Lastly, since deep learning modelling depends on large number of datasets, the enormous spatial and spectral information from the hyperspectral dataset effectively helped achieve a state-of-the-art performance of the proposed hybrid CNN model, as shown in Chapter 4. This shows that the decision to select a conventional ML or advanced deep learning architecture for analysis of plant phenotypic data is influenced by variables including the data size, the need for interpretability of the model, available computational resources and the objective of the studies.

### **6.3 Research implications- novelty, impact and contribution to knowledge**

The original contribution to knowledge from this research is the development of different image analysis pipelines to identify or predict nutrient and drought stresses in selected crops. Using two imaging sensing technologies, digital and hyperspectral cameras-classification and prediction models were developed to assess the nutritional and drought status of plants. A processing pipeline with a novel segmentation function to segment plants from different environmental scenarios was developed using digital imaging. Nutrient-sensitive (nitrogen and phosphorus) features were identified through a proposed feature engineering technique. These sensitive features can track both nitrogen and phosphorus deficiencies in plants. Also, plant features sensitive to drought under variable nitrogen content were identified using two sensitivity analysis tools. These features can be monitored to assess the early onset of drought in plants. Furthermore, these sensitive features can precisely identify nitrogen deficiency before they manifest. A proposed fast and lightweight multilayer perceptron model had similar performance as a state-of-the-art deep learning model, indicating the possibility of using it on edge devices for phenotyping the nutrient and drought status of crops.

In addition, a deep learning model was developed using hyperspectral datasets to detect the onset of nutrient deficiency in cowpea and quinoa. This is a promising tool for researchers to use to select spectral bands with more informative and discriminatory responses to nutrient (especially nitrogen) deficiencies. Calibrating it on quinoa plants, this tool gave information on the most relevant spectral bands that responds to changes in nutrient status at any plant growth stage. Further studies and calibration on other plants

will be needed to make this tool viable for research in image-based plant phenotyping. The rapid development of high-resolution imaging spectroscopy sensors opens new opportunities to investigate the reflectance spectra of individual plant organs, leading to an understanding of canopy biophysical and chemical characteristics. The application of the plant classification spectroscopy tool will be helpful to farm management practices and research, as it can automatically characterize the spectral variations of cowpea and quinoa leaves in multiple situations of nutrient deficiency.

Developing a simplistic model capable of identifying the early onset of drought stress is a contribution to the scientific community. This was achieved by modelling and proposing new VIs for drought stress. This study shows that combining the proposed VIs with selected known VIs yielded high accuracy in identifying plant drought stress. Using the combined VIs is essential in building lightweight models as this does not require extracting a full set of image features.

#### **6.4 Research Implication to Africa's Agriculture**

In light of Africa's projected population doubling by 2050, food security is a major priority. It is estimated that Africa must increase its food production by 50% in order to meet the demands of the growing population. Major plant nutrients including N and P are crucial for plant development and output. However, nutrient deficits, notably nitrogen and phosphorus, hinder production, particularly in Sub-Saharan Africa, where soils are frequently depleted of nutrients. One critical challenge of Africa's agricultural system is the excessive or under fertilizer application which degrades soil, contaminate water supplies, and increase greenhouse gas emissions.

This research enables targeted fertilization through the development of more precise and effective ways to identify phosphorus and nitrogen deficiencies in plants using simple RGB imaging analysis as well as complex spectroscopy analysis. The combination of machine learning with image analysis as used in this work will help farmers and researchers in identifying key physiological traits of plants which are sensitive to N and P deficiencies. This when used effectively can optimize the use of limited resources for plant yield maximization. Furthermore, upon further research, the combined utilization of plant spectrum characteristics with the power of machine learning can help plant breeders and researchers in Africa to predict plant yield and make more informed

fertilization decisions, lowering costs and environmental effects, which is critical for Africa's sustainable agricultural methods. This research contributes to Africa's sustainable agriculture by developing methods and techniques for precise detection of plant nutrient levels, which minimizes overapplication and thus reducing environmental damage. In effect, this research will help to improve long-term soil health in Africa by optimizing the use of N and P and other vital fertilizers.

This study proposes ways for diagnosing drought stress under different nitrogen levels which will help African farmers and researchers to identify with further studies, drought resistance germplasm, which can also be sustainable under different nitrogen levels. This in the long term will help African farmers to maintain consistent yields despite adverse weather conditions on the field. Additionally, leveraging technology for high-throughput phenotyping is essential, enabling researchers and breeders to evaluate plant features and their responses on a large scale. This is particularly useful in Africa, where plant breeding for nutrient-efficient and drought resistant cultivars has the potential to increase agricultural production. Improving the proposed methods and effectively implementing the proposed technologies in African breeding programs will promote the quick production of agricultural varieties tailored to the regions's specific environmental and resource restrictions. Finally, the tools and models developed in this study could be translated into field applications, which will enable African farmers to monitor crop health and make data driven decisions for crop management. African farmers and breeders may improve crop management by incorporating these techniques into local extension programs. Moreover, African breeding programs may use these high - throughput, imaging and machine learning based phenotyping approaches to choose nutrient-efficient crop varieties that are appropriate for local conditions, enhancing productivity at a regional scale

## 6.5 References

- Adams, J., Qiu, Y., Xu, Y., & Schnable, J. C. (2020). Plant segmentation by supervised machine learning methods. *Plant Phenome Journal*, 3(1), 1–11. <https://doi.org/10.1002/ppj2.20001>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Aug, 785–794. <https://doi.org/10.1145/2939672.2939785>

- Cloutis, E. A., Connery, D. R., Dover, F. J., & Major, D. J. (1996). Airborne multi-spectral monitoring of agricultural crop status: Effect of time of year, crop type and crop condition parameter. *International Journal of Remote Sensing*, 17(13), 2579–2601. <https://doi.org/10.1080/01431169608949094>
- Coy, A., Rankine, D., Taylor, M., Nielsen, D. C., & Cohen, J. (2016). Increasing the accuracy and automation of fractional vegetation cover estimation from digital photographs. *Remote Sensing*, 8(7), 21–25. <https://doi.org/10.3390/rs8070474>
- Duan, L., Han, J., Guo, Z., Tu, H., Yang, P., Zhang, D., Fan, Y., Chen, G., Xiong, L., Dai, M., Williams, K., Corke, F., Doonan, J. H., & Yang, W. (2018). Chapter 3. Novel digital features discriminate between drought resistant and drought sensitive rice under controlled and field conditions. *Frontiers in Plant Science*, 9(April). <https://doi.org/10.3389/fpls.2018.00492>
- Khan, M. A., & AlGhamdi, M. A. (2023). An intelligent and fast system for detection of grape diseases in RGB, grayscale, YCbCr, HSV and L\*a\*b\* color spaces. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-023-17446-8>
- Mishra, P., Lohumi, S., Ahmad Khan, H., & Nordon, A. (2020). Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. *Computers and Electronics in Agriculture*, 178(July), 105780. <https://doi.org/10.1016/j.compag.2020.105780>
- Mishra, P., Sadeh, R., Ryckewaert, M., Bino, E., Polder, G., Boer, M. P., Rutledge, D. N., & Herrmann, I. (2021). A generic workflow combining deep learning and chemometrics for processing close-range spectral images to detect drought stress in *Arabidopsis thaliana* to support digital phenotyping. *Chemometrics and Intelligent Laboratory Systems*, 216(May), 104373. <https://doi.org/10.1016/j.chemolab.2021.104373>
- Mu, Q., Kang, Z., Guo, Y., Chen, L., Wang, S., & Zhao, Y. (2021). Hyperspectral image classification of wolfberry with different geographical origins based on three-dimensional convolutional neural network. *International Journal of Food Properties*, 24(1), 1705–1721. <https://doi.org/10.1080/10942912.2021.1987457>
- Pandey, P., Ge, Y., Stoerger, V., & Schnable, J. C. (2017). High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging. *Frontiers in Plant Science*, 8(August), 1–12. <https://doi.org/10.3389/fpls.2017.01348>
- Pu, R. (2017). Hyperspectral remote sensing: Fundamentals and practices. In *Hyperspectral Remote Sensing: Fundamentals and Practices* (pp. 1–466). <https://doi.org/10.1201/9781315120607>
- Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., Shenderey, C., Dag, A., & Yermiyahu, U. (2013). Estimating olive leaf nitrogen concentration

- using visible and near-infrared spectral reflectance. *Biosystems Engineering*, 114(4), 426–434. <https://doi.org/10.1016/j.biosystemseng.2012.09.005>
- Sadeghi-Tehran, P., Virlet, N., Sabermanesh, K., & Hawkesford, M. J. (2017). Chapter 2. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. *Plant Methods*, 13(1), 1–17. <https://doi.org/10.1186/s13007-017-0253-8>
- Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., & Sun, Q. (2021). Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Computers and Electronics in Agriculture*, 180(October 2020), 105860. <https://doi.org/10.1016/j.compag.2020.105860>
- Siedliska, A., Baranowski, P., Zubik, M., Mazurek, W., & Sosnowska, B. (2018). Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biology and Technology*, 139(February), 115–126. <https://doi.org/10.1016/j.postharvbio.2018.01.018>
- Stuart, M. B., Stanger, L. R., Hobbs, M. J., Pering, T. D., Thio, D., McGonigle, A. J. S., & Willmott, J. R. (2020). Low-cost hyperspectral imaging system: Design and testing for laboratory-based environmental applications. *Sensors (Switzerland)*, 20(11), 1–13. <https://doi.org/10.3390/s20113293>
- Tan, Pang-Ning, Steinbach, M., Adeyeye Oshin, M., Kumar, V., & Vipin. (2005). *Introduction to Data Mining*.
- Velasco, L., & Möllers, C. (2000). Use of near-infrared reflectance spectroscopy to assess nitrogen concentration in different plant tissues of rapeseed. *Communications in Soil Science and Plant Analysis*, 31(19–20), 2987–2995. <https://doi.org/10.1080/00103620009370644>
- Wang, Y., Wang, D., Shi, P., & Omasa, K. (2014). Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. *Plant Methods*, 10(1). <https://doi.org/10.1186/1746-4811-10-36>
- Wijewardane, N. K., Zhang, H., Yang, J., Schnable, J. C., Schachtman, D. P., & Ge, Y. (2023). A leaf-level spectral library to support high-throughput plant phenotyping: predictive accuracy and model transfer. *Journal of Experimental Botany*, April. <https://doi.org/10.1093/jxb/erad129>
- Xiong, X., Zhang, J., Guo, D., Chang, L., & Huang, D. (2019). Non-invasive sensing of nitrogen in plant using digital images and machine learning for brassica campestris ssp. Chinensis L. *Sensors (Switzerland)*, 19(11). <https://doi.org/10.3390/s19112448>
- Xu, Z., Zhou, G., & Shimizu, H. (2010). Plant responses to drought and rewatering. *Plant Signaling and Behavior*, 5(6), 649–654. <https://doi.org/10.4161/psb.5.6.11398>

- Yamashita, H., Sonobe, R., Hirono, Y., Morita, A., & Ikka, T. (2020). CHAPTER 4: Dissection of hyperspectral reflectance to estimate nitrogen and chlorophyll contents in tea leaves based on machine learning algorithms. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-73745-2>
- Zadoks, J. C., & Board, E. (1999). Data sheet highlights close coupled pumps. *World Pumps*, 1999(388), 9. [https://doi.org/10.1016/s0262-1762\(99\)80614-2](https://doi.org/10.1016/s0262-1762(99)80614-2)
- Zhang, Y., Zha, Y., Jin, X., Wang, Y., & Qiao, H. (2022). Changes in Vertical Phenotypic Traits of Rice (*Oryza sativa* L.) Response to Water Stress. *Frontiers in Plant Science*, 13(July), 1–19. <https://doi.org/10.3389/fpls.2022.942110>
- Zhuang, S., Wang, P., Jiang, B., Li, M., & Gong, Z. (2017). Chapter 3. Early detection of water stress in maize based on digital images. *Computers and Electronics in Agriculture*, 140, 461–468. <https://doi.org/10.1016/j.compag.2017.06.022>

## **7 Conclusions and recommendations**

### **7.1 Summary**

Plant phenotyping provides essential information for selecting high-yield, disease and pest-resistant genotypes. However, using conventional methods is invasive, time-consuming and expensive, resulting in a bottleneck for plant breeding. High-precision plant phenotyping can be achieved through non-invasive image processing methods combined with machine learning algorithms. This thesis focuses on developing image processing and machine learning models for high-throughput phenotyping of the nutritional and drought status of plants under different environmental conditions. A key to good classification or predictive models is using the appropriate features sensitive to the phenotypic traits of interest to support different machine learning algorithms. Various methods were employed to understand the sensitivities of digital and hyperspectral imaging-based plant phenotypes and their responsiveness to the dynamics in nutrient and drought stress. The research findings revealed that colour and texture features are sensitive to nitrogen deficiency, while drought stress is more pronounced in texture features such as ASM when using digital imaging. Meanwhile, nitrogen stress was expressed more in the visible and SWIR regions of the electromagnetic spectrum. Various feature extraction techniques were employed at different stages to assess the importance of these features to the stress under study. Further analysis of these extracted data enabled the early detection of plant stress before becoming visible. Having robust features was essential for multi-year classification and prediction models, especially when dealing with data collected from multiple sensors and platforms over a period of time. This study shows that when trained with these features, traditional machine learning and deep learning models were robust in classifying nutrient deficiency and predicting drought stress in plants.

### **7.2 Key Objectives and Findings**

As mentioned in Chapter One, this thesis aims to develop an image processing and machine learning framework for phenotyping the nutrition and drought stress in selected crops. To achieve this objective, specific objectives were defined and evaluated. The specific objectives, general findings and contributions for each chapter are given below:



**Chapter 2:** The objective was to develop an image processing pipeline based on a novel segmentation approach to assess plant phenotypes responsive to nutrient deficiencies in cowpea and quinoa in the glasshouse and wheat on the field. An image processing workflow was proposed to segment field and glasshouse-based plants automatically. From the segmented images, colour features with a good relationship with SPAD values of plants were used to train ML models to predict plant SPAD readings of cowpea (glasshouse) and wheat (field). Since the segmentation tool was developed based on colour information, it faced several challenges, such as extreme light reflection. The tool was improved by including GLCM texture features in the training dataset. The improved segmentation tool had good performance as a state-of-the-art UNet deep learning tool. The proposed pipeline, which can be deployed on edge devices such as mobile phones, Raspberry Pi and robotic arms, is a practical solution for pre-processing digital images, segmenting and extracting features for high throughput phenotyping.

**Chapter 3:** In this Chapter, the sensitivities of selected plant features and their responsiveness to plant drought stress under variable nitrogen availability were sought. Two sensitivity analysis mechanisms were adopted: Morris (statistically based) and OAT (ML-based) sensitivity. From both analyses, colour features, especially DGCI and RGBVI, were identified as highly sensitive to nitrogen deficiency, which is consistent with several previous works. Moreover, the sensitivity analysis identified ASM and HOG as particularly responsive to drought-stress plants with changing nitrogen content, even though variables such as GPB, RGBVI, and HULL were all sensitive. When conventional ML models were trained with the extracted features, the performance was high when combined features were used as against the use of single features. Furthermore, a proposed 13-layer CNN model identified drought-stress plants achieving a convincing performance comparable to a state-of-the-art model (DenseNet) despite its low number of training parameters. This shows that the proposed CNN model could be used instead of other state-of-the-art models to identify plant drought stress. The proposed CNN model was further fused with the extracted features to identify drought stress in wheat. The results revealed a boost in the performance of the model compared to the previous model. Farmers and plant breeders could use the identified sensitive features to non-invasively detect the early onset of nitrogen deficiency or drought stress in plants by monitoring the changes in these features over time.

**Chapter 4:** This chapter aimed to develop an image analysis pipeline along with proposed deep learning models to characterize the spatial and spectral variations of plants for monitoring the dynamics of plant nutrients across selected plant growth stages using hyperspectral data. One challenge of hyperspectral data processing is illumination. This was managed by experimenting with various chemometrics approaches, including standard normal variate, second-order derivative and discriminant analysis. A hybrid CNN model was proposed comprising a 3D block and 2D separable CNN. The findings show that second derivative spectra remove noise and limit the impact of illumination on the spectra, achieving high classification accuracy when used to train the traditional and proposed CNN model. The proposed model was observed to outperform the conventional deep learning model. The use of separable convolution in the 2D block of the proposed model lowers the number of trainable parameters, reducing the processing time while improving the model's efficiency. A correlation features selection method was employed to select the wavelengths responsive to nitrogen and phosphorus in the cowpea and quinoa species. The experiment findings show that based on the selected wavebands from the second-order derivative of the reflectance spectra, the proposed model can reliably distinguish between plant nitrogen and phosphorus levels. The success of the waveband selection process indicates the potential of developing a multispectral sensor system with the selected wavebands as a viable alternative to hyperspectral imaging for nutrient stress detection.

**Chapter 5:** This chapter evaluates the use of hyperspectral data integrated with machine learning algorithms to identify drought stress in wheat under variable nitrogen content. An ensemble model was developed to select sensitive wavelengths responsive to nitrogen deficiency and drought stress. New vegetation indices were proposed using a combination of the selected wavelengths based on some adopted equations from the literature. Traditional ML models were trained with known VIs, proposed VIs and a combination of known and proposed VIs. From the results, the proposed VIs outperformed the known VIs, yielding  $> 0.94$  accuracies for all three models; however, the performance improved when they were trained with the combined VIs. The combined VIs were further used to train three regression models to predict  $g_s$  and  $P_n$ . From the results, the RFR had the best performance, indicating that this can be utilized as a proxy tool to predict  $g_s$  and  $P_n$  and to monitor drought stress in wheat.

### 7.3 Recommendations for Future Work

Applying image processing and machine learning methods to non-invasive phenotyping crops is no longer considered a potential solution or concept for high throughput plant phenotyping but a reality. The work in this thesis has significantly contributed to the development of image processing pipelines and data analysis algorithms for nutrient and drought stress analysis. It has also identified areas for future study, which include:

- To improve phenotyping capabilities for nutrient and drought stress analysis, a wide range of sensors for phenotyping are recommended. The development of low-cost specific sensors such as 3D light detection and ranging (LiDAR) and RGB-IR cameras makes them a practical choice for deployment in image-based plant phenotyping. 3D LiDAR provides detailed three-dimensional data of plant canopy and architecture. The RGB imaging used in this thesis provides a 2D analysis of plant phenotypes. Colour, texture and morphology were detected to be valuable for nutrient and drought stress analysis. With information on plant height volume and other 3D parameters, an insight into the dynamics of plants response to nutrient and drought stress could be achieved. Although hyperspectral imaging is made of a broad spectrum of wavebands offering many advantages for various applications in plant phenotyping, the equipment cost, limited spatial resolution and processing complexity are limitations. The findings in this study suggest that the visible and IR regions were sensitive to nutrient and drought stress, as reported in the literature. Hence, hybrid sensing cameras such as RGB-IR or RGB-NIR can be used as a multispectral tool for low-cost phenotyping plant nutrients and drought stress.
- To guarantee the precision, reliability and suitability for real-world applications of the developed tools, validations of these tools in field conditions are recommended. Most of the experiments in the thesis were conducted in glasshouse settings, which offer unique advantages and demerits. Although the glasshouse environment is controlled, providing uniformity in lighting and disease management, they may be limited, restricting the number of plants that can be grown and phenotyped simultaneously. In contrast, field experiments are often more scalable and represent real-world scenarios. Hence, the tools are

recommended for use on field crops to understand the influence of other environmental factors on the precision of the developed tools.

- This thesis focuses on analyzing multiple plant stresses: nutrient (nitrogen and phosphorus) and drought-nutrient (water stress and nitrogen deficiency). Most image-based phenotyping research involves a single stress analysis. Further work should consider the simultaneous analysis of multiple stresses. Using image processing techniques and computational tools, researchers can efficiently phenotype plants under various stress conditions. This allows for the quick and high-throughput evaluation of stress tolerance and adaptation in breeding programmes, ecological studies, and agricultural management practices.

## **7.4 Concluding Remarks**

Traditional phenotyping methods have been a bottleneck for non-destructive phenotyping of crops to improve yield. This PhD work has developed and evaluated various image processing pipelines for non-invasive nutrient and drought stress phenotyping in selected crops. Different image processing frameworks and powerful machine learning tools have been used to create pipelines for pre-processing, post-processing and data analysis to understand the dynamics of plant responses to nutrient and water variations in plants. The methods developed in this study offer a superior advantage in knowing and understanding plant physiology and spectral traits responsiveness to multiple stresses. They can also identify stressed or nutrient-deficient crops before they show signs of stress. However, further work is required using low-cost sensors and advanced machine-learning techniques to make these tools as practical as possible.

## Appendices

### Appendix A Ethical approval letter



24 March 2020

Dear Mr Okyere ,

Reference: CURES/9737/2020

Title: Image analysis for plant phenotyping--machine learning based methods for analysis of multi-model and multi-dimensional remote sensing data from high-throughput plant phenotyping

Thank you for your application to the Cranfield University Research Ethics System (CURES).

**We are pleased to inform you your CURES application, reference CURES/9737/2020 has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

## Appendix B Supplementary Materials for Chapter 2

**Table B-1** Training, validation and testing data the classification mode

| Features               | Total<br>Pixels | Training Data<br>(pixels) | Validation<br>Data (pixels) | Testing<br>Data |
|------------------------|-----------------|---------------------------|-----------------------------|-----------------|
| All Features (AF)      | 24,000,000      | 14,000,000                | 4,800,000                   | 4,800,000       |
| Selected Features (SF) | 12,000,000      | 7,200,000                 | 2,400,000                   | 2,400,000       |

**Table B-2** Training, validation and testing data for the different machine learning models

| Test Dataset | Model | Processing time (s) |         |
|--------------|-------|---------------------|---------|
|              |       | AF                  | SF      |
| 10           | MLP   | 5.150               | 2.450   |
|              | SVM   | 132.158             | 86.163  |
|              | RF    | 81.225              | 44.581  |
| 15           | MLP   | 9.443               | 2.931   |
|              | SVM   | 187.055             | 105.434 |
|              | RF    | 156.355             | 58.145  |
| 30           | MLP   | 17.220              | 6.052   |
|              | SVM   | 277.474             | 155.181 |
|              | RF    | 188.553             | 98.336  |

### B.1 Images of cowpea plants



21 DAS



28 DAS



35 DAS



42 DAS



49 DAS



56 DAS

(A) Low nitrogen high phosphorus (LNHP)



21 DAS



28 DAS



35 DA



42 DAS



49 DAS



56 DAS

(B) Low nitrogen low phosphorus (LNL)



21 DAS



28 DAS



35 DAS



42 DAS  
(C) High nitrogen high phosphorus (LNHP)



49 DAS



56 DAS



21 DAS



28 DAS



35 DAS





42 DAS                      49 DAS                      56 DAS  
(D)                      High nitrogen low phosphorus (HNLP)

**Figure B-1.** Examples of cowpea plants from the side view. (A) low N high P (LNHP), (B) high N high P (HNHP) (C) low N low P (LNLP) (D) high N low P (HNLP). N and P represents nitrogen and phosphorus respectively while DAS represent days after sowing

## B.2 Images of wheat (Field



183 DAS



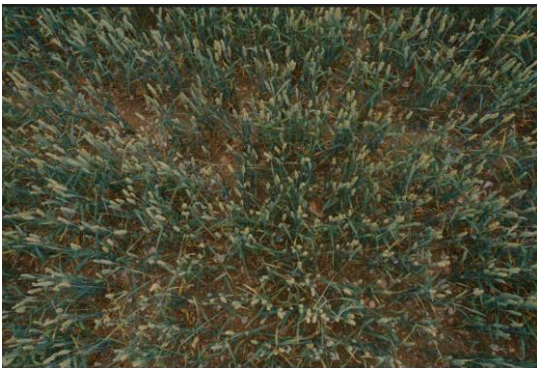
198 DAS



213 DAS



228 DAS



243 DAS



265 DAS

(A) Wheat plants with 3 ppm phosphorus level





183 DAS



198 DAS



213 DAS



228DAS



243 DAS



265 DAS

(B) Wheat plants with 60 ppm phosphorus level

**Figure B-2.** Sample images of the wheat plants under different phosphorus levels (A) wheat plants with 3 ppm and (B) wheat plants with 60ppm phosphorus levels. DAS indicate days after sowing.

### B.3 U-Net Architecture for Image Segmentation

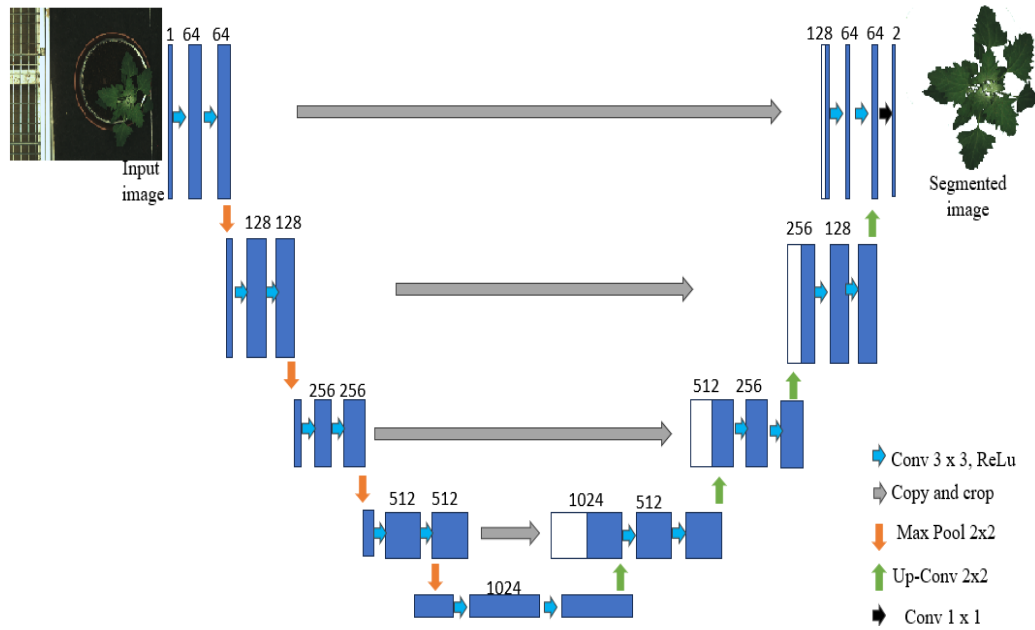


**Figure B-3.** U-Net Architecture (Ronneberger et al., 2015)

### B.4 VGG -16 network architecture

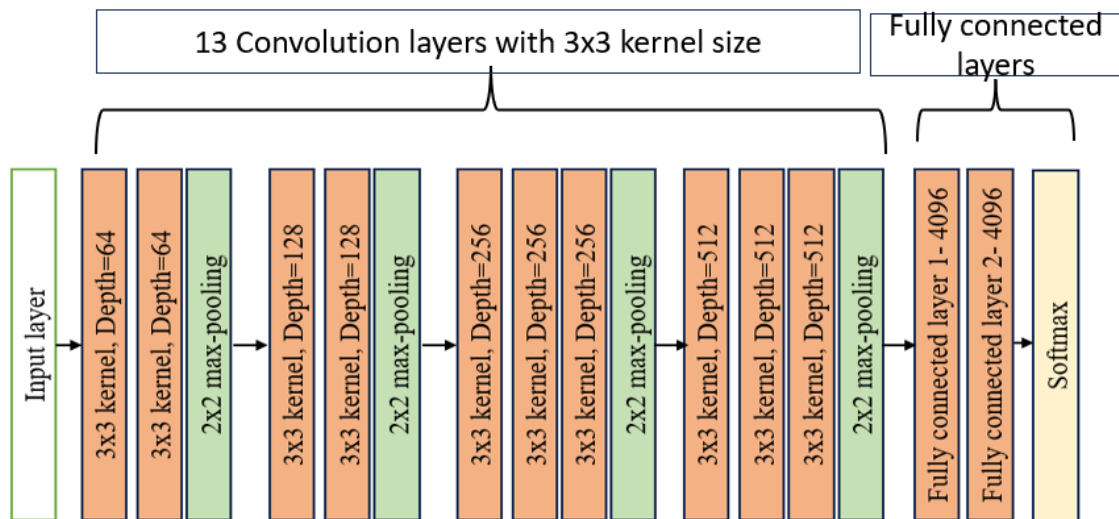


**Figure B-4.** VGG-16 model architecture – 13 convolutional layers and 2 Fully connected layers and 1 SoftMax classifier

## Appendix C - Supplementary Materials for Chapter 3

**Table C-1. The Proposed CNN model algorithm structure**

| Layer No. | Layers                         | Output shape | Parameters |
|-----------|--------------------------------|--------------|------------|
| 1         | conv2d (Conv2D)                | (68, 68, 8)  | 224        |
| 2         | max_pooling2d (MaxPooling2D)   | (34, 34, 8)  | 0          |
| 3         | batch normalization            | 34, 34, 8    | 32         |
| 4         | conv2d_1 (Conv2D)              | 32, 32, 16   | 1168       |
| 5         | max_pooling2d_1 (MaxPooling2D) | 16, 16, 16   | 0          |
| 6         | batch_normalization_1          | 16, 16, 16   | 64         |
| 7         | conv2d_2 (Conv2D)              | 14, 14, 32   | 4640       |
| 8         | max_pooling2d_2 (MaxPooling2D) | 7, 7, 32     | 0          |
| 9         | conv2d_3 (Conv2D)              | 5, 5, 64     | 18496      |
| 10        | max_pooling2d_3 (MaxPooling2D) | 2, 2, 64     | 0          |
| 11        | flatten                        | 256          | 0          |
| 12        | dense (Dense)                  | 128          | 32896      |
| 13        | dense_1 (Dense)                | 4            | 516        |

**Table C-2 Parameter estimates for mixed anova model with group and timepoint effects**

| Effect                  | Coef.  | Std.Err. | z      | P> z  | Range [0.025 - 0.975] |        |
|-------------------------|--------|----------|--------|-------|-----------------------|--------|
| Intercept               | 0.833  | 0.015    | 53.842 | 0.000 | 0.803                 | 0.863  |
| group[T.DSLN]           | -0.027 | 0.007    | -3.699 | 0.000 | -0.041                | -0.013 |
| group[T.WWHN]           | -0.002 | 0.008    | -0.296 | 0.767 | -0.018                | 0.013  |
| group[T.WWHN]           | -0.016 | 0.008    | -1.963 | 0.050 | -0.032                | 0.000  |
| timepoint               | 0.003  | 0.002    | 1.57   | 0.116 | -0.001                | 0.007  |
| group[T.DSLN]:timepoint | -0.003 | 0.001    | -4.146 | 0.000 | -0.005                | -0.002 |
| group[T.WWHN]:timepoint | 0.003  | 0.001    | 3.726  | 0.000 | 0.001                 | 0.005  |
| group[T.WWLN]:timepoint | 0.003  | 0.001    | 3.101  | 0.002 | 0.001                 | 0.004  |
| Group Var               | Var    | 0        | 0.030  |       |                       |        |



### C.1 Images of wheat for drought experiment



(i) WWLN      (ii) DSHN      (iii) DSLN      (iv) WWHN  
(a) 0 DADS



(i) WWLN      (ii) DSHN      (iii) DSLN      (iv) WWHN  
(b) 3 DADS



(i) WWLN      (ii) DSHN      (iii) DSLN      (iv) WWHN  
(c) 6 DADS



(i) WWLN

(ii) DSHN

(iii) DSLN

(iv) WWHN

(d) 9 DADS



(i) WWLN

(ii) DSHN

(iii) DSLN

(iv) WWHN

(d) 12 DADS



(i) WWLN

(ii) DSHN

(iii) DSLN

(iv) WWHN

(e) 15 DADS

**Figure C-1.** Samples of wheat plants in the glasshouse for the drought experiments. Samples were gathered over the 14 days experiment span (from 0 -15 DADS). The treatments include; WWLN, DSHN, DSLN and WWHN.

## C.2 Loss and accuracy curves

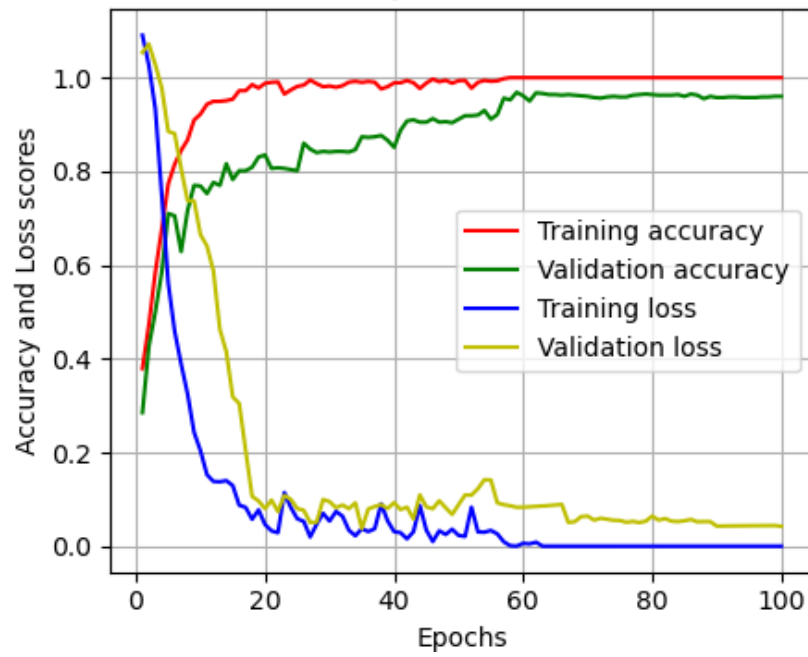


**Figure C-2** Loss and accuracy cuves on the training and validation sets of the proposed CNN model



## Appendix D Supplementary Materials for Chapter 4

**Table D-1** Selected growth stages of quinoa and cowpea using the BBCH system (DAT is the day after transplant)

| Quinoa       |     |                                                                     |
|--------------|-----|---------------------------------------------------------------------|
| Growth Stage | DAT | Description                                                         |
| I            | 19  | Nine pair of leaves visible                                         |
| II           | 29  | Nine side shoots visible                                            |
| III          | 36  | Stem elongation                                                     |
| IV           | 50  | Inflorescence present but still enclosed by leaves                  |
| V            | 69  | Complete anthesis: main inflorescence flowers with senesced anthers |
| Cowpea       |     |                                                                     |
| Growth Stage | DAT | Description                                                         |
| I            | 13  | 3 leaves (with stipules) unfolded, or 3 tendrils developed          |
| II           | 30  | Beginning of stem elongation                                        |
| III          | 51  | First flower buds visible outside leaves                            |
| IV           | 60  | First flowers open (sporadically within the population)             |
| V            | 72  | 20% of pods have reached typical length                             |

**Table D-2** Detailed architecture of HybridSN model

| Layer No. | Layer Type          | Layer details                                              | Activation |
|-----------|---------------------|------------------------------------------------------------|------------|
| 1         | Conv3D              | filters (8), kernel ( $3 \times 3 \times 7$ ), stride (1)  | ReLU       |
| 2         | Conv3D              | filters (16), kernel ( $3 \times 3 \times 5$ ), stride (1) | ReLU       |
| 3         | Batch Normalization | 256                                                        | -          |
| 4         | Max Pooling         | Pool size (2, 2, 2)                                        | -          |
| 5         | Conv3D              | filters (32), kernel ( $3 \times 3 \times 3$ ) stride (1)  | ReLU       |
| 6         | Conv3D              | filters (64), kernel ( $3 \times 3 \times 7$ ) stride (1)  | ReLU       |
| 7         | Batch Normalization | 128                                                        | -          |
|           | Max Pooling         | Pool size (2, 2, 2)                                        | -          |
| 8         | Flatten             | 516                                                        | -          |
| 9         | Dense               | Units (256), dropout (0.4)                                 | ReLU       |
| 10        | Dense               | Units (128), dropout (0.4)                                 | ReLU       |
| 11        | Dense               | Output units (4)                                           | SoftMax    |

**Table D-3** Architectural details of 3D-CNN model for classification of plant nutrient status

| Layer No. | Layer Type          | Layer details                                              | Activation |
|-----------|---------------------|------------------------------------------------------------|------------|
| 1         | Conv3D              | filters (8), kernel ( $3 \times 3 \times 7$ ), stride (1)  | ReLU       |
| 2         | Conv3D              | filters (16), kernel ( $3 \times 3 \times 5$ ), stride (1) | ReLU       |
| 3         | Batch Normalization | 256                                                        | -          |
| 4         | Max Pooling         | Pool size (2, 2, 2)                                        | -          |
| 5         | Conv3D              | filters (32), kernel ( $3 \times 3 \times 3$ ) stride (1)  | ReLU       |
| 6         | Conv3D              | filters (64), kernel ( $3 \times 3 \times 7$ ) stride (1)  | ReLU       |
| 7         | Batch Normalization | 128                                                        | -          |
|           | Max Pooling         | Pool size (2, 2, 2)                                        | -          |
| 8         | Flatten             | 516                                                        | -          |
| 9         | Dense               | Units (256), dropout (0.4)                                 | ReLU       |
| 10        | Dense               | Units (128), dropout (0.4)                                 | ReLU       |
| 11        | Dense               | Output units (4)                                           | SoftMax    |

**Table D-4** Architectural details of 2D-CNN model for classification of plant nutrient status

| Layer No. | Layer Type      | Layer details                         | Activation |
|-----------|-----------------|---------------------------------------|------------|
| 1         | Conv2D          | Filters (32), kernel ( $3 \times 3$ ) | ReLU       |
| 2         | Max Pooling     | Pool size (2, 2)                      | -          |
| 3         | Conv2D          | filters = 64, kernel = $3 \times 3$   | ReLU       |
| 4         | Max Pooling     | Pool size (2, 2)                      | -          |
| 6         | Flatten         | (256)                                 |            |
| 7         | Fully connected | No. of Units (128), dropout (0)       | ReLU       |
| 8         | Dense           | Output units (4)- number of classes   | SoftMax    |

## D.1 Preprocessing of spectral images

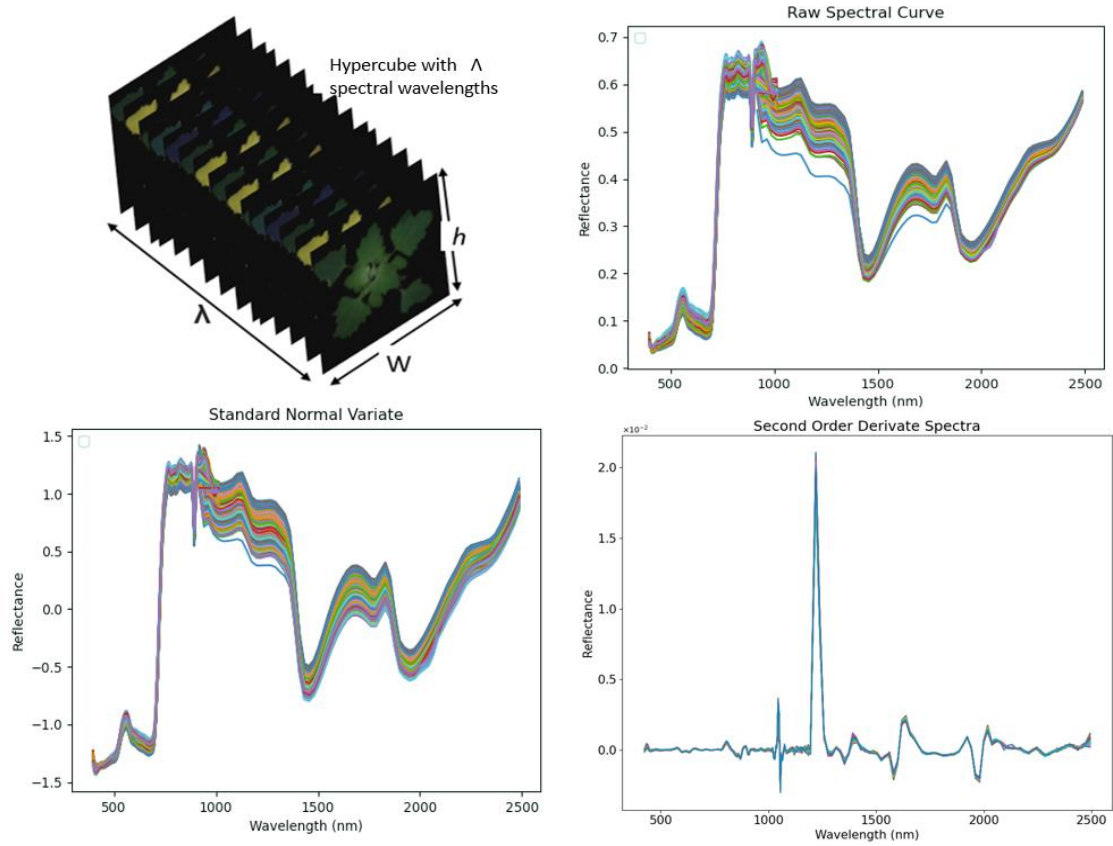


**Figure D-1** Hyperspectral data extraction and different transformation techniques. (a) hypercube data of cowpea and quinoa (b) raw spectral curve (c) SNV-based curve (d) first derivative curve

## D.2 Dimensionality technique (Quinoa)

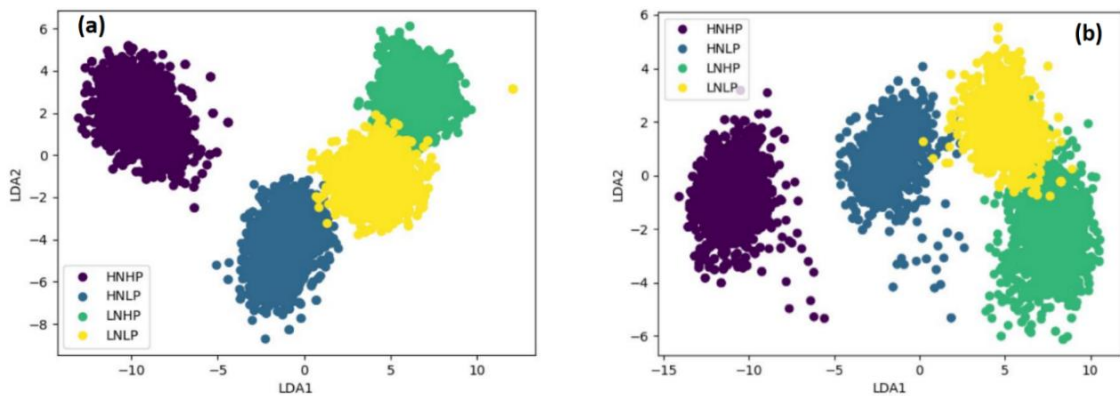


**Figure D-2** scores of scatter plots of the first and second LDA components (a) for cowpea and (b) for quinoa.

### D.3 Dimensionality technique (Cowpea)

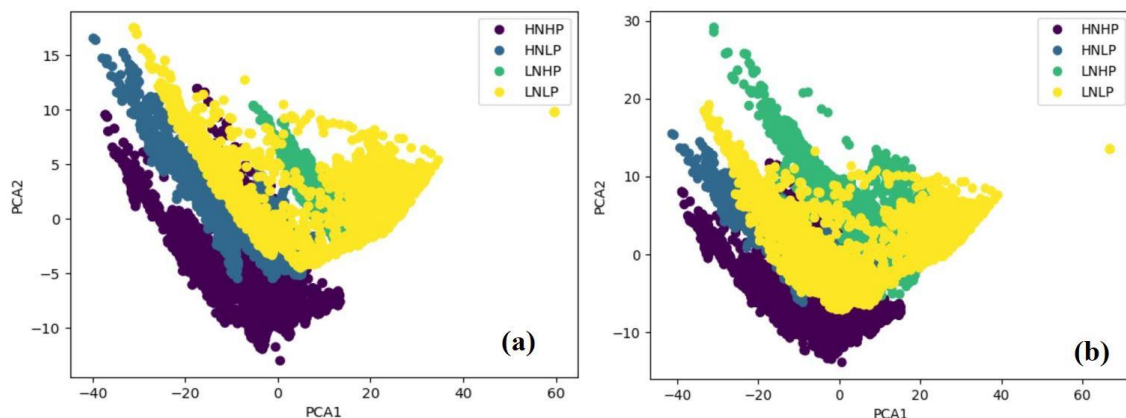


**Figure D-3:** Scores of scatter plots of the first and second LDA components (a) for cowpea and (b) for quinoa

### D.4 Patches extraction processes

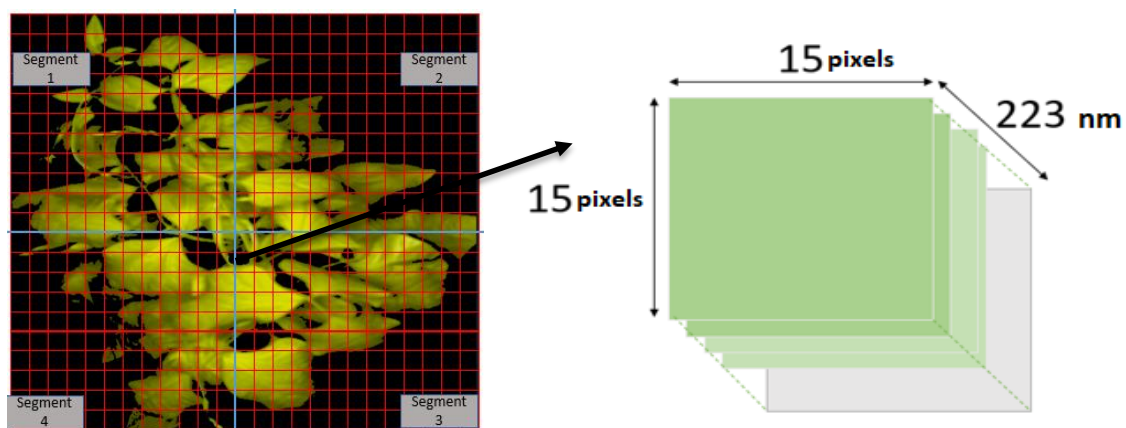
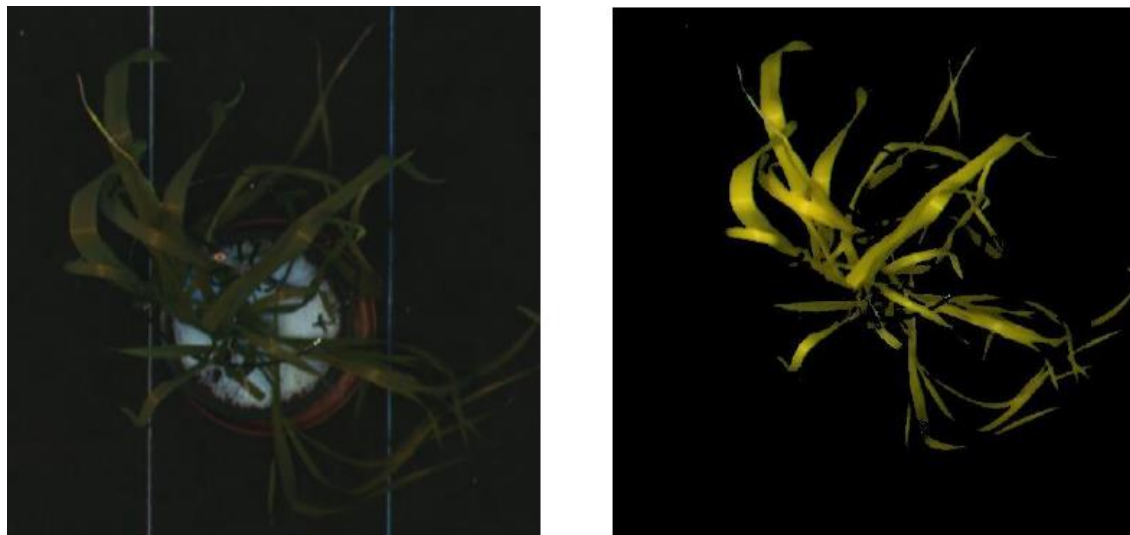


**Figure D-4** Pseudo 'RGB' image showing patches (15 x 15 x 223) for Cowpea.

## Appendix E Supplementary Materials for Chapter 5

### E.1 Segmentation of hyperspectral data



(a) Original pseudo RGB image

(b) Segmented pseudo RGB image

**Figure E-1.** Pseudo RGB image and corresponding Segmented Image (R: 600 nm, G:500 nm, B: 420 nm)

**Table E-1.** Machine learning models for the identification of drought stress in wheat under variable nitrogen levels. This shows the characteristics of each model and why they were chosen for this study.

| Model                        | Characteristics                                                                                                                                                                                                                                                                                                                                                                                      | Reason for choosing the model                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Support Vector Machine (SVM) | <ul style="list-style-type: none"> <li>• Supervised learning model for classification and re-gression tasks.</li> <li>• Principle is to find the optimal hyperplane that separates distinct sample classes with the highest margin</li> <li>• Effective when the number of features is greater than the number of samples</li> </ul>                                                                 | <ul style="list-style-type: none"> <li>• Can handle high-dimensional dataset</li> <li>• Good for linear and non-linear data and is robust to overfitting.</li> </ul>                                                                                                                                                                                                               |
| Random Forest (RF)           | <ul style="list-style-type: none"> <li>• A meta estimator that fits a number of decision tree classifiers on various sub-sampling of dataset increase predictive accuracy and control over-fit-ting.</li> <li>• Uses bootstrapped datasets and randomly selects subsets of features for tree construction.</li> <li>• Reduces overfitting by averaging the predictions of multiple trees.</li> </ul> | <ul style="list-style-type: none"> <li>• RF can identify sensitive/ important features of a da-taset which is useful for understanding driving fea- tures.</li> <li>• RF can handle noisy data and outliers, making it suit-able for this work</li> <li>• RF approach reduces the likelihood of overfitting compared to single decision trees hence used in this study.</li> </ul> |

## Deep Neural Net-work (DNN)

- Has multiple layers (input, hidden, output) with in-terconnected nodes (neurons).
  - Capable of modelling complex, non-linear relation-ships between inputs and outputs.
  - DNNs model is ideal when dealing with large, com-plex and non-linear datasets that some traditional ML models may be unable to handle.
  - DNN modelling is capable of automatically learning relevant features from raw data, reducing the need for manual feature engineering.
  - State-of-the-art performance:
  - Often achieves superior performance in tasks like image and speech recogni-tion, natural language processing, etc
  -
-

**Table E-4** Correlation coefficient of proposed indices with  $g_s$  and  $P_n$ 

| No. | RDI Equation                | Correlation coefficient (r) |       | No. | NDDI Equation                                               | Correlation (r) |       | No. | DDI Equation                | Correlation coefficient (r) |       |
|-----|-----------------------------|-----------------------------|-------|-----|-------------------------------------------------------------|-----------------|-------|-----|-----------------------------|-----------------------------|-------|
|     |                             | $P_n$                       | $g_s$ |     |                                                             | $P_n$           | $g_s$ |     |                             | $P_n$                       | $g_s$ |
| 1   | $\lambda 557 / \lambda 669$ | -0.34                       | -0.54 | 1   | $(\lambda 557 - \lambda 669) / (\lambda 557 + \lambda 669)$ | -0.65           | -0.6  | 1   | $\lambda 557 - \lambda 669$ | -0.28                       | -0.44 |
| 2   | $\lambda 557 / \lambda 674$ | -0.09                       | -0.04 | 2   | $(\lambda 557 - \lambda 674) / (\lambda 557 + \lambda 674)$ | -0.68           | -0.63 | 2   | $\lambda 557 - \lambda 674$ | 0.22                        | -0.41 |
| 3   | $\lambda 557 / \lambda 722$ | -0.58                       | -0.50 | 3   | $(\lambda 557 - \lambda 722) / (\lambda 557 + \lambda 722)$ | -0.72           | -0.66 | 3   | $\lambda 557 - \lambda 722$ | 0.51                        | -0.12 |
| 4   | $\lambda 557 / \lambda 940$ | -0.21                       | -0.54 | 4   | $(\lambda 557 - \lambda 940) / (\lambda 557 + \lambda 940)$ | -0.70           | -0.67 | 4   | $\lambda 557 - \lambda 940$ | 0.09                        | 0.04  |
| 5   | $\lambda 557 / \lambda 957$ | -0.55                       | -0.55 | 5   | $(\lambda 557 - \lambda 957) / (\lambda 557 + \lambda 957)$ | -0.44           | -0.23 | 5   | $\lambda 557 - \lambda 957$ | -0.15                       | -0.01 |
| 6   | $\lambda 557 / \lambda 636$ | -0.63                       | -0.55 | 6   | $(\lambda 557 - \lambda 636) / (\lambda 557 + \lambda 636)$ | -0.28           | -0.20 | 6   | $\lambda 557 - \lambda 636$ | -0.41                       | -0.39 |
| 7   | $\lambda 557 / \lambda 683$ | -0.45                       | -0.55 | 7   | $(\lambda 557 - \lambda 683) / (\lambda 557 + \lambda 683)$ | -0.65           | -0.60 | 7   | $\lambda 557 - \lambda 683$ | -0.33                       | -0.42 |
| 8   | $\lambda 557 / \lambda 542$ | -0.69                       | -0.55 | 8   | $(\lambda 557 - \lambda 542) / (\lambda 557 + \lambda 542)$ | -0.69           | -0.60 | 8   | $\lambda 557 - \lambda 542$ | -0.28                       | -0.42 |
| 9   | $\lambda 669 / \lambda 674$ | -0.69                       | -0.55 | 9   | $(\lambda 669 - \lambda 674) / (\lambda 669 + \lambda 674)$ | -0.61           | -0.60 | 9   | $\lambda 669 - \lambda 674$ | 0.45                        | -0.40 |
| 10  | $\lambda 669 / \lambda 722$ | -0.70                       | -0.55 | 10  | $(\lambda 669 - \lambda 722) / (\lambda 669 + \lambda 722)$ | -0.68           | -0.59 | 10  | $\lambda 669 - \lambda 722$ | 0.45                        | -0.38 |
| 11  | $\lambda 669 / \lambda 940$ | -0.72                       | -0.67 | 11  | $(\lambda 669 - \lambda 940) / (\lambda 669 + \lambda 940)$ | -0.47           | -0.58 | 11  | $\lambda 669 - \lambda 940$ | 0.05                        | -0.09 |
| 12  | $\lambda 669 / \lambda 957$ | -0.71                       | -0.63 | 12  | $(\lambda 669 - \lambda 957) / (\lambda 669 + \lambda 957)$ | -0.07           | -0.58 | 12  | $\lambda 669 - \lambda 957$ | 0.20                        | 0.03  |
| 13  | $\lambda 669 / \lambda 636$ | 0.20                        | 0.20  | 13  | $(\lambda 669 - \lambda 636) / (\lambda 669 + \lambda 636)$ | -0.78           | -0.63 | 13  | $\lambda 669 - \lambda 636$ | -0.05                       | -0.01 |
| 14  | $\lambda 669 / \lambda 683$ | -0.24                       | -0.08 | 14  | $(\lambda 669 - \lambda 683) / (\lambda 669 + \lambda 683)$ | -0.49           | -0.56 | 14  | $\lambda 669 - \lambda 683$ | -0.48                       | -0.45 |
| 15  | $\lambda 669 / \lambda 542$ | 0.61                        | 0.58  | 15  | $(\lambda 669 - \lambda 542) / (\lambda 669 + \lambda 542)$ | -0.41           | -0.30 | 15  | $\lambda 669 - \lambda 542$ | -0.48                       | -0.43 |
| 16  | $\lambda 674 / \lambda 722$ | 0.08                        | 0.45  | 16  | $(\lambda 674 - \lambda 722) / (\lambda 674 + \lambda 722)$ | 0.22            | 0.16  | 16  | $\lambda 674 - \lambda 722$ | -0.35                       | -0.40 |
| 17  | $\lambda 674 / \lambda 940$ | 0.66                        | 0.06  | 17  | $(\lambda 674 - \lambda 940) / (\lambda 674 + \lambda 940)$ | -0.06           | -0.15 | 17  | $\lambda 674 - \lambda 940$ | -0.62                       | -0.36 |
| 18  | $\lambda 674 / \lambda 957$ | -0.21                       | -0.01 | 18  | $(\lambda 674 - \lambda 957) / (\lambda 674 + \lambda 957)$ | -0.75           | -0.52 | 18  | $\lambda 674 - \lambda 957$ | -0.01                       | -0.07 |
| 19  | $\lambda 674 / \lambda 636$ | -0.55                       | -0.07 | 19  | $(\lambda 674 - \lambda 636) / (\lambda 674 + \lambda 636)$ | -0.47           | -0.53 | 19  | $\lambda 674 - \lambda 636$ | 0.08                        | 0.03  |
| 20  | $\lambda 674 / \lambda 683$ | 0.58                        | -0.11 | 20  | $(\lambda 674 - \lambda 683) / (\lambda 674 + \lambda 683)$ | -0.59           | -0.49 | 20  | $\lambda 674 - \lambda 683$ | -0.05                       | -0.01 |
| 21  | $\lambda 674 / \lambda 542$ | -0.53                       | -0.14 | 21  | $(\lambda 674 - \lambda 542) / (\lambda 674 + \lambda 542)$ | -0.62           | -0.48 | 21  | $\lambda 674 - \lambda 542$ | -0.44                       | -0.38 |
| 22  | $\lambda 722 / \lambda 940$ | -0.07                       | -0.17 | 22  | $(\lambda 722 - \lambda 940) / (\lambda 722 + \lambda 940)$ | -0.74           | -0.65 | 22  | $\lambda 722 - \lambda 940$ | -0.33                       | -0.35 |
| 23  | $\lambda 722 / \lambda 957$ | -0.22                       | -0.19 | 23  | $(\lambda 722 - \lambda 957) / (\lambda 722 + \lambda 957)$ | -0.67           | -0.66 | 23  | $\lambda 722 - \lambda 957$ | -0.52                       | -0.30 |
| 24  | $\lambda 722 / \lambda 636$ | -0.62                       | -0.20 | 24  | $(\lambda 722 - \lambda 636) / (\lambda 722 + \lambda 636)$ | -0.53           | -0.63 | 24  | $\lambda 722 - \lambda 636$ | -0.01                       | -0.05 |
| 25  | $\lambda 722 / \lambda 683$ | -0.15                       | -0.10 | 25  | $(\lambda 722 - \lambda 683) / (\lambda 722 + \lambda 683)$ | -0.23           | -0.63 | 25  | $\lambda 722 - \lambda 683$ | 0.01                        | 0.03  |



|    |                             |       |       |    |                                                             |       |       |    |                             |       |       |
|----|-----------------------------|-------|-------|----|-------------------------------------------------------------|-------|-------|----|-----------------------------|-------|-------|
| 26 | $\lambda 722 / \lambda 542$ | 0.08  | 0.06  | 26 | $(\lambda 722 - \lambda 542) / (\lambda 722 + \lambda 542)$ | -0.03 | -0.63 | 26 | $\lambda 722 - \lambda 542$ | -0.19 | -0.01 |
| 27 | $\lambda 940 / \lambda 957$ | -0.45 | -0.28 | 27 | $(\lambda 940 - \lambda 957) / (\lambda 940 + \lambda 957)$ | -0.78 | -0.63 | 27 | $\lambda 940 - \lambda 957$ | -0.41 | -0.31 |
| 28 | $\lambda 940 / \lambda 636$ | -0.37 | -0.57 | 28 | $(\lambda 940 - \lambda 636) / (\lambda 940 + \lambda 636)$ | -0.74 | -0.61 | 28 | $\lambda 940 - \lambda 636$ | -0.26 | -0.26 |
| 29 | $\lambda 940 / \lambda 683$ | -0.29 | -0.22 | 29 | $(\lambda 940 - \lambda 683) / (\lambda 940 + \lambda 683)$ | -0.7  | -0.63 | 29 | $\lambda 940 - \lambda 683$ | -0.22 | -0.03 |
| 30 | $\lambda 940 / \lambda 542$ | -0.68 | -0.62 | 30 | $(\lambda 940 - \lambda 542) / (\lambda 940 + \lambda 542)$ | -0.56 | -0.63 | 30 | $\lambda 940 - \lambda 542$ | 0.23  | 0.03  |
| 31 | $\lambda 957 / \lambda 636$ | -0.72 | -0.65 | 31 | $(\lambda 957 - \lambda 636) / (\lambda 957 + \lambda 636)$ | -0.72 | -0.67 | 31 | $\lambda 957 - \lambda 636$ | -0.16 | -0.01 |
| 32 | $\lambda 957 / \lambda 683$ | -0.33 | -0.61 | 32 | $(\lambda 957 - \lambda 683) / (\lambda 957 + \lambda 683)$ | 0.22  | 0.35  | 32 | $\lambda 957 - \lambda 683$ | -0.22 | -0.19 |
| 33 | $\lambda 957 / \lambda 542$ | -0.23 | -0.23 | 33 | $(\lambda 957 - \lambda 542) / (\lambda 957 + \lambda 542)$ | -0.51 | 0.01  | 33 | $\lambda 957 - \lambda 542$ | -0.53 | -0.02 |
| 34 | $\lambda 636 / \lambda 683$ | -0.25 | -0.20 | 34 | $(\lambda 636 - \lambda 683) / (\lambda 636 + \lambda 683)$ | -0.75 | -0.67 | 34 | $\lambda 636 - \lambda 683$ | 0.22  | 0.03  |
| 35 | $\lambda 636 / \lambda 542$ | -0.74 | -0.67 | 35 | $(\lambda 636 - \lambda 542) / (\lambda 636 + \lambda 542)$ | -0.5  | -0.6  | 35 | $\lambda 636 - \lambda 542$ | -0.25 | -0.01 |
| 36 | $\lambda 683 / \lambda 542$ | -0.68 | -0.60 | 36 | $(\lambda 683 - \lambda 542) / (\lambda 683 + \lambda 542)$ | -0.55 | -0.6  | 36 | $\lambda 683 - \lambda 542$ | -0.21 | -0.01 |

---

**Table E-2.** Parameter estimates for mixed anova model with group and timepoint effects for P<sub>n</sub> values

|                         | Coef.  | Std.Err. | z      | P> z  | [0.025 - | 0.975] |
|-------------------------|--------|----------|--------|-------|----------|--------|
| Intercept               | 26.899 | 4.217    | 6.378  | 0.000 | 18.634   | 35.165 |
| group[T.DSLN]           | -2.126 | 5.964    | -0.356 | 0.721 | -13.816  | 9.563  |
| group[T.WWHN]           | 1.992  | 5.964    | 0.334  | 0.738 | -9.698   | 13.681 |
| group[T.WWLN]           | -5.341 | 5.964    | -0.896 | 0.370 | -17.031  | 6.34   |
| timepoint               | -1.681 | 0.179    | -9.389 | 0.000 | -2.032   | -1330  |
| timepoint:group[T.DSLN] | 0.317  | 0.253    | 1.251  | 0.211 | -0.180   | 0.813  |
| timepoint:group[T.WWHN] | 1.307  | 0.253    | 5.161  | 0.000 | 0.810    | 1.803  |
| timepoint:group[T.WWLN] | 0.981  | 0.253    | 3.873  | 0.000 | 0.484    | 1477   |
| Group Var               | 15142  |          |        |       |          |        |

**Table E-3.** Parameter estimates for mixed anova model with group and timepoint effects for g<sub>s</sub> values

| Effects                            | Coefficient | Standard Error | P-value |
|------------------------------------|-------------|----------------|---------|
| Intercept                          | 0.257       | 0.055          | 0.000   |
| group[T.DSLN]                      | 0.002       | 0.078          | 0.981   |
| group[T.WWHN]                      | 0.074       | 0.078          | 0.339   |
| group[T.WWLN]                      | -0.015      | 0.078          | 0.846   |
| timepoint[T.12 DADS]               | -0.235      | 0.039          | 0.000   |
| timepoint[T.15 DADS]               | -0.246      | 0.039          | 0.000   |
| timepoint[T.3 DADS]                | -0.012      | 0.039          | 0.764   |
| timepoint[T.6 DADS]                | -0.125      | 0.039          | 0.001   |
| timepoint[T.9 DADS]                | -0.104      | 0.039          | 0.007   |
| group[T.DSLN]:timepoint[T.12 DADS] | 0.050       | 0.055          | 0.361   |
| group[T.WWHN]:timepoint[T.12 DADS] | 0.215       | 0.055          | 0.000   |
| group[T.WWLN]:timepoint[T.12 DADS] | 0.194       | 0.055          | 0.000   |
| group[T.DSLN]:timepoint[T.15 DADS] | -0.003      | 0.055          | 0.953   |
| group[T.WWHN]:timepoint[T.15 DADS] | 0.223       | 0.055          | 0.000   |
| group[T.WWLN]:timepoint[T.15 DADS] | 0.199       | 0.055          | 0.000   |
| group[T.DSLN]:timepoint[T.3 DADS]  | -0.004      | 0.055          | 0.941   |
| group[T.WWHN]:timepoint[T.3 DADS]  | 0.012       | 0.055          | 0.832   |
| group[T.WWLN]:timepoint[T.3 DADS]  | 0.025       | 0.055          | 0.645   |
| group[T.DSLN]:timepoint[T.6 DADS]  | 0.111       | 0.055          | 0.044   |
| group[T.WWHN]:timepoint[T.6 DADS]  | 0.171       | 0.055          | 0.002   |
| group[T.WWLN]:timepoint[T.6 DADS]  | 0.126       | 0.055          | 0.022   |
| group[T.DSLN]:timepoint[T.9 DADS]  | 0.044       | 0.055          | 0.424   |
| group[T.WWHN]:timepoint[T.9 DADS]  | 0.124       | 0.055          | 0.024   |
| group[T.WWLN]:timepoint[T.9 DADS]  | 0.110       | 0.055          | 0.045   |
| Group Var                          | 1.000       | 67108864       | 0.999   |

**Table E-5** Anova results for F1-score for all the models

| Training dataset | Models             |                    |                    |
|------------------|--------------------|--------------------|--------------------|
|                  | RF                 | SVM                | DNN                |
| Known Vis        | 0.925 <sup>a</sup> | 0.881 <sup>b</sup> | 0.933 <sup>a</sup> |
| Proposed Vis     | 0.911 <sup>a</sup> | 0.93 <sup>b</sup>  | 0.949 <sup>c</sup> |
| Combined Vis     | 0.984 <sup>a</sup> | 0.982 <sup>a</sup> | 0.979 <sup>a</sup> |
| PCA Features     | 0.962 <sup>a</sup> | 0.94 <sup>b</sup>  | 0.900 <sup>c</sup> |